PREFACE

In the auricular structure introduced by this University for students of Post-Graduate degree programme, the opportunity to pursue Post-Graduate course in Subject introduced by this University is equally available to all learners. Instead of being guided by any presumption about ability level, it would perhaps stand to reason if receptivity of a learner is judged in the course of the learning process. That would be entirely in keeping with the objectives of open education which does not believe in artificial differentiation.

Keeping this in view, study materials of the Post-Graduate level in different subjects are being prepared on the basis of a well laid-out syllabus. The course structure combines the best elements in the approved syllabi of Central and State Universities in respective subjects. It has been so designed as to be upgradable with the addition of new information as well as results of fresh thinking and analysis.

The accepted methodology of distance education has been followed in the preparation of these study materials. Co-operation in every form of experienced scholars is indispensable for a work of this kind. We, therefore, owe an enormous debt of gratitude to everyone whose tireless efforts went into the writing, editing and devising of a proper lay-out of the materials. Practically speaking, their role amounts to an involvement in invisible teaching. For, whoever makes use of these study materials would virtually derive the benefit of learning under their collective care without each being seen by the other.

The more a learner would seriously pursue these study materials the easier it will be for him or her to reach out to larger horizons of a subject. Care has also been taken to make the language lucid and presentation attractive so mat they may be rated as quality self-learning materials. If anything remains still obscure or difficult to follow, arrangements are there to come to terms with them through the counselling sessions regularly available at the network of study centres set up by the University.

Needless to add, a great deal of these efforts is still experimental—in fact, pioneering in certain areas. Naturally, there is every possibility of some lapse or deficiency here and there. However, these do admit of rectification and further improvement in due course. On the whole, therefore, these study materials are expected to evoke wider appreciation the more they receive serious attention of all concerned.

Professor (Dr.) Manimala Das
Vice-Chancellor

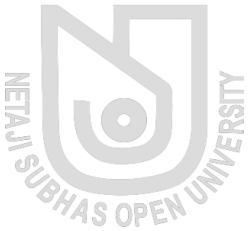
First Reprint : December, 2008

Printed in accordance with the regulations and financial assistance of
the Distance Education Council, Government of India

POST GRADUATE ZOOLOGY
[M.Sc]

PAPER : GROUP
PGZO - 3 : B

	Writer	Editor
Part-I		
Units 1-4	Dr. Kamalesh Mishra	
Units 5-8	Prof. Chandrasekhar Chakraborty	Prof. Chandrasekhar Chakraborty
Units 9-12	Prof. Anil Kumar Saha	
Part-II		
Units 1-8	Dr. Sanjib Kumar Das	Dr. Subir Chandra Dasgupta


NETAJI SUBHAS OPEN UNIVERSITY

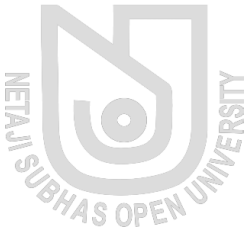
GROUP
B(I)

Part-I : Cytogenetics

GROUP
B(II)
Molecular Biology

GROUP
B (I)

Part-I : Cytogenetics


NETAJI SUBHAS OPEN UNIVERSITY

Unit 1 Biology of Chromosomes

Structure

1.1 Molecular anatomy of eukaryotic chromosomes

The proteins that bind to the DNA to form eukaryotic chromosomes are traditionally divided into two general classes: *histories and the nonhistone chromosomal proteins.* The complex of both classes of protein with the nuclear DNA of eukaryotic cells is known as chromatin. The total mass of histones in chromatin is about equal to that of the DNA. Histones are responsible for the first and most basic level of *chromosome* organization, the nucleosome, which was discovered in 1974. At interphase nuclei most of the chromatin is in the form of a fibre with a diameter of about 30 nm. If this chromatin is subjected to treatments that cause it to unfold partially, it can be seen under the electron microscope as a series of "beads on a string". The string is DNA, and each bead is a "nucleosome core particle" that consists of DNA wound around a protein core formed from histones. The beads on a string represent the first level of chromosomal DNA packing.

The structural organization of nucleosomes was determined after first isolating them from unfolded chromatin by digestion with particular enzymes (called nucleases) that break down DNA by cutting between the nucleosomes. After digestion for a short period, the exposed DNA between the nucleosome core particles, the linker DNA, is degraded. Each individual nucleosome core particle consists of a complex of eight histone proteins—two molecules each of histones H2A, H2B, H3, and H4—and double-stranded DNA that are 146 nucleotide pairs long. The histone octamer forms a protein core around which the double-stranded DNA is wound (Fig. 1.1).

Each nucleosome core particle is separated from the next by a region of linker DNA, which can vary in length from a few nucleotide pairs up to about 80. (The term nucleosome technically refers to a nucleosome core particle plus

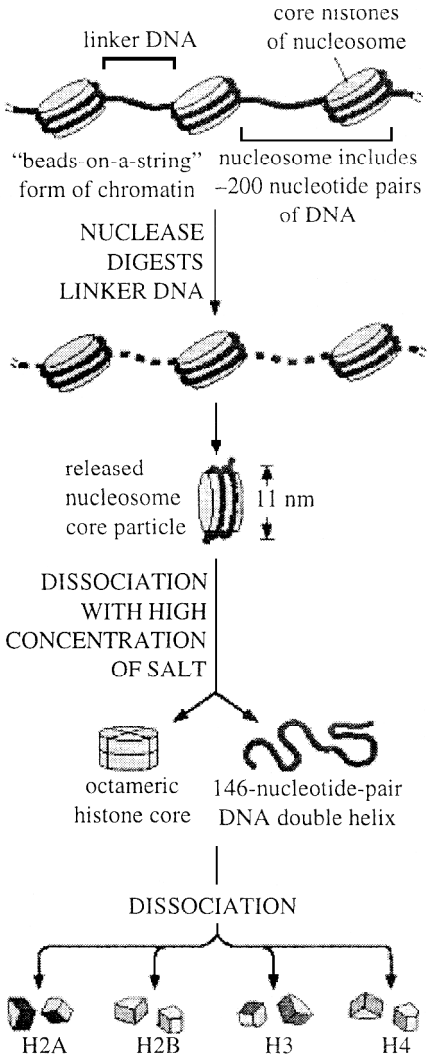


Fig. 1.1 Structural organization of the nucleosome A nucleosome contains a protein core made of eight histone molecules. As indicated, the nucleosome core particle is released from chromatin by digestion of the linker DNA with a nuclease, an enzyme that breaks down DNA. (The nuclease can degrade the exposed linker DNA but cannot attack the DNA wound tightly around the nucleosome core.) After dissociation of the isolated nucleosome into its protein core and DNA, the length of the DNA that was wound around the core can be determined. This length of 146 nucleotide pairs is sufficient to wrap 1.65 times around the histone core

one of its adjacent DNA linkers, but it is often used synonymously with nucleosome core particle.) On average, therefore, nucleosomes repeat at intervals of about 200 nucleotide pairs.

The structure of the nucleosome core particle reveals how DNA is packaged

The high-resolution structure of a nucleosome core particle, solved in 1997, revealed a disc-shaped histone core around which the DNA was tightly wrapped 1.65 turns in a left-handed coil. All four of the histones that make up the core of the nucleosome are relatively small proteins (102-135 amino acids), and they share a structural motif, known as the *histone fold,* formed from three **a** helices connected by two loops (Figure 1.2).

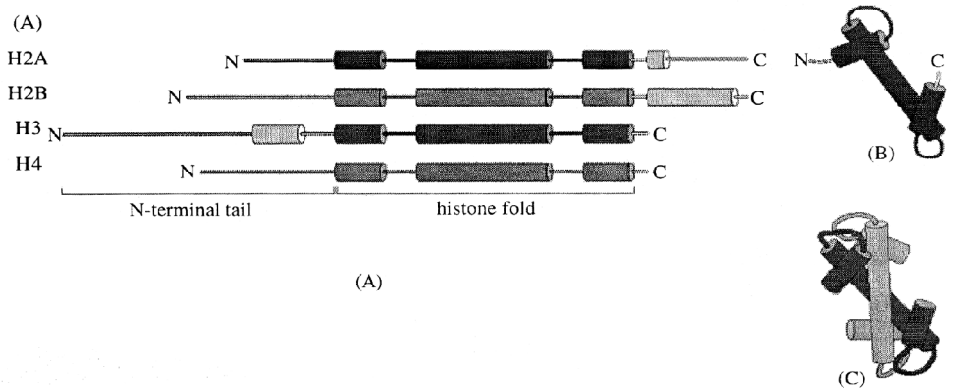


Fig. 1.2 The overall structural organization of the core histones, (A) Each of the core histones contains an N-terminal **tail,** which is subject to several forms of covalent modification, and a histone fold region, as indicated. (B) The structure of the histone fold, which is formed by all four of the core histones. (C) Histones 2A and 2B form a dimer through an interaction known as the "handshake." Histones H3 and H4 form a dimer through the same type of interaction, as illustrated in Figure 1.3

In assembling a nucleosome, the histone folds first bind to each other to form H3-H4 and H2A-H2B dimers, and the H3-H4 dimers combine to form tetramers. An H3-H4 tetramer then further combines with two H2A-H2B dimers to form the compact octamer core, around which the DNA is wound (Fig. 1.3).

The interface between DNA and histone is extensive ; 142 hydrogen bonds are formed between DNA and the histone core in each nucleosome. Nearly half of these bonds form between the amino acid backbone of the histones and the phosphodiester backbone of the DNA. Numerous hydrophobic interactions and salt linkages also hold DNA and protein together in the nucleosome. These numerous interactions explain in part why DNA of virtually any sequence can be bound on a histone octamer core. The path of the DNA around the histone core is not smooth; rather, several kinks are seen in the DNA, as expected from the nonuniform surface of the core. In addition to its histone fold, each of the core histones has a long N-terminal amino acid "tail", which extends out from the DNA-histone core (see Figure 1.3). These histone tails are subject to several different types of covalent modifications, which control many aspects of chromatin structure. The histones are among the most highly conserved eukaryotic proteins. This strong evolutionary conservation suggests that the functions of histones involve nearly all of their amino acids, so that a change in any position is deleterious to the cell. Despite the high conservation of the core histones, many

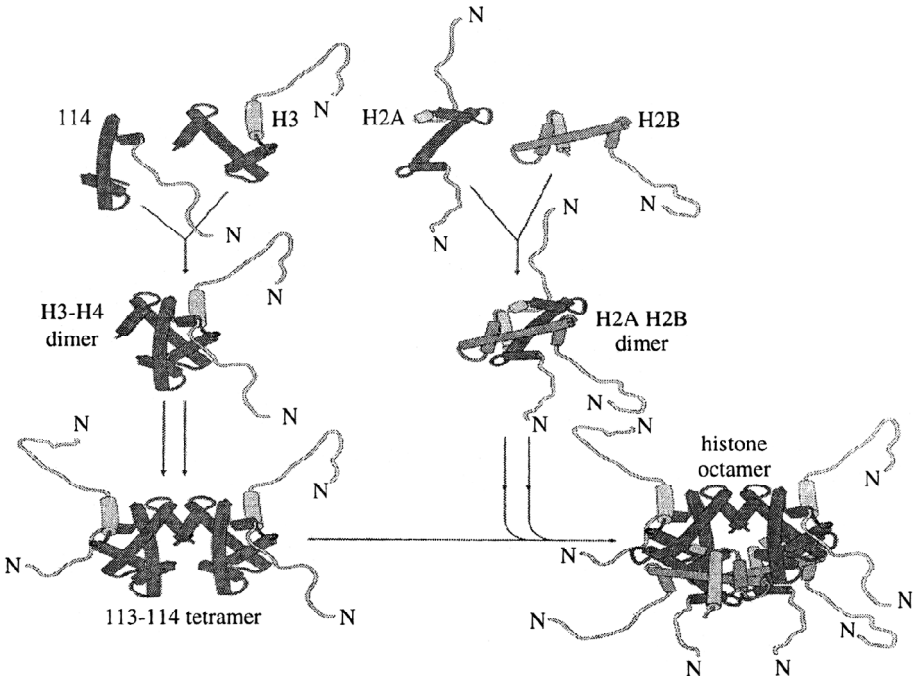


Fig. 1.3 The assembly of a histone octamer. The histone H3-H4 dimer and the H2A-H2B dimer are formed from the handshake interaction. An H3-H4 tetramer forms the scaffold of the octamer onto which two H2A-H2B dimers are added, to complete the assembly. Note that all eight N-terminal tails of the histones protrude from the disc-shaped core structure. In the x-ray crystal, most of the histone tails were unstructured (and therefore not visible in the structure), suggesting that their conformations are highly flexible. (Adapted from figures by J. Waterborg.)

eukaryotic organisms also produce specialized variant core histones that differ in amino acid sequence from the main ones. It is thought that nucleosomes that have incorporated these variant histones differ in stability from regular nucleosomes, and they may be particularly well suited for the high rates of DNA transcription and DNA replication that occur during these early stages of development.

The positioning of nucleosome on DNA is determined by both DNA flexibility and other DNA-bound proteins

Two main influences determine where nucleosomes form in the DNA. One is the difficulty of bending the DNA double helix into two tight turns around the outside of the histone octamer, a process that requires substantial compression of the minor groove of the DNA helix. Because A-T-rich sequences in the minor groove are easier to compress than G-C-rich sequences, each histone octamer tends to position itself on the DNA so as to maximize A-T-rich minor grooves on the inside of the DNA coil. Thus, a segment of DNA that contains short

A-T-rich sequences spaced by an integral number of DNA turns is easier to bend around the nucleosome than a segment of DNA lacking this feature. In addition, because the DNA in a nucleosome is kinked in several places, the ability of a given nucleotide sequence to accommodate this deformation can also influence the position of DNA on the nucleosome.

The second, and probably most important, influence on nucleosome positioning is the presence of other tightly bound proteins on the DNA. Some bound proteins favour the formation of a nucleosome adjacent to them. Others create obstacles that force the nucleosomes to assemble at positions between them. Finally, some proteins can bind tightly to DNA even when their DNA-binding site is part of a nucleosome. The exact positions of nucleosomes along a stretch of DNA therefore depend on factors that include the DNA sequence and the presence and nature of other proteins bound to the DNA.

Nucleosomes are usually packed together into a compact chromatin fibre

The nucleosomes are packed on top of one another, generating regular arrays in which the DNA is even more highly condensed. Thus, when nuclei are very gently lysed onto an electron microscope grid, most of the chromatin is seen to be in the form of a fiber with a diameter of about 30 nm, which is considerably wider than chromatin in the "beads on a string" form. Several models have been proposed to explain how nucleosomes are packed in the 30-nm chromatin fiber; the one most consistent with the available data is a series of structural variations known collectively as the zigzag model. In reality, the 30-nm structure found in chromosomes is probably a fluid mosaic of the different zigzag variations.

Several mechanisms probably act together to form the 30-nm fiber from a linear string of nucleosomes. First, an additional histone, called histone HI, is involved in this process. H1 is larger than the core histones and is considerably less well conserved. A single histone HI molecule binds to each nucleosome, contacting both DNA and protein, and changing the path of the DNA as it exits from the nucleosome. Although it is not understood in detail how H1 pulls nucleosomes together into the 30-nm fiber, a change in the exit path in DNA seems crucial for compacting nucleosomal DNA so that it interlocks to form the 30-nm fibre (Fig. 1.4).

A second mechanism for forming the 30-nm fiber probably involves the tails of the core histones, which, as we saw above, extend from the nucleosome. It is thought that these tails may help attach one nucleosome to another—thereby allowing a string of them, with the aid of histone H1, to condense into the 30-nm fibre (Fig. 1.5).

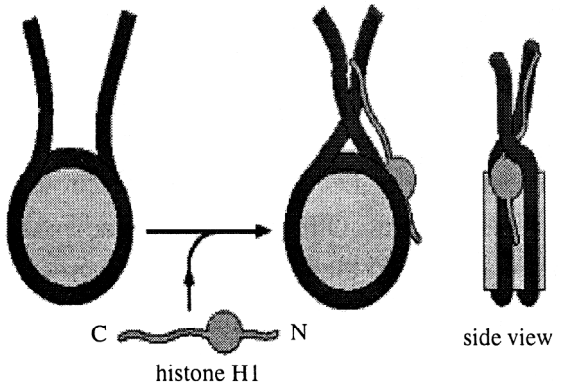


Fig. 1.4 A speculative model for how histone H1 could change the path of DNA as it exits from the nucleosome. Histone H1 (green) consists of a globular core and two extended tails. Part of the effect of H1 on the compaction of nucleosome organization may result from charge neutralization: like the core his-tones, H1 is positively charged (especially its C-terminal tail), and this helps to compact the negatively charged DNA. Unlike the core histones, H1 does not seem to be essential for cell viability; in one ciliated protozoan the nucleus expands nearly two fold in the absence of H1, but the cells otherwise appear normal

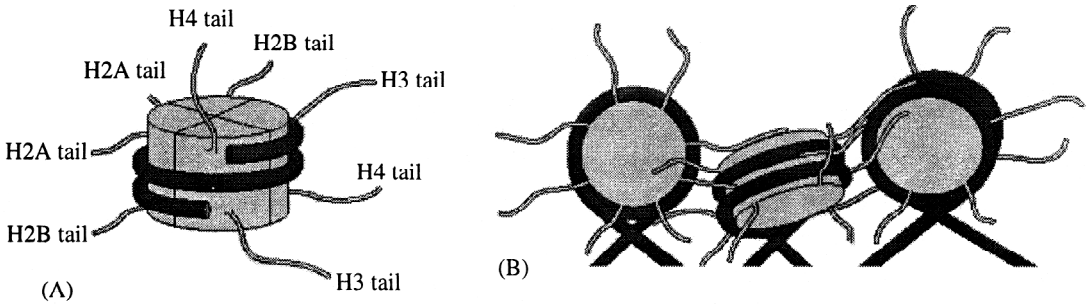


Fig. 1.5 A speculative model for histone tails in the formation of the 30-nm fiber. (A) The approximate exit points of the eight histone tails, four from each histone subunit, that extend from each nucleosome. In the high-resolution structure of the nucleosome, the tails are largely unstructured, suggesting that they are highly flexible. (B) A speculative model showing how the histone tails may help to pack nucieosomes together into the 30-nm fiber. This model is based on (1) experimental evidence that histone tails aid in the formation of the 30-nm fiber, (2) the x-ray crystal structure of the nucleosome, which showed that the tails of one nucleosome contact the histone core of an adjacent nucleosome in the crystal lattice, and (3) evidence that the histone tails interact with DNA

ATP-driven chromatin remodeling machines change nucleosome structure

Eukaryotic cells contain chromatin remodeling complexes, protein machines that use the energy of ATP hydrolysis to change the structure of nucieosomes temporarily so that DNA becomes less tightly bound to the histone core. The remodeled state may result from movement of the H2A-H2B dimers in the nucleosome core; the H3-H4 tetramer is particularly stable and would be difficult to rearrange (Fig. 1.3).

The remodeling of nucleosome structure has two important consequences. First, it permits ready access to nucleosomal DNA by other proteins in the cell, particularly those involved in gene expression, DNA replication, and repair. Even after the remodeling complex has dissociated, the nucleosome can remain

in a "remodeled state" that contains DNA and the full complement of histones—but one in which the DNA-histone contacts have been loosened; only gradually does this remodeled state revert to that of a standard nucleosome. Second, remodeling complexes can catalyze changes in the positions of nucleosomes along DNA (Fig. 1.6); some can even transfer a histone core from one DNA molecule to another.

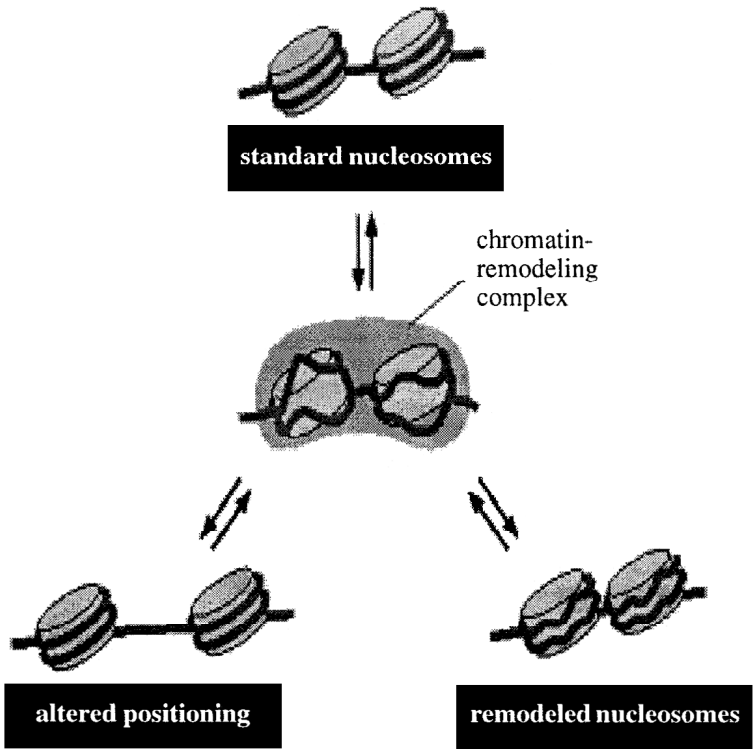


Fig. 1.6 Model for the mechanism of some chromatin remodeling complexes. In the absence of remodeling complexes, the interconversion between the three nucleosomal states shown is very slow because of a high activation energy barrier. Using ATP hydrolysis, chromatin-remodeling complexes (green) create an activated intermediate (shown in the center of the figure) in which the histone-DNA contacts have been partly disrupted. This activated state can then decay to any one of the three nucleosomal configurations shown. In this way, the remodeling complexes greatly increase the rate of interconversion between different nucleosomal states. The remodeled state, in which the histone-DNA contacts have been loosened, has a higher free energy level than that of standard nucleosomes and will slowly revert to the standard nucleosome conformation, even in the absence of a remodeling complex. Cells have many different chromatin remodeling complexes, and they differ in their detailed biochemical properties; for example, not all can change the position of a nucleosome, but all use the energy of ATP hydrolysis to alter nucleosome structure. (Adapted from R.E. Kingston and G.J. Narlikar, Genes Dev. 13:2339-2352, 1999.)

Cells have several different chromatin remodeling complexes that differ subtly in their properties. Most are large protein complexes that can contain more than ten subunits. It is likely that they are used whenever a eucaryotic cell needs direct access to nucleosome DNA for gene expression, DNA replication, or DNA repair. Different remodeling complexes may have features specialized for each of these roles. It is thought that the primary role of some remodeling complexes is to allow access to nucleosomal DNA, whereas that of others is to re-form nucleosomes when access to DNA is no longer required (Fig. 1.7).

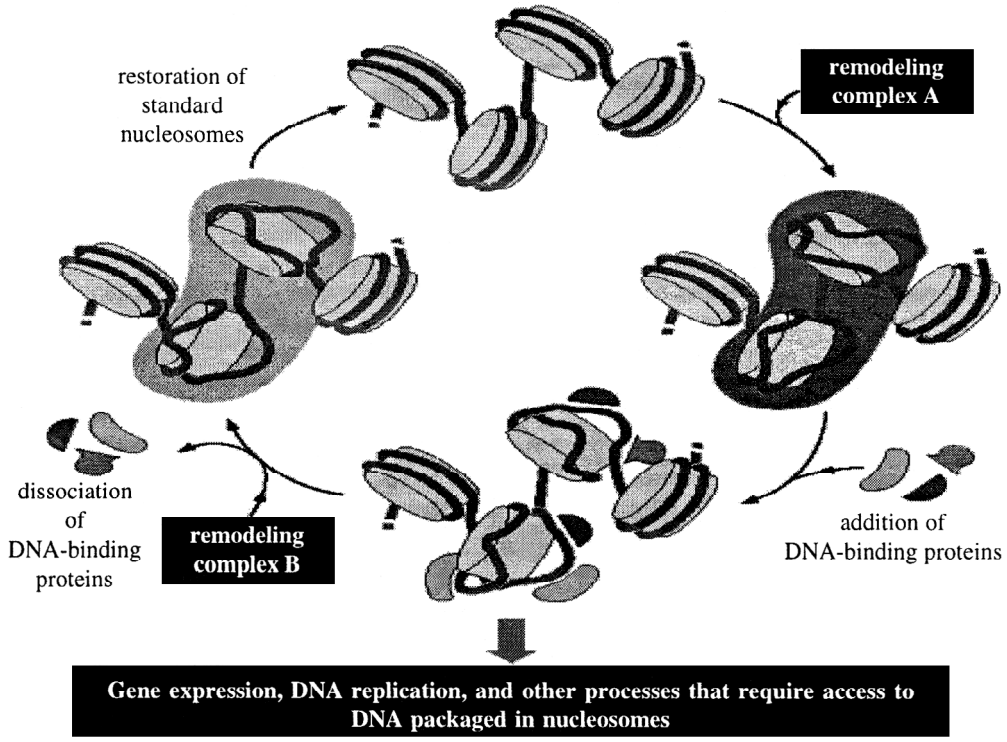
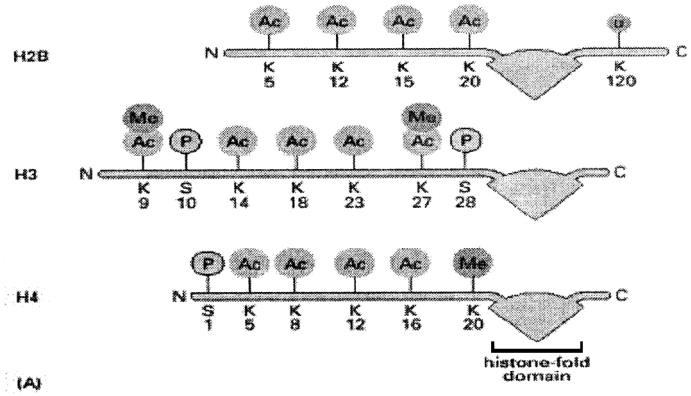


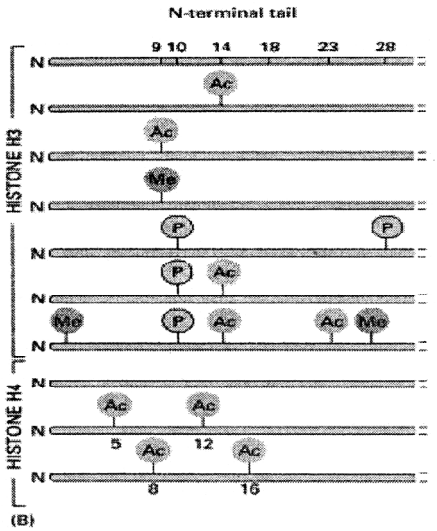
Fig. 1.7 A cyclic mechanism for nucleosome disruption and re-formation. According to this model, different chromatin remodeling complexes disrupt and re-form nucleosomes, although, in principle, the same complex might catalyze both reactions. The DNA-binding proteins could function in gene expression, DNA replication, or DNA repair, and in some cases their binding could lead to the dissociation of the histone core to form nucleosome-free regions of DNA, (Adapted from A. Travers, Cell 96:311-314, 1999.)

Covalent modification of the histone tails can profoundly affect chromatin

The N-terminal tails of each of the four core histones are highly conserved in their sequence, and perform crucial functions in regulating chromatin structure.

Each tail is subject to several types of covalent modifications, including acetyJation of lysine: methylation of lysines, and phosphorylation of serines (Fig. 1.8).





modification state	"meaning"
unmodified	gene silencing?
acetylated	gene expression
acetylated	histone deposition
methylated	gene-silencing/ heterochromatin
phosphorylated	mitosis/meiosis
phosphorylated/ acetylated	gene expression
higher-order combinations	?
unmodified	gene silencing?
acetylated	histone deposition
acetylated	gene expression

Fig. 1.8 Covalent modification of core histone tails

Histones are synthesized in the cytosol and then assembled into nucleosomes. Some of the modifications of histone tails occur just after their synthesis, but before their assembly. The modifications that concern us, however, take place once the nucleosome has been assembled. These nucleosome modifications are added and removed by enzymes that reside in the nucleus; for example, acetyl groups are added to the histone tails by histone acetyl transferases (HATs) and taken off by histone deacetylases (HDACs). The various modifications of the histone tails have several important consequences, histone acetylation tends to destabilize chromatin structure, perhaps in part because adding an

acetyl group removes the positive charge from the Iysine, thereby making it more difficult for histones to neutralize the charges on DNA as chromatin is compacted. However, the most profound effect of modified histone tails is their ability to attract specific proteins to a stretch of chromatin that has been appropriately modified.

The enzymes that modify (and remove modifications from) histone tails are usually multisubunit proteins, and they are tightly regulated. They are brought to a particular region of chromatin by other cues, particularly by sequence-specific DNA-binding proteins. It is likely that histone-modifying enzymes and chromatin remodeling complexes work in concert to condense and recondense stretches of chromatin; for example, evidence suggests that a particular modification of the histone tail attracts a particular type of remodeling complex. Moreover, some chromatin remodeling complexes contain histone modification enzymes as subunits, directly connecting the two processes.

1.2 Metaphase chromosome : Centromere, Kinetochore, Telomere and its maintenance

1.2.1 Centromere

The region of the chromosome that is responsible for its segregation at mitosis and meiosis is called the **centromere.** It is associated with two important features :

It contains the site at which the sister chromatids are held together prior to the separation of the individual chromosomes.

The term "centromere" historically; has been used in both the functional and structural sense to describe the feature of the chromosome responsible for its movement.

The centromere is essential for segregation, as shown by the behavior of chromosomes that have been broken. **Acentric fragment** does not become attached to the mitotic spindle. There can be only one centromere per chromosome. In some species the centromeres are "diffuse", which creates a different situation. Only discrete centromeres have been analyzed at the molecular level,

The regions flanking the centromere often are rich in satellite DNA sequences and contain a considerable amount of constitutive heterochromatin.

1.2.2 Kinetochore

Within the centromeric region, a darkly staining fibrous object of diameter or length ~400 nm can be seen. This object is called as Kinetochore. This

Kinetochore appears to be directly attached to the microtubules. The Kinetochore provides the MTOC on a chromosome.

Genetic engineering has produced plasmids of yeast that are replicated like chromosomal sequences. However, they are unstable at mitosis and meiosis, segregate erratically. Fragments of chromosomal DNA have been isolated by virtue of their ability to confer mitotic stability on these plasmids.

A CEN fragment is defined by its ability to confer stability upon such a plasmid. A *CEN* fragment derived from one chromosome can replace the centromere of another chromosome with no apparent consequence. This suggests that centromeres are interchangeable. *They are used simply td attach the chromosome to the spindle, and play no role in distinguishing one chromosome from another.*

The sequences required for centromeric function fall within a stretch of ~120 bp. The centromeric region is packaged into a nuclease-resistant structure, and it binds a single microtubule. The *S*. cerevisiae centromeric region has three types of sequence element may be distinguished in the *CEN* region.

CDE-I is a sequence of 9 bp that is conserved with minor variations at the left boundary of all centromeres.

CDE-II is a >90% A-T-rich sequence of 80-90 bp found in all centromeres; its function could depend on its length rather than exact sequence. Its base composition may cause some characteristic distortions of the DNA double helical structure.

CDE-III is an 11 bp sequence highly conserved at the right boundary of all centromeres. Sequences on either side of the element are less well conserved, and may also be needed for centromeric function.

A 240 kD complex of three proteins, called Cbf-III, binds to CDE-III. Mutations in the components of the genes coding for Cbf-III block chromosome movement at mitosis. A protein complex with motor activity connects the centromeric region of a chromosome to microtubules and contributes to movement on the mitotic spindle. The yeast S. *pombe* have the centromeres within regions of 40-100 kb that consist largely or entirely of repetitious DNA. The significance of the difference between the short centromeric regions in S. *cerevisiae* and the long regions in *S. pombe* is not clear. The common feature is that the DNA consists of noncoding sequences that are repetitive. Attempts to localize centromeric functions in *Drosophila* chromosomes suggest that they are dispersed in a large region, consisting of 200-600 kb. The large size of this type of centromere suggests that it is likely to contain several separate specialized functions, including sequences required for Kinetochore assembly, sister chromatid pairing, etc.

The primary modification comprising the constitutive heterochromatin of primate centromeres is a satellite DNA, which consists of tandem arrays of a 170 bp repeating unit. There is significant variation between individual repeats, although those at any centromere tend to be better related to one another than to members of the family in other locations. It is not clear whether the a satellite sequences themselves provide this function, or whether other sequences are embedded within the a satellite arrays.

1.2.3 Telomere

Essential feature in all chromosomes is the **telomere.** This "seals" the end. Telomere must be a special structure, because chromosome ends generated by breakage are "sticky" and tend to react with other chromosomes, whereas natural ends are stable. Two criteria in identifying a telomeric sequence :

It must lie at the end of a chromosome (or, at least, at the end of an authentic linear DNA molecule).

It must confer stability on a linear molecule.

Several telomeric sequences have been obtained from genomes of lower eukaryotes. In plant and man the construction of the telomere seems to follow a universal principle. Each telomere consists of a long series of short, tandemly repeated sequences. Table 1.1 lists the repeating units that have been identified at the ends of the linear DNA molecules. All can be written in the general form C_n $(A/T)_m$, where n > 1 and m is 1-4.

Within the telomeric region is a specific array of discontinuities, taking the form of single-strand breaks whose structure prevents them from being sealed by the ligase enzyme. They may be organized in a hairpin so that they are not recognized by nucleases.

Table 1.1 Telomeres have a common type of short tandem repeat. The repeating unit gives the sequence of one strand, in the 5'-3' direction form the telomere toward the centromere

Telomere	**Repeating unit**
Ciliate (*Tetrahymena*) macronucleus	CCCCAA
Ciliate (*Oxytricha*) macronucleus	CCCCAAAA
Trypanosoma minichromosome	CCCTA
Slime molds (*Dictyostelium*) rDNA	CCCTA
Yeast (*Saccharomyces*) chromosome	C_{2-3} A(CA)l-3
Plant (*Arabidopsis*) chromosome	C_3 TA_3
Human chromosome	C_3 TA_2

Addition of telomeric repeats to the end of the chromosome in every replication cycle could solve the problem of end replicating. The addition of repeats by de novo synthesis would counteract the loss of repeats resulting from failure to replicate up to the end of the chromosome. Extension and shortening would be in dynamic equilibrium.

The overall length of the telomere is under genetic control. If telomeres are continually being lengthened (and shortened), their exact sequence may be irrelevant. All that is required is for the end to be recognized as a suitable substrate for addition. This explains how the ciliate *telomere* functions in yeast. Extracts of *Tetrahymena* contain an enzyme, called telomerase, that uses the 3 '-OH of the G+T telomeric strand as a primer for synthesis of tandem TTGGGG repeats. Only dGTP and dTTP are needed for the activity. The telomerase is a large ribonucleoprotein. It contains a short RNA component, 159 bases long in *Tetrahymena,* 192 bases long in *Euplotes.* Each RNA includes a sequence of 15-22 bases that is identical to two repeats of the C-rich repeating sequence given in Table 1.1. This RNA provides the template for synthesizing the G-rich repeating sequence, to which it is complementary. Bases are added individually, in the correct sequence. The enzyme progresses discontinuously. The telomerase is a specialized example of a reverse transcriptase. The protein component provides the catalytic activity of reverse transcriptase, and is (presumably) confined to acting upon the RNA template provided by the nucleic acid component.

The structure of the telomere is organized as single-stranded extension of the G-T-rich strand, usually for 14-16 bases.

A model for the structure of the end proposes the existence of a "quartet" of G residues, formed by an association of one G from each repeating unit. The association between the G residues requires that two of them change the orientation of the base with regard to the sugar (from the usual anti to be usual syn configuration). Since each repeating unit has more than one G, more than one quartet could be formed if other G residues associate, in which case quartets might be stacked upon one another in a helical manner.

It is not known how the complementary (C-A-rich) strand of the telomere is assembled, but we may speculate that it could be synthesized by using the 3'-OH of a terminal G-T hairpin as a primer for DNA synthesis.

1.3 Heterochromatin and Euchromatin

1.3.1 Heterochromatin

Light-microscope studies in the 1930s distinguished between two types of chromatin in the interphase nuclei of many higher eukaryotic cells: a highly condensed form, called ***heterochromatin***, and all the rest, which is less condensed, called ***euchromatin*** which is composed of the types of chromosomal structures—30-nm fibers and looped domains. Heterochromatin, in contrast, includes additional proteins and probably represents more compact levels of organization that are just beginning to be understood. In a typical mammalian cell, approximately 10% of the genome is packaged into heterochromatin. Although present in many locations along chromosomes, it is concentrated in specific regions, including the centromeres and telomeres.

Heterochromatin is classified as :—

(i) Constitutive
(ii) Facultative

(i) Constitutive heterochromatin :

It consists of specific regions that are not expressed. They include satellite DNAs, and could play a structural role in the chromosome. Often these sequences are concentrated in specific regions, typically around the centromere,

(ii) Facultative heterochromatin :

It takes the form of entire chromosomes that are inactive in one cell lineage, although they can be expressed in other lineages. The example *parexcellence* is the mammalian X-chromosome, one copy of which (selected at random) is entirely inactive in a given female. (This compensates for the presence of two X chromosomes, compared with the one present in males.) The inactive X chromosome is perpetuated in a heterochromatic state, while the active X chromosome is part of the euchromatin. Here it is possible to see a correlation between transcriptional activity and structural organization when the *identical DNA sequences are involved in both states.*

Most DNA that is folded into heterochromatin does not contain genes. However, genes that do become packaged into heterochromatin are usually resistant to being expressed, because heterochromatin is unusually compact. Regions of heterochromatin are responsible for the proper functioning of telomeres and centromeres (which lack genes), and its formation may even help protect the genome from being overtaken by "parasitic" mobile elements of DNA. Moreover, a few genes require location in heterochromatin regions if they are to be expressed. In fact, the term *heterochromatin* (which was first defined cytologically) is likely to encompass several distinct types of chromatin structures

whose common feature is an especially high degree of organization. When agene that is normally expressed in euchromatin is experimentally relocated into a region of heterochromatin, it ceases to be expressed, and the gene is said to be *silenced.* These differences in gene expression are examples of position effects, in which the activity of a gene depends on its position along a chromosome. Many ***position effects*** exhibit an additional feature called position effect *variegation,* which result from patches of cells in which a silenced gene has become reactivated; once reactivated, the gene is inherited stably in this form in daughter cells. The study of position effect variegation has revealed two important characteristics of heterochromatin. First, heterochromatin is dynamic; it can "spread" into a region and later "retract" from it at low but observable frequencies. Second, the state of chromatin—whether heterochromatin or euchromatin—tends to be inherited from a cell to its progeny. These two features are responsible for position effect variegation, as explained in Figure 1.9.

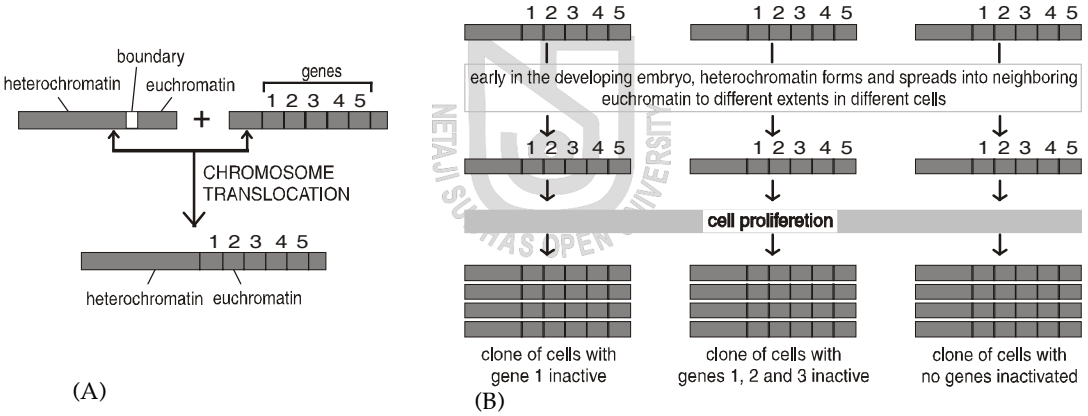


Fig. 1.9 The cause of position effect variegation in *Drosophila.* (A) Heterochromatin is normally prevented from spreading into adjacent regions of euchromatin by special boundary DNA sequences. In flies that inherit certain chromosomal rearrangements, however, this barrier is no longer present. (B) During the early development of such flies, heterochromatin can spread into neighboring chromosomal DNA, proceeding for different distances in different cells. This spreading soon stops, but the established pattern of heterochromatin is inherited, so that large clones of progeny cells are produced that have the same neighboring genes condensed into heterochromatin and thereby inactivated. Although "spreading" is used to describe the formation of new heterochromatin near previously existing heterochromatin, the term may not be wholly accurate. There is evidence that during expansion, heterochromatin can "skip over" some regions of chromatin, sparing the genes that lie within them from repressive effects. One possibility is that heterochromatin can expand across the base of some DNA loops, thus bypassing the chromatin contained in the loop

The ends of chromosomes have a special form of heterochromatin

The molecular nature of heterochromatin is probably best understood in the simple yeast *S. cerevisiae.* Mutations in any one of a set of yeast Silent information regulator (Sir) proteins prevent the silencing of genes located near

telomeres, thereby allowing these genes to be expressed. Analysis of these proteins has led to the discovery of a telomere-bound Sir protein complex that recognizes underacetylated N-terminal tails of selected histones (Fig. 1.10A). One of the proteins in this complex is a highly conserved histone deacetylase known as Sir2, which has homologs in diverse organisms, including humans, and presumably has a major role in creating a pattern of histone under acetylation unique to heterochromatin.

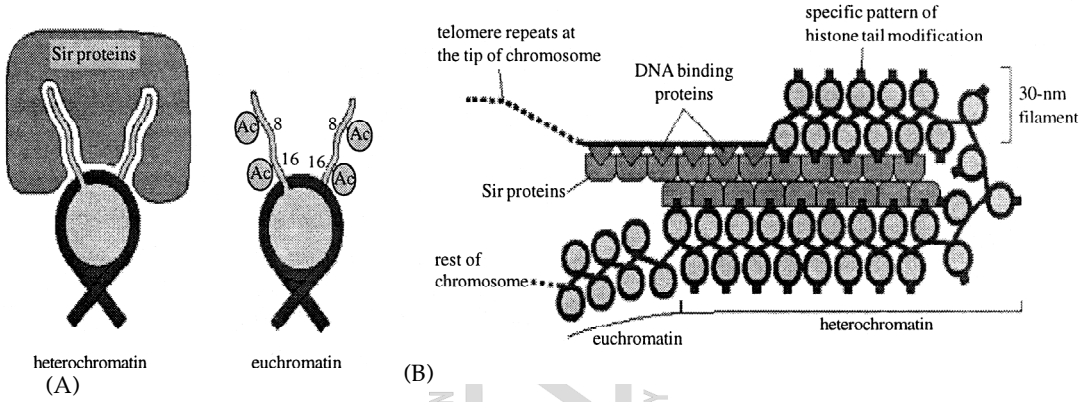


Fig. 1.10 Speculative model for the heterochromatin at the ends of yeast chromosomes. (A) Heterochromatin is generally underacetylated, and underacetylated tails of histone H4 are proposed to interact with a complex of Sir proteins, thus stabilizing the association of these proteins with nucleosomes. Although shown as fully unacetylated, the exact pattern of histone H4 tail modification required to bind to the Sir complex is not known with certainty. In some organisms, the methylation of lysine 9 of histone H3 is also a critical signal for heterochromatin formation. In euchromatin, histone tails are typically highly acctylated. Those of H4 are shown as partially acetylated but, in reality, the acetylation state varies across euchromatin. (B) Specialized DNA-binding proteins (*blue triangles*) recognize DNA sequences near the ends of chromosomes and attract the Sir proteins, one of which (Sir2) is a NAD^+ dependent histone deacetylase. This then leads to the cooperative spreading of the Sir protein complex down the chromosome. As this complex spreads, the deacetylation catalyzed by Sir2 helps create new binding sites on nucleosomes for more Sir protein complexes. A "fold back" structure of the type shown may also form

But how is the Sir2 protein delivered to the ends of chromosomes in the first place ? A DNA-binding protein that recognizes specific DNA sequences in yeast teiomeres also binds to one of the Sir proteins, causing the entire Sir protein complex to assemble on the telomeric DNA. The Sir complex then spreads along the chromosome from this site, modifying the N-terminal tails of adjacent histones to create the nucleosome-binding sites that the complex prefers. This "spreading effect" is thought to be driven by the cooperative binding of adjacent Sir protein complexes, as well as by the folding back of the chromosome on itself to promote Sir binding in nearby regions (see Fig. 1.10B). In addition, the formation of heterochromatin probably requires the action of chromatin remodeling complexes to readjust the positions of nucleosomes as they are packed together.

Whatever the precise mechanism of heterochromatin formation, it has become clear that covalent modifications of the nucleosome core histones have a critical role in this process. Of special importance in many organisms are the *histone methyl transferases,* enzymes that methylate specific lysines on histones including lysine 9 of histone H3 . This modification is "read" by heterochromatin components (including HPl in *Drosophila*) that specifically bind this modified form of histone H3 to induce the assembly of heterochromatin. It is likely that a spectrum of different histone modifications is used by the cell to distinguish heterochromatin from euchromatin.

Centromeres are also packaged into heterochromatin

In many complex organisms, including humans, each centromere seems to be embedded in a very large stretch of heterochromatin that persists throughout interphase. The stmcture and biochemical properties of this so-called *centric heterochromatin* are not well understood, but, like other forms of heterochromatin, it silences the expression of genes that are experimentally placed into it. It contains, in addition to histones (which are typically underacetylated and methylated in heterochromatin), several additional structural proteins that compact the nucleosomes into particularly dense arrangements.

Heterochromatin may provide a defense mechanism against mobile DNA elements

DNA packaged in heterochromatin often consists of large tandem arrays of short, repeated sequences that do not code for protein, as we saw above for the heterochromatin of mammalian centromeres. This suggests that some types of repeated DNA may be a signal for heterochromatin formation. This feature, called *repeat-induced gene silencing,* may be a mechanism that cells have for protecting their genomes from being overtaken by mobile genetic elements. These elements can multiply and insert themselves throughout the genome.

Once a cluster of such mobile elements has formed, the DNA that contains them would be packaged into heterochromatin to prevent their further proliferation. The same mechanism could be responsible for forming the large regions of heterochromatin that contain large numbers of tandem repeats of a simple sequence, as occurs around centromeres.

1.3.2 Euchromatin

Euchromatin is a lightly packed form of **chromatin** that is rich in gene concentration, and is often (but not always) under active **transcription.** Unlike **heterochromatin,** it is found in both **eukaryotes** and **prokaryotes.**

Structure

The structure of euchromatin is reminiscient of an unfolded set of beads along a string, where those beads represent nucleosomes. Nucleosomes consist of eight proteins known as histones, with apprxomiately 145bp of DNA wound around them; in euchromatin this wrapping is loose so that the raw DNA may be accessed. Each core histone possesses a 'tail' structure which can vary in several ways; it is thought that these variations act as "master control switches" which determine the overall arrangement of the chromatin. In particular, it is believed that the presence of-methylated lysine 4 on the histone tails acts as a general marker for euchromatin. The exact organization of the DNA within euchromatin is not known, but with the electron microscope it is possible to see loops of DNA within the euchromatin regions, each loop between 40 and 100 kb in length and predominantly in the form of the 30 nm chromatin fiber. The loops are attached to the nuclear matrix via AT-rich DNA segments called **matrix-associated regions (MARs)** or **scaffold attachment regions (SARs).**

Appearance

Euchromatin generally appears as light-colored bands when stained in **GTG banding** and observed under an **optical microscope;** in contrast to **heterochromatin,**
which stains darkly. This lighter staining is due to the less compact structure of euchromatin. It should be noted that in **prokaryotes,** euchromatin is the only form of chromatin present; this indicates that the heterochromatin structure evolved later along with the **nucleus,** possibly as a mechanism to handle increasing genome size and therefore a decrease in safety/managability.

Function

Euchromatin participates in the active transcription of DNA to mRNA products. The unfolded structure allows gene regulatory proteins and RNA polymerase complexes to bind to the DNA sequence, which can then initiate the transcription process. Not all euchromatin is necessarily transcribed, but in general that which is not is transformed into **heterochromatin** to protect the genes while they are not in use. There is therefore a direct link to how actively productive a cell is and the amount of euchromatin that can be found in its nucleus. It is thought that the cell uses transformation from euchromatin into heterochromatin as a method of controlling gene expression and replication, since such processes behave differently on densely compacted chromatin- this is known as the 'accessibility hypothesis'.

1.4 Giant chromosomes : Polytene and Lampbrush chromosomes

1.4.1 Polytene chromosomes

In dividing diploid cells the DNA synthetic phase (S phase) is regularly followed by mitosis (M phase). The alternation of Gl, S, G2, M and Gl phases is called the **cell cycle.** The process of recurrent duplication cycle without consequent mitosis is called endoreduplication. Many of the cells of certain fly larvae grow to an enormous size through multiple cycles of DNA synthesis without cell division. The resulting giant cells contain as much as several thousand times the normal DNA complement. Cells with more than the normal DNA complement are said to be **polyploid** when they contain increased numbers of standard chromosomes. In several types of secretory cells of fly larvae, however, all the homologous chromosome copies are held side by side, creating a single polytene chromosome. The fact that, in some large insect cells, polytene chromosomes can disperse to form a conventional polyploid cell demonstrates that these two chromosomal states are closely related, and that the basic structure of a polytene chromosome must be similar to that of a normal chromosome. Instances of polyploid chromosomes in Drosophila include ovary nurse cells, follicle cells surrounding oocytes, abdominal histoblasts (see **Escargot),** fat body cells, gut cells, and cells of the late prepupal **salivary gland.** During the process of polyploidization, chromosomes become multistranded.

Polytene chromosomes are large and precisely aligned side-by-side adherence of individual chromatin strands greatly elongates the chromosome axis and prevents tangling. Polyteny has been most studied in the salivary gland cells of *Drosophila* larvae, in which the DNA in each of the four *Drosophila* chromosomes has been replicated through 10 cycles without separation of the daughter chromosomes, so that **1024** (2^{10}) identical strands of chromatin are lined up side by side.

When viewed in the light microscope, distinct alternating dark **bands** and light **interbands** are visible. Each band and interband represents a set of **1024** identical DNA sequences arranged in register. About **95 %** of the DNA in polytene chromosomes is in bands, and **5%** is in interbands. The chromatin in each band appears dark, either because it is much more condensed than the chromatin in the interbands, or because it contains a higher proportion of proteins, or both. Depending on their size, individual bands are estimated to contain 3000-300,000 nucleotide pairs in a chromatin strand. The bands of *Drosophila* polytene chromosomes can be recognized by their different thickness and spacings, and each one has been given a number to generate a chromosome "map." There are

approximately 5000 bands and 5000 interbands in the complete set of *Drosophila* polytene chromosomes.

Polyploid chromosomes exhibit a banded structure that is reproducible from individual to individual. In Drosophila there are thousands of recognizable bands. In situ hybridization of cloned complementary DNA of identified genes to banded polyploid chromosomes allows the localization of genes to individual chromosome bands. Chromosomal rearrangements are easily documented by comparing the order of bands between individuals, lines or even species. The degree of rearrangement observed between species is indicative of their evolutionary distance. *Drosophila melanogaster* has four pairs of chromosomes, three pairs of autosomes and a pair of sex chromosome.

The reference system proposed by Bridges divides the limbs of salivary gland chromosomes into 102 sections called "divisions" designated by number from 1 to 102. Each of the five main limbs (X, 2L, 2R, 3L, and 3R) contains 20 divisions; the short chromosome 4 contains only two divisions. The divisions are started with a prominent band and divided further into 6 subdivisions, each designated with capital letters from A to F. Each subdivision starts with a sharp band. Thus each individual band of salivary gland chromosomes can be identified by giving the division number, subdivision, and the number of the band starting from the beginning of the subdivision. Bridges presents the following minimum numbers of bands for the salivary gland chromosomes of *Drosophila melanogaster*: 537 bands for the X chromosome, 1032 bands for the second chromosome, 1047 bands for the third chromosome, and 34 bands for the fourth chromosome, totalling a minimum of 2650 bands for the whole genome. In this initial count doublets were listed as single bands; more recent interpretations give the total number of bands as 3286 (Sorsa, 1988).

In late prepupal salivary gland chromosomes, not all DNA in each of the chromosomes is polyploid. Approximately a third of the Drosophila genome is represented by heterochromatin, and heterochromatic regions are underrepresented in polytene chromosomes as these regions do not undergo endoreduplication. For example, the *rolled* locus is found in a heterochromatic region of chromosome 2 that is considered to remain condensed (and for the most part transcnptionally inactive) throughout all or most of the cell cycle, *rolled* lies in what is considered to be alpha heterochromatin, a chromosome region that makes up the chromocenter of polytene salivary gland chromosomes. The chromocenter is thought to be made up of DNA and protein in a dense, tightly knit structure that is transcnptionally inactive. Such heterochomatic regions, which make up 30% of the *Drosophila* genome, have a much lower density of genes as compared to euchromatin. *Rolled* gene activity is unusual in that it requires the surrounding heterochromatin for gene function. Rolled gene

activity is severly impaired by bringing *rolled* close to any euchromatic position. However, these position effects can be reversed by chromosomal rearrangements that bring the *rolled* gene closer to any block of autosomal or X chromosome heterochromatin (Eberl, 1993).

Both bands and interbands in Polytene chromosomes contain genes

Since the number of bands in *Drosophila* chromosomes was once thought to be roughly equal to the number of genes in the genome, it was initially thought that each band might correspond to a single gene; however, we now know this simple idea is incorrect. There are nearly three times more genes in *Drosophila* than chromosome bands, and genes are found in both band and interband regions. Moreover, some bands contain multiple genes, and some bands seem to lack genes altogether.

It seems likely that the band-interband pattern reflects different levels of gene expression and chromatin structure along the chromosome, with genes in the less compact interbands being expressed more highly than those in the more compact bands. The remarkable appearance of fly polytene chromosomes is thought to reflect the heterogeneous nature of the chromatin compaction found along all interphase chromosomes. The remarkable appearance of fly polytene chromosomes is thought to reflect the heterogeneous nature of the chromatin compaction found along all interphase chromosomes.

Individual Polytene chromosome bands can unfold and refold as a unit

A major factor controlling gene expression in the polytene chromosomes of *Drosophila* is the insect steroid hormone ***ecdysone***, the levels of which rises and falls periodically during larval development. When ecdysone concentrations rise, they induce the expression of genes coding for proteins that the larva requires for each molt and for pupation. As the organism progresses from one developmental stage to another, distinctive ***chromosome puffs*** arise and old puffs recede as new genes become expressed and old ones are. turned off. Most puffs arise from the decondensation of a single chromosome band.

Puffing is the term that describes structural changes in polytene chromosomes. If one observes polytene chromosomes during the late prepupal stage, different bands appear to be puffed up. Puffs, then, afford **a** view of the temporal sequence of gene activation. A temporal pattern to puffing in the salivary glands of late prepupal flies is inducible by ecdysone injection and is therefore under control of the **ecdysone receptor.** A small number of genes react by puffing within minutes of exposure to ecdysone, and a much larger number (>100) react within hours. It is hypothesized that the time sequence of puffing represents a genetic hierarchy of gene activation. Early puffs are independent of protein

synthesis while late puffs require prior protein synthesis (Ashburner, 1990).

In recent years, transcription factors and chromosomal proteins have been localized to various bands. Binding of these proteins is thought to have functional significance and to reflect the activity of these proteins in gene regulation. For more information on the binding of various proteins and RNA species to bands, see **HP1/ Su(var)205, polycomb, male sex lethal 2,** and **suppressor of hairy wing.**

An example of binding of specific proteins to polytene chromosomes is found in a study of the protein CHDl (chromo-ATPase/helicase-DNA-binding domain). Proteins related to CHDl via the helicase domain have been shown to exist in large multiprotein complexes. For example SNF2/SWI2/Brm proteins are thought to participate in ATP-dependent remodeling of chromatin. Antibodies to CHDl localize this protein to extended chromatin (interbands) and regions associated with high transcriptional activity (puffs) on polytene chromosomes from salivary glands. These observations support the idea that CHDl functions to alter chromatin structure in a way that facilitates gene expression (Stokes, 1996).

Polyploidization by endoreduplication requires regulation of the **cell cycle.** What makes one region of the chromosome become polyploid while another remains underreplicated. Information about the roles of cell cycle genes in the regulation of polyploidization can be found in **cyclin E, Escargot,** and **origin recognition complex 2.**

Electron micrographs of certain puffs, called Balbiani rings, of *Chiwnomus* salivary gland polytene chromosomes show the chromatin arranged in loops, much like those observed in the amphibian lampbrush chromosomes. Each loop contains a single gene. When not expressed, the loop of DNA assumes a thickened structure, possibly a folded 30-nm fiber, but when gene expression is occurring, the loop becomes more extended. Both types of loops contain the four core histones and histone HI.

It seems likely that the default loop structure is a folded 30-nm fiber and that the histone modifying enzymes, chromatin remodeling complexes, and other proteins required for gene expression all help to convert it to a more extended form whenever a gene is expressed.

1.4.2 Lampbrush chromosomes

In 1882 Fleming first observed these chromosomes in urodele amphibian ovary. Riickert (1892) first described in great detail in shark oocytes. He coined the name “Lampbrush Chromosome” because of their brush-like appearance. These chromosomes’ occur at diplotene stage of meiotic prophase in oocytes of all animal species, in spermatocytes of several species and even in giant nucleus

of unicellular algae *Acetabularia*. These chromosomes are characterized by several lateral projections called "Lateral Loops". They are very large and best seen in salamander oocytes because of their high DNA content.

Morphology : Lampbrush chromosomes are extensible and elastic. These chromosomes with well-developed lateral loops can be stretched to about 2½ times of original length. Since, these chromosomes are found in meiotic prophase they are present in the form of bivalents in which maternal and paternal chromosomes are held together by chiasmata. The axis of each chromosome consists of a row of granules or chromomeres and from which lateral loops extend. :

(i) Centromeres : These are round, smooth, and Fuelgen positive and bear no lateral loop. In many species of urodele, centromeres are identifiable chromosome landmarks as "axial bars", formed by the amalgamation of neighbouring chromomeres, whereas in certain species of urodeles centromeres are not flanked by axial bars and are difficult to be identified, flank them. In such urodele species partner centromere do not fuse whereas urodele having axial bars centric fusion occurs.

(ii) Telomeres : Ends of Lampbrush chromosomes are occupied by distinctive telomeres consisting of a small fuelgen positive part closely applied to the surface of a smooth round Fuelgen negative part. Fuelgen negative material can be digested by proteolytic enzyme. Like centromere, Telomere do not possess lateral loop. They are of different sizes, large in T. c. *cristatus* and small in T. c. *karelini.* In some urodele fusion b/w telomeres are common in T. c. *aristatus* while in T. c. *karetini* fusions are rare.

(iii) Lateral loops : Loops are always symmetrical. Each chromosome having two of them, one for each chromatid. Loops can be distinguished by size, thickness and by several other morphological characteristics. Each loop appears at a constant position in the chromosome and there are about 10,000 loops /chromosome set. Each loop has axis formed by a single DNA molecule. About 5-10% DNA is present in the lateral loops the rest is condensed in chromomeres of chromosome axis, which is transcriptionally inactive.

Types of loops :

1. **Normal loop :** Most loops can form one pattern termed as normal loop while other loops are distinguished by their matrix deposition.
2. **Granulous loop :** They are so called because they accumulate granule at the distal end of the fine fibers projection from other classes of lateral loop differ from granulous loop in material accumulation.

(a) In some cases matrix plastering the axis but leaving the tips of the projecting fibers visible.

(b) In other cases, matrix fusion is irregular over the loop length. Such loops have uneven outlines.

3. Lumpy Loop : Situated on either side of centromere much degree of matrix fusion, usually so great that the loop pattern is wholly observed. Sometimes sister lumpy loop may fuse together so that instead of a pair of loops a single amorphous body is present.

4. Giant Loops : They are much larger than lumpy are matrix that they accumulate is exceedingly heterogeneous in texture. Each loop has its own developmental sequence of extension and regression. For example, giant granular loops are already of full size in very young oocytes and remain the same throughout oocyte development. The giant fusing loops are small in small oocytes and regress only just before ovulation. The granulous loops are largest in young oocytes is not regress early.

Unineme theory and C. value paradox :

The results of most of the earlier studied have revealed that each loop has just 1 DNA molecule as a major finding because it showed a single thread of DNA runs through each chromatid and lead to the elaboration of unineme model concept of chromosome structure.

A matrix covers each loop that consists of RNA transcripts with RNA binding proteins attached to them. In general ribonucleoprotein matrix is asymmetrical being thickness at one end of the loop than at the others. RNA synthesis starts at the thinner end and progresses to the thicker end. Many of the loops correspond to a single transcriptional unit while the other loop contains several units of transcription. Some t unite on lampbrush chromosome are extremely long i.e. over lOOum in length i.e. (lu-m of DNA = 3000 bases) why they are so enormous? Even more puzzling is that the length of loop increases with C. value as a result Salamander has 10 times longer transcription units than those of a frog although both code for a similar gene product. We have no answer to this paradox but it may be possibly connected with inefficient termination of transcription in oocytes.

The results of *in situ* hybridization studies have revealed that long Lampbrush transcripts are due to failure of termination i.e. transcription eventually stops where the next t. unit is reached. However the function of these long transcripts remain unexplainable, though the majority of them are degraded in the nucleus but presumably some RNA has some role to play in preparing an oocyte for the journey that an egg undertakes after fertilization i.e. development of a new organism.

Anyway, at the cellular level of analysis both Lampbrush chromosome and

p.c. provide remarkably favorable opportunities to study the mechanisms responsible for gene ordered synthesis. Most of the recent exciting advances in our understanding of the nature and mode of action of the genetic material have come from the genetic studies on microorganisms. Cytologists have contributed rather little to this advances.

I.5 Suggested questions

1. Describe the structure of nucleosome along with diagrams.
2. State and explain the ways of chrornatin remodeling.
3. Sate the different types of covalent modifications in histone tails and its significance.
4. Explain the significance of centormeric sequence in chromosomal segregation.
5. How is the length of telomere maintained in eukaryotic systems?
6. Explain "position effect variegation" with example.
7. How is heterochromatinization brought about?
8. Elucidate polytene chromosome structural organization.
9. Explain chromosomal puffs.
10. State the morphology of lampbrush chromosomes.
11. Validate unineme theory by lampbrush chromosome structure.

Unit 2 Sex Chromosomes, Sex Determination and Dosage Compensation

Structure

2.1 Introduction

In multicellular organisms sex is determined by many different mechanisms, which vary greatly. Of the various mechanisms of sex determination known till date, sex-chromosomal method of determination is perhaps the best understood and intriguing. Here, sex of an individual is determined by the presence or absence of its species-specific sex chromosomes. In this system of sex determination there are defined set of autosomes and well-defined pair of *allosomes* (sex chromosomes). The allosomes may be of one kind (e.g. in C. *elegans,* Grasshopper etc. has only X chromosome; thus two sexes are determined by either XX or XO) or of two different kinds (e.g. *Drosophila* has both X and Y chromosomes and sexes are determined by XX or XY).

The paradox for such mechanism of sex determination is in the fact that either of the two sexes have different sex chromosomal constitution, leading to differential allosomal gene dosages. In many organisms there are two X chromosomes in female and one X in male. Therefore, it is essential to make a balance between the products of the genes of two X chromosomes and the products of one X chromosome. The mechanism by which the balance between two dosages and one dose is maintained is known as dosage compensation. This is done either by suppressing the activities of the genes of one of the two X chromosomes of the female (inactivation of one of the female X chromosome) or by hyperactivation of the male X chromosome. This would thus require dosage compensation to negate the genie imbalance for the sex chromosomes. Although sex determination pathway and dosage compensation are different pathways they may have few steps in common but must not be considered to be same under any circumstances.

2.2 Sex determination and dosage compensation in *Caenorhabditis elegans*

Caenorhabditis elegans has two sexes: hermaphrodites and males. Hermaphrodites are essentially female animals that produce sperm during larval development and oocytes during adulthood. Hence, hermaphrodites are capable of self-fertilization, as well as cross-fertilization by males. Although some adult structures such as the pharynx are similar in males and hermaphrodites, most tissues and many aspects of behavior are different.

The pathway is not as linear and that several loops and branches in the pathway play important roles in specifying sexual development.

2.2.1 Control of xol-1 by the X:A Ratio

The primary signal for sex determination is the ratio of X chromosomes to sets of autosomes, which causes XX animals to become hermaphrodites and XO animals to become males. Early in development this ratio regulates the activity of *xol-l* (Fig.2.1), a key developmental switch gene that controls both sex determination and dosage compensation, *xol-1* encodes a novel protein, and during early embryogenesis, high levels of XOL-1 protein activity promote male development and low levels promote hermaphrodite development. The male specifying *xol-l* transcript is not needed after the end of gastrulation.

The early time at which *xol-1* acts strongly suggests that it is a direct target of the X : A signal. This signal must involve elements on the X and elements autosomes that are compared.

The X chromosome signal is polygenic, and that the combined action of these X signal element is required to inhibit *xol-1* activity in hermaphrodites. At least four different regions, regions 1-4, of the X contain signal elements, and two of these elements have been identified molecularly: *sex-land fox-1.* Increasing

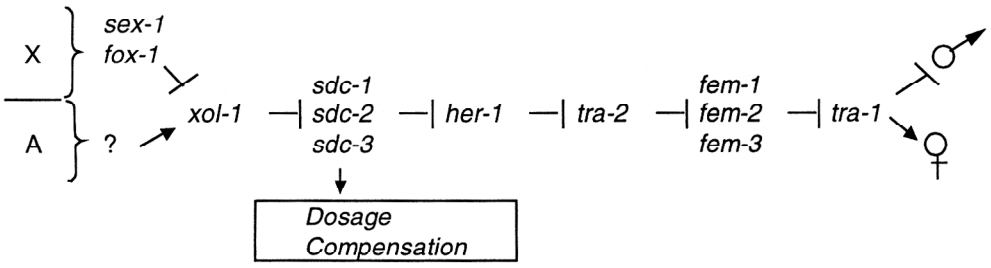


Fig. 2.1 The basic sex determination and dosage compensation pathways in *C. elegans*

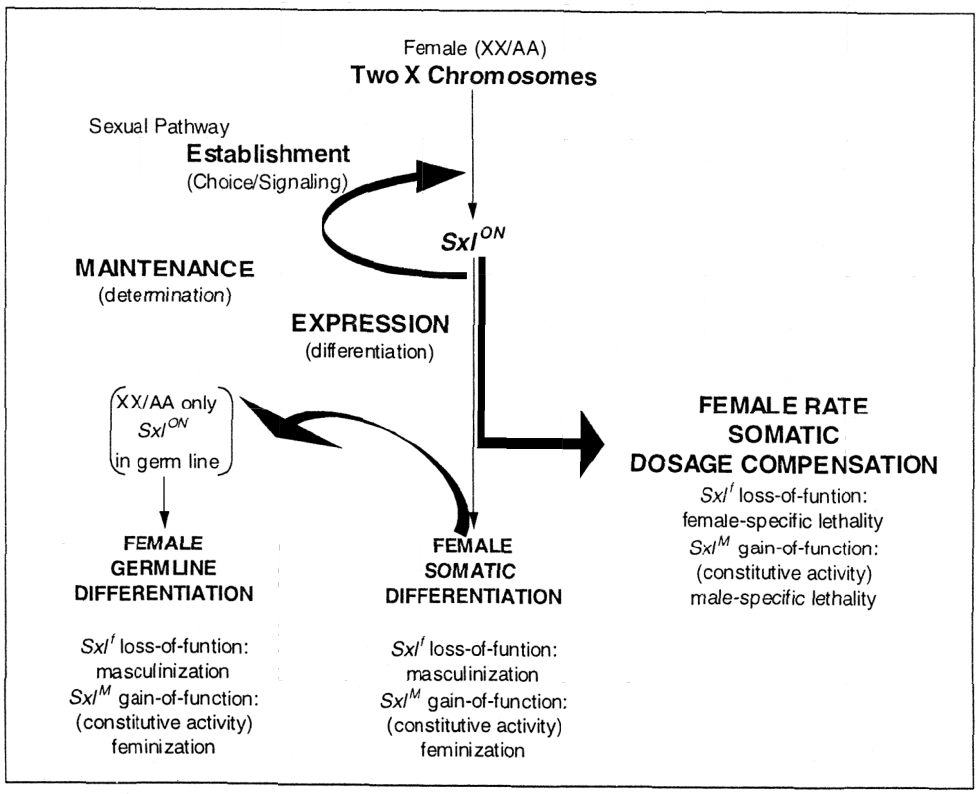


Fig. 2.2 The female-specific developmental switch gene, *Sex-lethal*, counts X chromosomes early in development to establish the choice between male and female alternative pathways of development at the cellular level. Through an autoregulatory feedback loop, *Sxl* subsequently maintains this choice throughout development, and ultimate, directs sexually dimorphic aspects of differentiation through its effects on different sets of subordinate genes downstream. Because one of these sets controls the vital process of X chromosome dosage compensation, misregulation of *Sxl* caused by upsets in the X chromosome counting process is lethal to one sex or the other. This lethality obscures potential effects on sexual phenotype. Loss-of-function {f-type) mutations in this gene are deleterious to chromosomal females (XX), while gain-of-function mutations (M-type) lead to constitutive expression and arc deleterious to chromosomal males (XY). Control ofS.t/in the germ line requires sex-specific input from the soma

the dose of these elements in XO animals represses *xol-1,* promotes hermaphrodite development and causes death because dosage compensation is activated. Decreasing their dose in XX animals activates *xol-1,* promotes male development, and causes death due to failure to initiate dosage compensation. To date, no autosomai signal elements have been identified. Evidence indicates that *sex-1* regulates the transcription *of xol-1.*

In contrast to *sex-1, fox-1* and region 2 act postranscriptionally to regulate *xol-1* expression. The *fox-1* gene encodes a protein with ribonuclear protein (RNP)

motifs, suggesting that it might bind the *xol-1* RNA. It is possible that the OF-1 protein regulates *xol-1* alternative splicing, or it might govern another aspect of *xol-1* mRNA metabolism.

Combinatorial effect of these regulatory mechanisms allows the worm to discriminate accurately between small differences in the X : A.

2.2.2 Control of the *sdc* genes by *xol-1*

Three genes are required in XX animals to promote both hermaphroditic development and dosage compensation—*sdc-], sdc-2* and *sdc-3.* The primary means by which XOL-1 transmits the X:A signal appears to be by negative regulation of *sdc-2,* as *sdc-2* is not expressed in wild-type XO embryos, but is expressed in *xol-1* XO embryos.

Null mutations in *sdc-2* and *sdc-3* have no effect on XO animals but cause complete reversal of sexual fate n XX animals; null mutations in *sdc-1* cause only a partial reversal of sexual fate. The *sdc* genes control XX hermaphrodite development by regulating the expression of the downstream sex-determining gene, *her-1*, a gene required for male development.

SDC-2 and SDC-3 might act in a complex to directly repress *her-1* transcription. This is supported by the finding that SDC-2, is highly charged and contains coiled-coil motifs, is targeted to transgenic copies of the *her-1* promoter. Moreover, this localization is blocked by specific *sdc-3* mutations, called *sdc-3 (Tra* alleles. SDC-3 is a novel protein that contains two functional domains. The first is a zinc finger motif that is required for dosage compensation but not for sex determination. The second resembles a **myosin** ATPase domain, and is necessary for sex determination but not for dosage compensation. The *sdc-3 (Tra)* mutations after this letter domain. The role of SDC-1 in regulating sexual development is less clear. SDC-1 has seven zinc fingers and resembles TFIIIA.

2.2.1 Somatic sex determination

Regulation of TRA-2A by HER-1

her-1 is required for male development, as mutations in *her-1* cause XO animals to develop as hermaphrodites but do not affect dosage compensation. The *her-1* gene is predicted to encode a novel protein with an ami no-terminal signal sequence and protein cleavage and glycosylation sites, suggesting that HER-1 is a secreted protein.

HER-1 promotes male development by repressing the activity of *tra-2.* Major transcript of this gene, *tra-2,* encodes a transmembrane protein, a direct interaction between secreted HER-1 and TRA-2A.

Regulation of sexual fate by *tra-2*

TRA-2A might be processed to release its intracellular domain, TRA-2ic. Production of TRA-2ic might occur by the action of TRA-3, a member of the calpain protease family. The *tra-3* gene is necessary for hermaphroditic development. TRA-3 can proteolytically cleave TRA-2A to release TRA-2ic in insect cells. How HER-1 inhibits TRA-2A activity is unclear. One possibility is that HER-1 inhibits production of TRA-2ic by TRA-3. *tra-2* activity is controlled not only at the level of protein processing or protein interaction, but also at the translational level by two elements called *tra Gli* elements (TGEs) which are located in the 3' untranslated region (3'UTR) of the *tra-2* message. Epistasis test show that *tra-2* promotes hermaphrodite development by inhibiting the activity of three genes, *fem-1, fem-2* and *fem-3.* TRA-2A does not transcription ally regulate these genes, as they are all expressed at high levels in both sexes. Instead it appears that TRA-2A or TRA-2ic inhibits FEM activities by protein-protein interaction. FEM-2 can bind FEM-3, so they might interact to promote male development. In addition, TRA-2A and TRA-2ic can bind FEM-3, which they might inactivate to allow hermaphrodite development.

Regulation of TRA-1A activity by the FEM proteins

The final gene in the sex-determination pathway is *tra-1,* which acts cell autonomously to promote hermaphrodite development. Although genetic experiments indicate that the FEM proteins promote male development by inhibiting TRA-1 A activity, how they do so is a mystery. They are unlikely to act transcriptionally, as *ra-1* mRNA levels do not differ between males and hermaphrodites. The phosphatase activity of Fem-2 is necessary for its activity, so it is possible that FEM-2 control the activity of TRA-1 A by altering its phosphorylation state. Alternatively the FEM proteins might control sexual

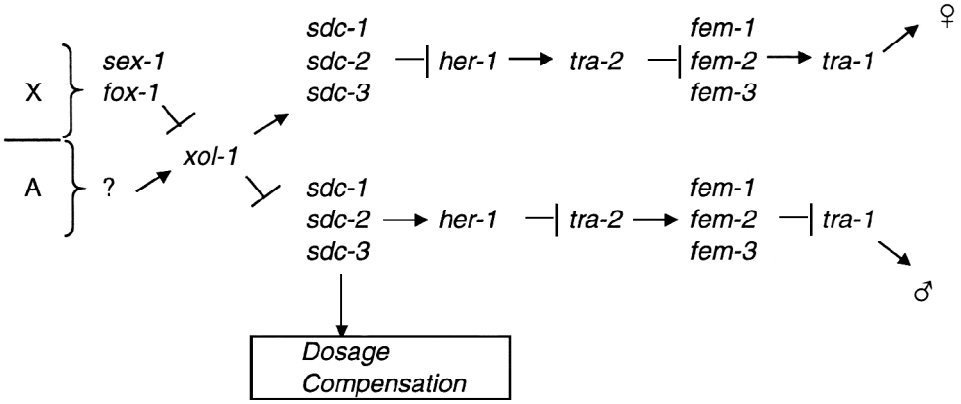


Fig. 2.3 Schematic diagram representing both male and female specific sex determination cascade in C. *elegans*

development by regulating the nuclear levels of TRA-1A. Recent analyses have revealed that hermaphrodite tissues have higher TRA-1A nuclear levels that male tissue. Thus TRA-1A transcriptional regulatory activity might be specified by nuclear versus cytoplasmic distribution of the protein. Furthermore, the FEM proteins might regulate nuclear import or export of TRA-1A, as TRA-1A is almost completely nuclear in loss-of-function *fem-1* animals. The entire sex-determination cascade can be summarized as Figure 2.2.

2.3 Sex determination and dosage compensation in *Drosophila*

In 1921 Bridges proposed that '..Sex in D. *melanogaster* is determined by a balance between the genes contained in the X chromosomes and those contained in the autosomes. It is not the simple possession of two X chromosomes that makes a female, and of one that makes a male'.

The discovery of numerator and denominator elements appears to validate the concept of sex determination of; as a ratio-measuring process. Proposed regulation cascade for *Drosophila* sex determination is schematically given below (Fig. 2.4)

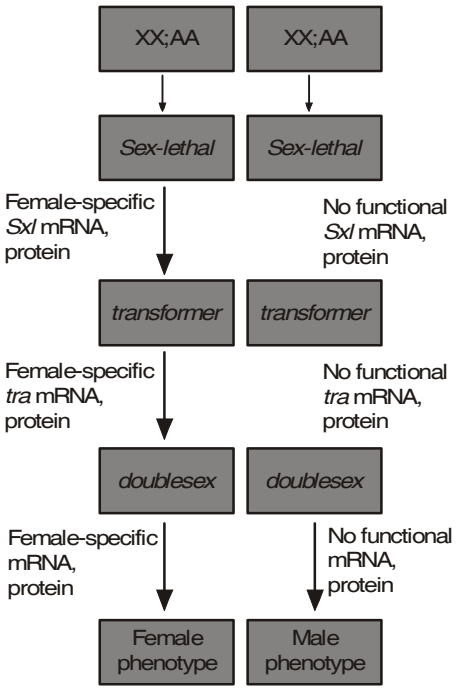


Fig. 2.4 Sex determination in *Drosophila.* This simplified scheme shows that the X-to-autosome ratio is monitored by the Sex-*lethal* gene. If this gene is active, it processes the *transformer* nRNA into a functional female-specific message. In the presence of the female-specific Transformer protein, the *doublesex* gene transcript is processed in a fern ale-specific fashion. The female-specific Doublesex protein is a transcription factor that leads to the production of the female phenotype. If the *transformer* gene does not make a female-specific product (i.e., if the *Sex-lethal* gene is not activated), the *doublesex* transcript is spliced in the male-specific manner, leading to the formation of a male-specific Doublesex protein. This is a transcription factor that generates the male phenotype

Sxl, the target of the somatic sex determination signal

Sexual differentiation and dosage compensation are a consequence of *Sxl* being turned on in diplo-X individuals and remaining off in haplo-X individuals. In somatic cells, the active (female) state is maintained by a positive feedback loop (Fig.2.2), rather than continued input from the X-linked genes that trigger the initial activation.

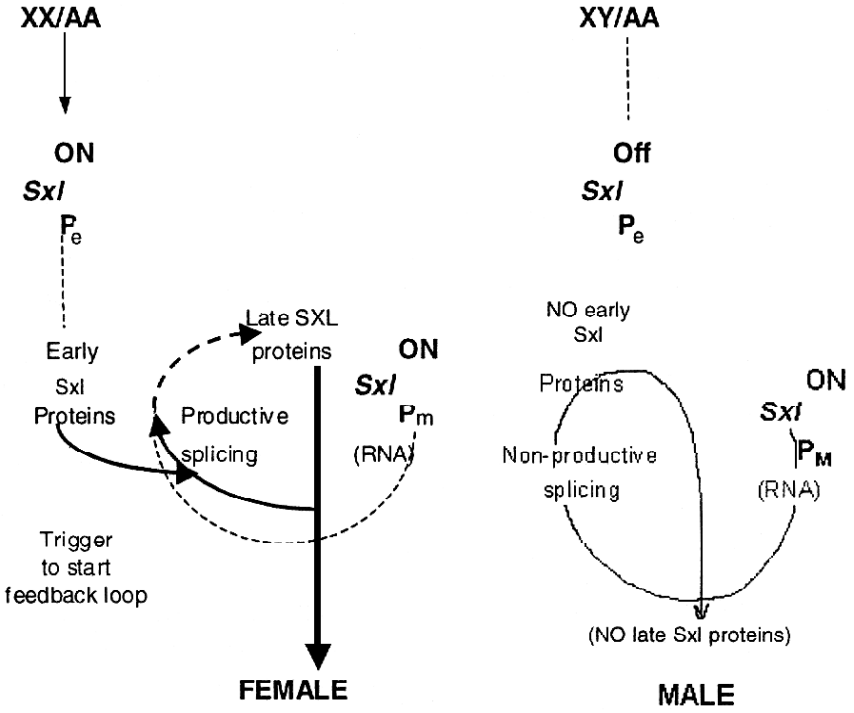


Fig. 2.5 Molecular steps in the operation of Sxl to establish and maintain a sexual pathway choice in somatic cells that is appropriate for their X chromosome dose

Figure 2.5 summarizes the current model for how, in the soma, a brief, very early effect of X chromosome dosage has a long-lasting effect on the activity of *Sxl*. The double dose of X chromosomes in females activates an *Sxl* 'establishment' promoter, $P_{e'}$ before the blastoderm stage, resulting in production of the 'early' *Sxl* proteins. In contrast, the single X chromosome dose in males leaves P_e inactive; hence males fail to produce early *Sxl* proteins.

Transition from the sexual pathway establishment (signaling) level of *Sxl* regulation to the pathway maintenance (determination) level reflects a switch in *Sxl* promoters and an attendant shift from transcriptional regulation to regulation at the level of RNA splicing. At the blastoderm stage, P_e shuts down and a 'maintenance' promoter, P_m, located 5 kb upstream, becomes active in both sexes and remains active throughout the rest of development. In contrast to

transcripts from P_e, transcripts from Pm can be spliced into mRNA encoding full-length Sxl proteins only if Sxl proteins (early or late) are already present.

The early burst of P_e expression in diplo-X individuals generates a pulse of Sxl early protein that directs the splicing of P_m-derived transcripts to eliminate a male-specific, translation-terminating exon that would otherwise block synthesis of active Sxl proteins. The Sxl proteins generated from Pm-derived transcripts then maintain productive RNA splicing through a positive auto regulatory feed back loop, and act on downstream genes to elicit female differentiation and suppress X chromosomes hyper activation.

In the absence of Sxl protein, downstream gene targets involved in both sexual differentiation and dosage compensation are expressed in a male-specific manner.

Numerator and denominator elements

The first candidate was *sisterless-a.* A zygotically acting positive regulator of Sxl. The gene was subsequently defined as 'numerator in experiments at that also revealed a second numerator elements, sis-b. the behavior of these two genes is characterized by reciprocal, zygotic dose-dependent *Sxl*-based, sex specific lethality.

One additional numerator element, the apir-rule segmentation gene runt, the first known denominator element, the semi lethal behavioral mutant *deadpan*. As a denominator element, *dpn* shows reciprocal, zygotic dose-dependent, *Sxl*-based, sex specific lethality that is the inverse of that for numerator elements; duplications, rather than deletions of this autosomal gene kill females.

Maternal involvement in progeny sex determination

Mother plays an important role in, sex determination by building into the egg the biochemical machinery the embryo needs to count its X chromosomes and decide whether to activates *Sxl*. Maternal daughterless da is a positive regulator that is necessary but not sufficient for Sxl activation. Without da activity cannot activate *Sxl* after fertilization, regardless of their X chromosome dosage. All progeny develop as males.

Analysis of the cDNA from *Sxl* mRNA shows that the *Sxl* mRNA of males differs from the *Sxl* mRNA of females. This is the result of differential RNA processing. Moreover, the Sxl protein appears to bind to its own mRNA precursor to splice it in the female manner. Since males do not have any available Sxl protein, their new 5x7 transcripts are processed in the male manner. The male *Sxl* mRNA is non functional. While the female-specific *Sxl* transcript contains a translation termination codon (UGA) after amino acid 48. The differential RNA processing that puts this termination codon into the male-specific mRNA. In

males, the nuclear transcript is spliced in a manner that yields three exons, and the termination codon is within the central exon. In females, RNA processing yields only two exons, and the male-specific central exon is now spliced out as a large intron. Thus, the female-specific mRNA lacks the termination codon. The protein made by the female-specific *Sxl* transcript can be predicted from its nucleotide sequence. This protein would contain two regions that are important for binding to RNA. Bell and colleagues (1988) have proposed that there are tow targets for the RNA-binding protein encoded by *Sxl.* One of these targets is the pre-mRNA of *Sxl* itself. This would be the mechanism that would maintain the female state of the pathway after the initial activating event had passed. The second target of the female-specific Sxl protein would be the pre-mRNA of the next gene on the pathway, *transformer.*

The transformer genes

The *Sxl* gene regulates somatic sex determination by controlling the processing of the *transformer* gene transcript. The *transformer* gene (*tra*) is alternatively spliced in males and females. There is a female-specific mRNA and also a nonspecific mRNA that is found in both females and males. The non-specific tra mRNA contains a termination codon early in the message, making the protein non-functional. The second exon of the non-specific mRNA has the termination codon. This exon is not utilized in the female-specific message. The female-specific protein form the *Sxl* gene activates a female-specific 3' splice site in the transformer pre-mRNA, causing it to be processed in a way that splices out the second exon. To do this, the *Sxl* protein blocks the binding of splicing factor U2AF to the nonspecific splice site by specifically binding to the polypyrimidine tract adjacent to it. This causes U2AF to bind to the lower-affinity (female-specific) 3' splice site and generate a female-specific mRNA. The protein encoded by this message is critical in female sex determination.

The female-specific tra product acts in concert with the *transformer-2 (tra2)* gene to help genrate the female phenotype. The *tra2* gene is eonstitutively active and makes the same protein product in both males and females. This Tra2 protein, like that of the female-specific *Sxl* protein, contains and RNA-binding domain. It is proposed tha the *tra2* gene can bind to the transcript of the *doublesex* gene, but only in the presence of the female-specific Tra protein.

Doublesex : the switch gene of sex determination

The *doubelsex* gene is active in both males and females, but its primary transcript is processed in a sex-specific manner. Male and female transcripts are identical through the first three exons. The 3'exons differ markedly. What is an exon for the female-specific transcripts is part of the untraslated 3'end of the male-specific message.

The alternative RNA processing appears to be the result of the *transformer* genes. The Tra2 and female-specific Tral proteins bind specifically to a DNA sequence adjacent to the female-specific 3′splice site of the *dsx* pre-mRNA, and they recruit nonspecific splicing factors to this site. If *tra* is not produced, the *doublesex* transcript is spliced in the male-specific manner. The downstream 3' splice site is used, and a male-specific transcript is made. This encodes an active protein that inhabits female traits and promotes male traits. The Transformer proteins bid to sequences within the female-specific exon and activate the female-specific 3′ splice site. This activation of an otherwise unused female-specific 3′ splice site produces an mRNA encoding a female-specific protein that activates female-specific genes (such as those of the yolk proteins) and inhibits male development.

The functions of the Doublesex proteins can be seen in the formation of the *Drosophila* genitalia.

Target genes for the sex determination cascade

Numerous proteins in *Drosophila* are present in one sex and not the other. In females, these include yolk proteins and eggshell (chorion) proteins. In males,

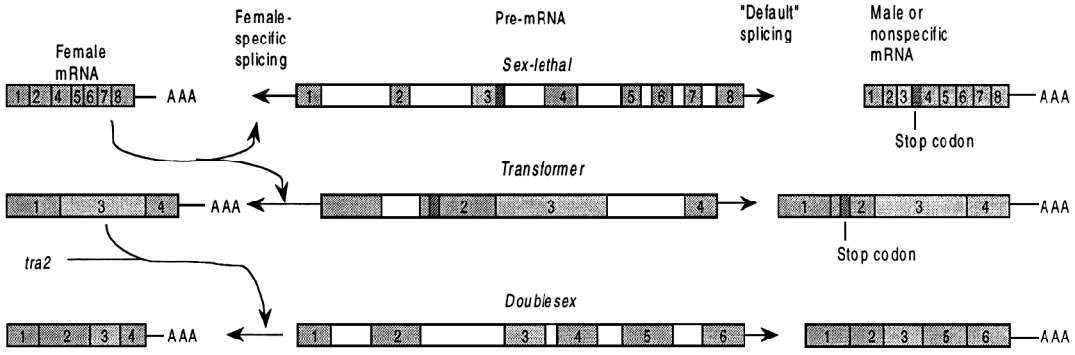


Fig. 2.6 The pattern of sex-specific RNA splicing in three major *Drosophila* sex-determining genes. The pre-mRNAs are located in the center of the diagram and are identical in both male and female nuclei. In each case, the female-specific transcript is shown at the left, while the default transcript (whether male or non-specific) is shown to the right. Exons are numbered, and the positions of the termination codons and poly(A) sites are marked. (After Baker, 1989)

the sex combs of the legs are sex -specific structure. Both the male and female *doublesex* transcripts bind to three sites within the 127-base-pair enhancer of the *yolk* protein genes. Their binding and mutagenesis studies demonstrate that the male-specific Doublesex product inhibits transcription by its binding to these sites, whereas the female-specific Doublesex protein activates gene transcription from the same sites.

2.4 Genetic regulation of sex determination and gonadal differentiation in humans

It is clear that only a small region of the Y-chromosome is endowed with the gene(s) for sex determination. The mystery of the sex reversed (Sxr) male mice having XX-chromosome constitution was resolved unambiguously by Singh and Jones (1982) who showed that one of the 2 X-chromosomes in the Sxr males carried a min]ute segment of the short arm of the Y-chromosome.

Testis determining gene on the Y-chromosome (TDY)

Subsequent studies led to the discovery of SRY gene (sex determining region of the Y) that coded for an HMG domain DNA binding protein. SRY is present on the short arm of human Y, and is conserved. The expression of Sry occurred from the onset of differentiation of medulla (primordial testis) in the genital ridge. Clinching evidence in favour of Sry as the male determining gene came from the sex reversal of the XX embryos to male through insertion of only a 14kb fragment of Y-chromosomal DNA having Sry. Transgenic mouse males are sterile the role of Sry as the male determining factor is confirmed. This single-exon gene codes for a protein that has a 79 amino acid long HMG (High mobility group proteins)-domain. The proteins harbouring this domain constitute a SOX (Sry-box) family of transcription factors that bind to a heptamer of nucleotides A/TAACAAT. Mutations in the HMG domain or in the upstream promoter region of Sry have been shown to result in XY pseudohermaphrodites, gonadal dysgenesis and other gonadal pathologies. The mouse Sry carries a polyglutamine stretch, which is absent in the human Sry.

Autosomal genes

XY individuals have however been reported with anomalies of gonad and urogenital system, in spite of having completely normal SRY (Sry) gene. More than 75% of the XY individuals suffering from acute dwarfism due to Campomelic dysplasia (a rate skeletal disorder) are hermaphrodite with ambiguous genitalia. This disorder is caused by mutation in a gene SOX9. Importance of SOX9 (Sox9 in mouse) in testis differentiation as well as subsequent organogenesis of the male genital system has since been established. An X-linked transcription factor, DAX1 (Xp21.2-22.2), has repressive effect on male development. The XY, SRY-positive pseudohermaphrodite individuals were found to have a duplication of DAX1 on the single X-chromosome, showing dosage-dependent effect of DAX1. Individuals suffering from WAGR (Wilms aniridia, genitourinary malformation, menl&l retardation) and Denys Drash syndromes both fail to have genital as

well as the renal development and are caused due to mutation in the WT-1 gene. Loss of function mutation in the autosomalWT-1, the Wilms1 tumour gene, also leads to gonadal dysgenesis. Similarly, mutation in SF-1 (orphan steroid factor-1 gene), has a broader effect on the mesonephric gonadal complex.

Interaction of genes in sexual differentiation

SF-1 and Sox9 bring about the activation of MIS gene, whose product leads to the regression of MD. These two transcription factors bind to the MIS promoter to induce its activity. WT-1 and GATA-4 act as cofactors of SF-1 and facilitate its binding to the MIS promoter. The X-linked, Dax-1, acts as a repressor of SF-1 in the female. Thus in the absence (or low level) of Sox9 and Sf-1, MIS (Mullerian Inhibiting Substances) is not produced and there is no regression of the Mullerian duct. Dax-1 is expressed also in males coincidentally with Sry, Sox9 and Sf-1 but its level is much lower than that in the female, the dosage of Dax-1 vis-a-vis Sry, Sox9, perhaps makes the difference between the differentiation of the male and female sex.

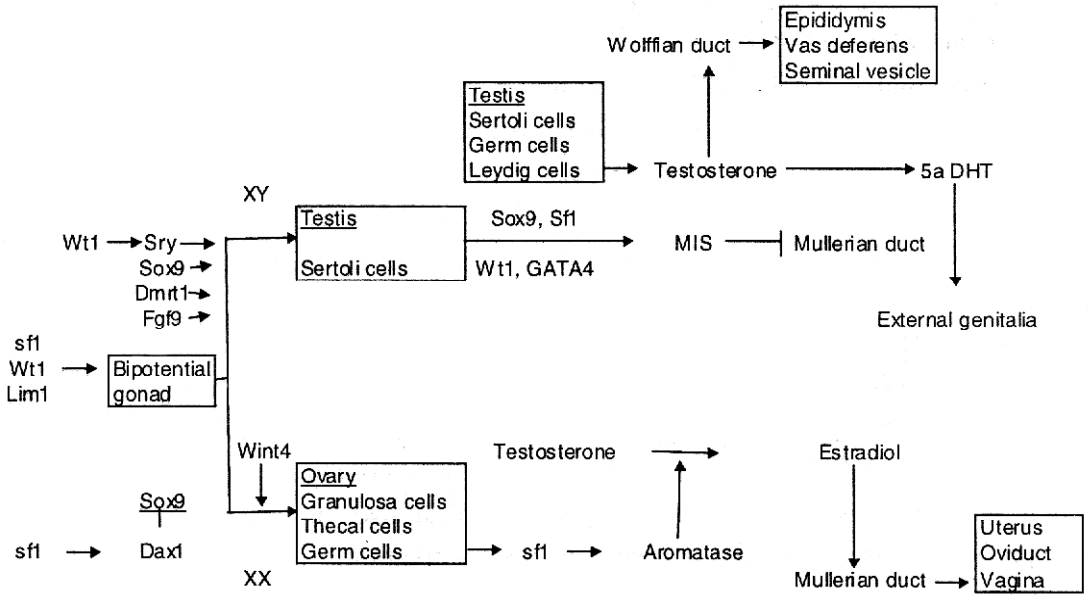


Fig. 2.7 Cartoon giving a simplified scheme of the genes participating in sex determination, and differentiation of the urogenital system in mammals

Obviously, Daxl is an "anti-male" rather than "female" gene. Its elevated level in female must be blocking not omy SF-1 induced activation of MIH but also testicular development.

More genes in the gonadal pathway

In addition a number of cases of sex reversals and ambiguous genitalia have been observed through mutations in other parts of the genome. Wnt4, coding for a cell-signaling molecule, is likely to have a positive role in female determination, unlike the "anti-testis" DAXl.

Like SF-1 artd WT-1, genes, LIM1 and LIM9, also act in the mesonephros to trigger the differentiation of GR (GR = Gonadal Ridge). Fgf9 mediates the migration of mesonephric cells in the testicular sex cords in GR of XY fetus. Therefore its absence is not detrimental to ovary in females. The XY-male specific expression of Vanin-1, a cell membrane asso

ciated G-protein, during testis differentiation may also be playing an important role in directing the mesonephric cells towards the testicular sex cords in the genital ridge.

It is obvious from the foregoing that though autosomal and X-linked genes play a role in the differentiation of the urogenital system, SRY is the key gene that switches the bipotential genital ridge towards testis determination. DNA as well as RNA binding functions are assigned to Sry (and other Sox proteins). It is intriguing that it occurs only in mammals. Functional SRY initiates a chemotactic action that leads to the migration of mesonephric cells into GR. It has also been demonstrated that the Sry expression in GR is not only temporally but also spatially restricted. The most important, role of Sry in gonadogenesis is to decisively direct the bipotential precursor cells in GR to for m the Sertoli cells instead of the follicular cells that form ovary. Centrally of SRY in testis determination is established. Nevertheless, which genes are the Sry-specific target gene(s) still remains a mystery. In a recent experiment, introduction of WT1 - promoter-driven Sox9 fusion gene succeeded in imparting male phenotype to the XX mouse, suggesting that if the Sox9 level could be raised then Sry was redundant for testis formation. The immediate conclusion from this evidence is that SOX9 may be the immediate target of SRY. It also suggests that WT-1 which is expressed upstream of Sry is involved in its activation.

Although many sex-chromosomal as well as autosomal genes play a role in mammalian (human) sex determination pathway, only few of them have' been tabulated below (Table 2.1).

Table 2.1 Genes Commonly Involved in Sex Determination in Man, Other Mammals and Lower Vertebrates

Genes	Gene product	Mutant Phenotype in man/mouse	Mutational mechanisms	Orthologue in other vertebrates
SRY	protein with an HMG domain	XY gonadal dysgenesis	loss of function	None (C. versicolor?)
WT1	Zn finger	WAGR syndrome Denys-Drash syndrome, Agenesis of gonads, kidney	Haplo-insufficiency, Dominant negative, loss of function	TSD-reptiles, birds
SF-1	orphan steroid receptor	Agenesis of urogential system	loss of function	TSD-reptiles, birds
SOX9	SRY-like HMG domain	XY-sex reversal, Campometic dysplasia,	loss of function	fish (Sox9a, b), reptiles, birds
DAX1/DSS	nuclear receptor	AHC, HGG XY SEX reversal	loss of function duplication	reptiles, birds
MIS	TGF-β gene family	persistance of Mullerian duct	loss of function	birds, TSD-reptiles
DMRT1	protein with dm-vdomain	XY-sex reversal	loss of function	fish, amophibia, reptiles, birds
Aromafase	enzyme in steroid biosynthesis	female to male sex reveral in birds & reptiles	loss of function	reptiles, birds
Wint4	signal transduction	XX, Leydig like cells	loss of function	—
Liml, Lirn9	homebox	Lim-1-loss of gonad, kidney Lim9-gonadal agenesis	loss of function	—
Fgf9	growth factor	XY abnormal gen italia	loss of function	—
Vanin2	G-protein	XY-abnormal genitalia	loss of function	—

2.5 Suggested questions

1. Describe the molecular mechanisms that regulate *Sxl-* promoter activity and its maintenance.
2. With the help of diagrams, describe the events of alternative splicing that lead to sex determination in *Drosophila.*
3. Briefly describe the cascade of events that lead to sex determination in C. *elegans.*
4. Explain the role of xol-J in C. *elegans sex* determination.
5. State and explain the effect on sex determination and dosage compensation due to mutation in the following genes:
   (a) Loss of function of *fox-l* in XX
   (b) Gain of function of *Xol* in XO
   (c) Gain of function of *Sxl* in XY
   (d) Loss of function of *da* in XX and XY
   (e) Gain of function of *Sdc* in XX
6. With the help of diagram briefly describe the initial pathway that regulates sex determination in Human.

Unit 3 Imprinting of Genes, Chromosomes and Genomes

Structure

3.1 Introduction

In humans and other mammals, several biallelic genes are known where the expression of one parental allele, either the paternal or the maternal allelebut not both, is *normally* repressed in some cells (**allelic exclusion**). In such cells the relevant gene is said to exhibit functional hemizygosity; *even although the sequences of both parental alleles are perfectly consistent with normal gene expression and may even be identical*. In some cases the allelic exclusion may be a property of select cells or tissues while in other cells of the same individual both alleles may be expressed normally. A variety of different expression mechanisms can be involved and two broad classes of mechanism are involved :

- *Allelic exclusion according to parent of origin* (*imprinting*). The choice of which of the two inherited copies is expressed is not random. This means that for some genes the allele whose expression is repressed is always the paternally inherited allele; in others it is always the maternally inherited allele.
- *Allelic exclusion independent of parent of origin*. Here the decision as to which of the two alleles is repressed is initially made randomly, but afterwards that pattern of allelic exclusion is transmitted stably to daughter cells following cell division. A variety of different mechanisms may be involved. A unique form of control is the programmed DNA rearrangements.

3.2 Genomic imprinting

Genomic imprinting is an epigenetic phenomenon, which, in most cases, is believed to occur during gametogenesis. Genomic imprinting occurs when both maternal and paternal alleles are present, but one allele will be expressed while the other remains inactive. The most prominent assumption is that this process is necessary for development and may somehow regulate growth in the embryo and neonate. Some

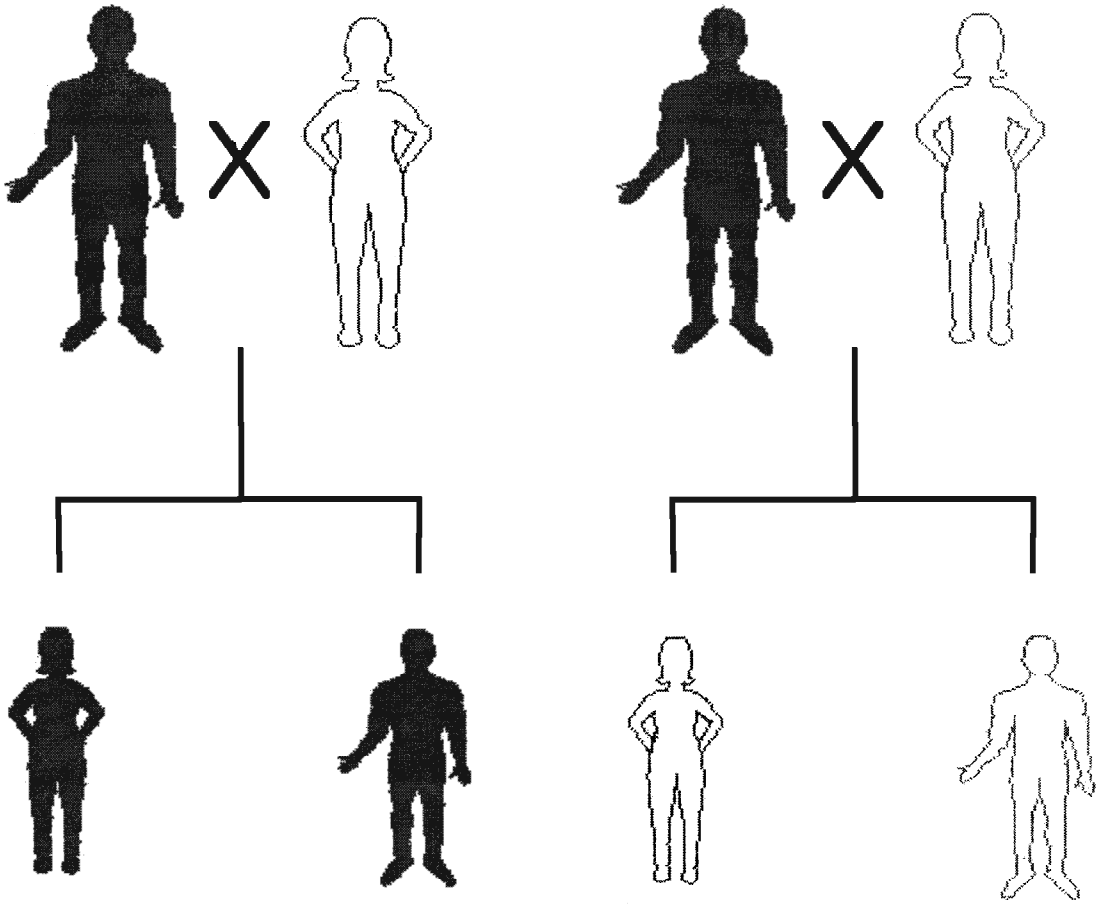


Fig. 3.1 Two examples of a hypothetical imprinted gene responsible for body color. (LEFT) In this example the pigment gene is maternally imprinted (maternal allele is inactivated). Matings between a male who possesses the allele for pigment and a female who possesses the allele for no pigment produces offspring that show only the pigmented phenotype. In this example, the mother's allele is imprinted and inactivated in the offspring. Therefore, the only actively-expressing allele is the father's pigment allele, which is not imprinted in the offspring. (RIGHT) In this example the pigment gene is paternally imprinted (paternal allele is inactivated). Matings between a male who possesses the allele for pigment and a female who possesses the allele for no pigment produces offspring that show only the pigmented phenotype. In this example, the father's allele is imprinted and inactivated in the offspring. Therefore, the only actively expressing allele is the mother's no pigment allele, which is not imprinted in the offspring. (Figure courtesy of Ross McGowan, Dept. Zoology, University of Manitoba)

mechanism must be able to distinguish between maternally and paternally inherited alleles: as chromosomes pass through the male and female germlines they must acquire some imprint to signal a difference between paternal and maternal alleles in the developing organism (Fig. 3.1).

An optimal method for gene imprinting, at least in maintaining the imprinted status, is allele-specific DNA methylation. The imprinting of several

imprinted genes has been shown to be disrupted in mutant mice that are deficient in the ***Dnmt 1*** cytosine methyltransferase gene and all imprinted genes are characterized by CG-rich regions of differential methylation. This process is carried out with the enzyme DNA methyltransf erase (DNA MTase) in mammals. DNA MTase acts on the DNA sequence 5'-CpG-3'. Some organisms (primarily higher eukaryotes) have aggregates of CpG (known as CpG islands) in their genomes. These islands are rarely methylated in animal cells. This may be due to the bound transcription factors that block DNA MTase. De novo methylation and maintenance of methylation are two distinct processes that are required for establishment and mitotic inheritance of tissue specific methylation patterns. *Dnmt1* is the major maintenance methyltransferase. ***Dnmt3a*** and ***Dnmt3b*** are essential for de novo methylation. And Sequences that are methylated are usually not active (Gold and Pedersen, 1994). Recent investigations, however, have shown that this is not always the case (Li *et al*., 1993).

It has been postulated that if a mutation was introduced to the DNA MTase gene in the embryonic stem cells of mice, the methylation of CpG would be abnormal, and gene expression would be affected. The mutation of the DNA MTase gene was caused by homologous recombination. The three genes used in this experiment were HI9, Igf2 (insulin-like growth factor) and Igf2r (Igf2 receptor).

For the H19 gene, it is the maternal allele that is expressed, while the paternal allele is silent. It should be noted that the inactive paternal allele is methylated while the maternal allele is not. It was shown that typical DNA methylation is a requirement to keep the paternal allele inactive for the H19 gene, a result that is consistent with the hypothesis.

In contrast to the H19 gene, the Igf2 gene is expressed only from a methylated paternal allele. It has now been concluded that a normal level of DNA methylation is needed for expression of the paternal Igf2 allele.

The gene Igf2r (insulin like growth factor receptor) is expressed from a methylated maternal allele. DNA methylation by DNA MTase is a requirement for the expression of the Igf2r gene.

Disruption of primary imprinting during oocyte growth leads to the modified expression of imprinted genes during embryogenesisThus far, experiments have not demonstrated how the imprinting process is regulated.lt has been proposed that further research should be attempted to discover and isolate imprinting genes . Already, some progress is being made in these areas. For instance, genomic imprinting has been implicated in cancer and shown to be involved in chromosomal deletion syndromes, such as Prader-Willi and Angelman Syndromes (Peterson and Sapienza, 1993).

3.3 Uniparental disomy and genomic imprinting

Uniparental disomy refers to the presence of two copies of a chromosome (or part of a chromosome) from one parent and none from the other. Several additional disorders resulting from uniparental disomy of single genes or multiple genes (including whole chromosomes) have been reported. A readily detectable adverse outcome of uniparental disomy is the consequence of a newly recognized phenomenon *called genetic or genomic imprinting.*

The first recognized example of such human abnormality resulting from the presence of uniparental disomy of an imprinted part of the genome was in Prader-Willi syndrome (PWS). Uniparental maternal disomy for chromosome 15

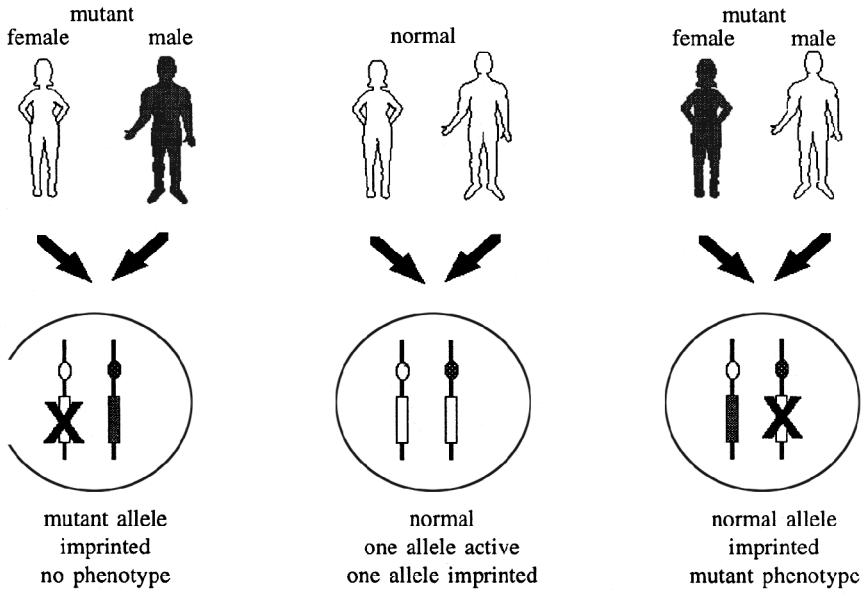


Figure 3.2 Schematic representation of the phenotypic effects of maternal imprinting of a mutant allele. Darkened body indicates individual that is mutant for the hypothetical imprinted locus. A cross is used to indicate the imprinted/inactive allele. (CENTER) Both parents are homozygous for the normal allele at the imprinted locus. Although only one allele is active (the paternal copy) in the offspring produced from these parents, it must be a normal allele and therefore all offspring will have a normal phenotype. (LEFT) The mother is homozygous mutant at the imprinted locus, and the father is normal. Since this hypothetical locus is maternally imprinted, the maternal mutant copy will be inactivated in their offspring and the paternal normal copy will be the only active allele. The offspring will be phenotypically normal, and the mutant allele will appear to be a recessive mutation. (RIGHT) The mother is homozygous normal at the imprinted locus, and the father is homozygous mutant. The maternal normal allele is imprinted and inactivated in the offspring of these parents. The only allele that is active is the mutant paternal copy. Therefore, all offspring produced from these parents will display the mutant phenotype, and the mutant allele will appear to be a dominant mutation. (Figure courtesy of Ross McGowan, Dept. Zoology, University of Manitoba)

is thought to cause Prader-Willi syndrome because there is absence of needed paternally contributed genes in the critical PWS region (del 15q 11 -q 13). The paternal contribution is hypothesized to be necessary because the homologous maternally derived genes are inactivated or imprinted (perhaps by methylation).

Interestingly, a very different disorder called Angelman syndrome also involves imprinting of the same chromosome region - only in Angelrnan syndrome the maternal contribution of the critical region is missing. The terminology used to describe the role of imprinting in these two disorders is somewhat confusing but goes as follows.

It is hypothesized that the critical genetic region which determines Prader-Willi syndrome is *maternally* imprinted (i.e. inactivated when inherited from the mother), whereas the critical region which determines Angelman syndrome is *paternally* imprinted (i.e. inactivated when inherited from the father). Both disorders result when the expected active genetic contribution from one parent is missing, either by deletion or uniparental disomy (Fig.3.2).

Interestingly, a number of human congenital tumors show evidence of genomic imprinting. For example, in cells from Wilms' tumor, loss of the maternal chromosome 11 is common. This suggests that the maternal chromosome 11 has a tumor suppresser role not present on the paternal 11. This phenomenon in relation to cancer is referred to as "loss of heterozygosity".

3.4 Suggested questions

1. How is genomic imprinting established and maintained?
2. Explain the phenomena and importance of uniparental disomy.
3. Differentiate paternal and maternal imprinting.

Unit 4 Somatic Cell Genetics

Structure

4.1 Cell fusion and hybrids-agents and mechanisms of fusion

Cultured animal cells infrequently undergo cell fusion spontaneously. The fusion rate, increases greatly in the presence of certain viruses that have a lipoprotein envelope similar to the plasma membrane of animal cells. Cell fusion is also promoted by polyethylene glycol, which causes the plasma membranes of adjacent cells to adhere to each other and to fuse. As most fused animal cells undergo cell division, the nuclei eventually fuse, producing viable cells with a single nucleus that contains chromosomes from both "parents." The fusion of two cells that are genetically different yields a hybrid cell called a **heterokaryon** (Fig. 4.1).

Because some **somatic cell** from animals can be cultured from single cells in a well-defined medium, it is possible to select for genetically distinct cultured animal cells. Genetic studies of cultured animal cells are called *somatic-cell genetics* to distinguish them from *classical genetics,* which deals with whole organisms derived from **germ cells** (sperm and eggs).

Assigning genes to chromosomes

The technique of somatic cell hybridization is extensively used in human genome mapping, but it can in principle be used in many different animal systems. The procedure uses cells growing in culture. A virus called the Sendai virus has a useful property that makes the mapping technique possible.

If suspensions of human and mouse ceils are mixed together in the presence of Sendai virus that has been inactivated by ultraviolet light, the virus can mediate fusion of the cells from the different species. The initial fusion products are described as heterokaryons because the cells contain both a human and a rodent nucleus. Eventually, heterokaryons proceed to mitosis, and the two nuclear

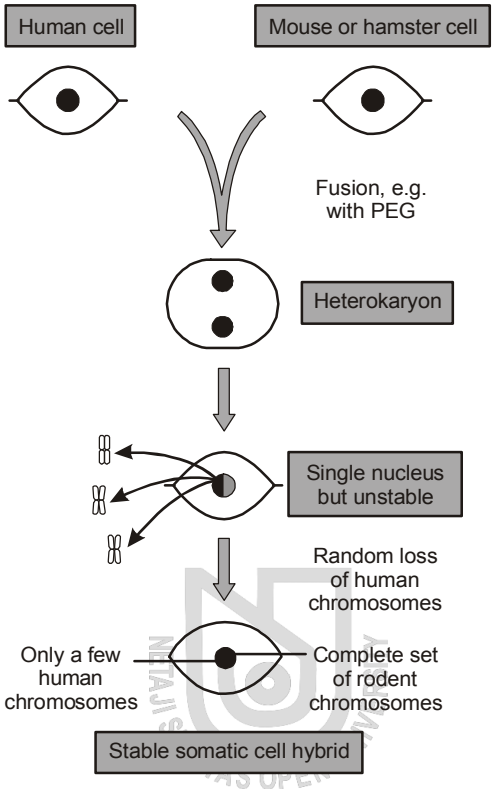


Fig. 4.1 Fusion of cells from different species can result in stable somatic cell hybrids. The example shows how stable human-rodent somatic cell hybrids can be generated following initial fusion using polyethylene glycol (PEG). For reasons that are not understood, human chromosomes are selectively lost from the initial fusion products. The loss occurs essentially at random so that eventually the stable products of a single fusion experiment will include a variety of cells with different complements of human chromosomes. They can be cloned to establish individual cell lines with a specific complement of human chromosomes. The identity of the human chromosomes can be established by PCR-based typing for chromosome-specific markers

envelopes dissolve. Because the mouse and human chromosomes are recognizably different in number and shape, the two sets in the hybrid cells can be readily distinguished. However, in the course of subsequent cell divisions, for unknown reasons the human chromosomes are gradually eliminated from the hybrid at random.

The loss of human chromosomes can be arrested in the following way to encourage the formation of a stable partial hybrid. The cells used are mutant for some biochemical function; so, if the cells are to grow, the missing function must be supplied by the other genome. This selective technique results in the maintenance of hybrid cells that have a complete set of mouse chromosomes

and a small number of human chromosomes, which vary in number and type from hybrid to hybrid but which always include the human chromosome carrying the wild-type allele defective in the mouse genome.

4.2 Heterokaryon-selecting hybrids and chromosome segregation

In cells, DNA can be made either de novo ("from scratch") or through a salvage pathway that uses molecular skeletons already available. The selective technique involves the application of a chemical, aminopterin that blocks the de novo synthetic pathway, confining DNA synthesis to the salvage pathway. Two essential salvage enzymes, thymidine kinase (TK) and hypoxanthine-guanine phosphoribosyl transferase (HGPRT), are relevant to the system, as shown in the following two reactions :

$$\text{thymine} \xrightarrow{\text{TK}} \text{thymidilic acid (a DNA precursor)}$$

$$\text{hypoxanthine} \xrightarrow{\text{HGPRT}} \text{inosinic acid (a DNA precursor)}$$

The mouse cell line to be fused is genetically unable to make TK because it is homozygous for the allele tkr, whereas the human cell line is genetically unable to make HGPRT because it is homozygous at another locus for the allele hgprT. So the genotypes of the two fusing cell lines are :

Mouse : tk^- / tk^-; $hgprt^+ / hgprt^+$

Human : tk^+ / tk^+; $hgprt^- / hgprt^-$

Because each is deficient for one enzyme, neither the mouse nor the human cells are able to make DNA individually. In. the hybrid cells, however, the tk^+ alleje complements the $hgprt^+$ allele, so the cells can make both enzymes. Therefore, DNA is synthesized and the cells can proliferate. Most human chromosomes are eliminated from the hybrid cell cultures because their loss has no effect on the cultures' ability to grow. But, to continue to grow in medium containing hypoxanthine, aminopterin, and thymidine (*HAT medium*), a hybrid culture must retain at least one of the human chromosomes that carry the tk^+ allele.

4.2.1 Selecting for the chromosome contents of hybrids

Hybrids can be selected for retention of a given human chromosome or chromosome fragments if it corrects an otherwise lethal abnormality in the rodent cell. Frequently used systems include :

- **Hybrid cells often are selected on HAT medium :** The medium most often used to select hybrid cells is called HAT *medium,* because it contains hypoxanthine (a purine), aminopterin, and thymidine. Normal cells can grow in HAT medium because even though aminopterin blocks de novo synthesis of purines and TMP, the thymidine in the media is transported into the cell and converted to TMP by TK and the hypoxanthine is transported and converted into usable purines by HGPRT. On the other hand, neither TK nor HGPRK cells can grow in HAT medium because each lacks an enzyme of the salvage pathway. However, hybrids formed by fusion of these two mutants will carry a normal *TK* gene from the HGPRK parent and a normal *HGPRT* gene from the TK parent. The hybrids thus will produce both functional salvage-pathway enzymes and grow on HAT medium. Likewise, hybrids formed by fusion of mutant cells and normal cells can grow in HAT medium. Somatic cell hybrids can be forced to retain human chromosome 17 by using thymidine kinase deficient (TK) rodent cells and growing the hybrids in *HAT* (hypoxanthine-aminopterin-thymidine) medium.
- **G418 selection** : Hybrids can be selected for the presence of a particular human chromosome segment if it has been tagged by incorporation of a neomycin resistance (neo^R) gene. The neomycin analog G418 kills nonresistant cells. NeoR is a typical example of a dominant selectable marker.

Selecting for the chromosome contents of hybrids

Hybrids can be selected for retention of a given human chromosome or chromosome fragments if it corrects an otherwise lethal abnormality in the rodent cell. Frequently used systems include :

- **HAT selection :** Somatic cell hybrids can be forced to retain human chromosome 17 by using thymidine kinase deficient (TK) rodent cells and growing the hybrids in *HAT* (hypoxanthine-aminopterin-thymidine) medium. TK cells are killed in HAT medium, but are rescued by the human TK gene on chromosome 17.
- **G418 selection :** Hybrids can be selected for the presence of a particular human chromosome segment if it has been tagged by incorporation of a neomycin resistance (neo^R) gene. The neomycin analog G418 kills nonresistant cells. Neo^R is a typical example of a dominant selectable marker.

4.2.2 Somatic cell hybrid panels can permit chromosomal localization of any human DNA sequence

The human chromosomes in somatic cell hybrids can conveniently be identified by PCR screening with sets of chromosome-specific primers. By

collecting hybrid cell lines with different human chromosome contents it is possible to generate a hybrid cell panel that can be used to map any human DNA sequence to a specific chromosome. To do this, each of the hybrid cell lines is tested for the presence of the human sequence of interest. A PCR assay can be used with primers specific for that sequence or the relevant DNA sequence can be labeled and used as a hybridization probe.

4.3 Radiation hybrids, hybrid panels and gene mapping

Subchromosomal mapping is possible using hybrid cells containing defined portions of a human chromosome

Conventional somatic cell hybrids are a relatively crude tool for physical mapping. More refined mapping is possible using hybrids that contain only part of a particular human chromosome. Translocation hybrids and deletion hybrids are made using donor human cells that have a chromosomal translocation or deletion. To be useful, the hybrids must lack the normal homolog of the chromosome of interest. Such hybrids can be used for subchromosomal mapping of a human sequence-tagged site or biochemical marker. (Fig. 4.2). They are especially useful for defining the sequences removed by microdeletions, by segregating the deletion-carrying chromosome away from its normal homolog.

Chromosome-mediated gene transfer

One of the first techniques to use this approach was **chromosome-mediated gene transfer (CMGT)**. Fragments of purified mitotic chromosomes from a donor, such as a human fibroblast, are coprecipitated with calcium phosphate on to the surface of a recipient rodent cell line in monolayer culture. Human chromosome fragments enter the recipient cells, such as mouse fibroblasts, and integrate into the chromosomes, resulting in stable transformation. As a result, hybrids can be established that retain segments of human DNA (**transgenomes**) of a size that is useful for mapping (usually in the range of 1-50 Mb). However, the transgenomes are prone to frequent rearrangements, so CMGT is more suited to functional assays of complex loci than as a mapping tool.

Irradiation fusion gene transfer

The most valuable hybrids for gene mapping are radiation hybrids. Donor cells are subjected to a lethal dose of radiation, which fragments their chromosomes. The average size of a fragment is a function of the dose of radiation. After irradiation the donor cells are fused with recipient cells of a different species. A selection system is used to pick out recipient cells that have taken

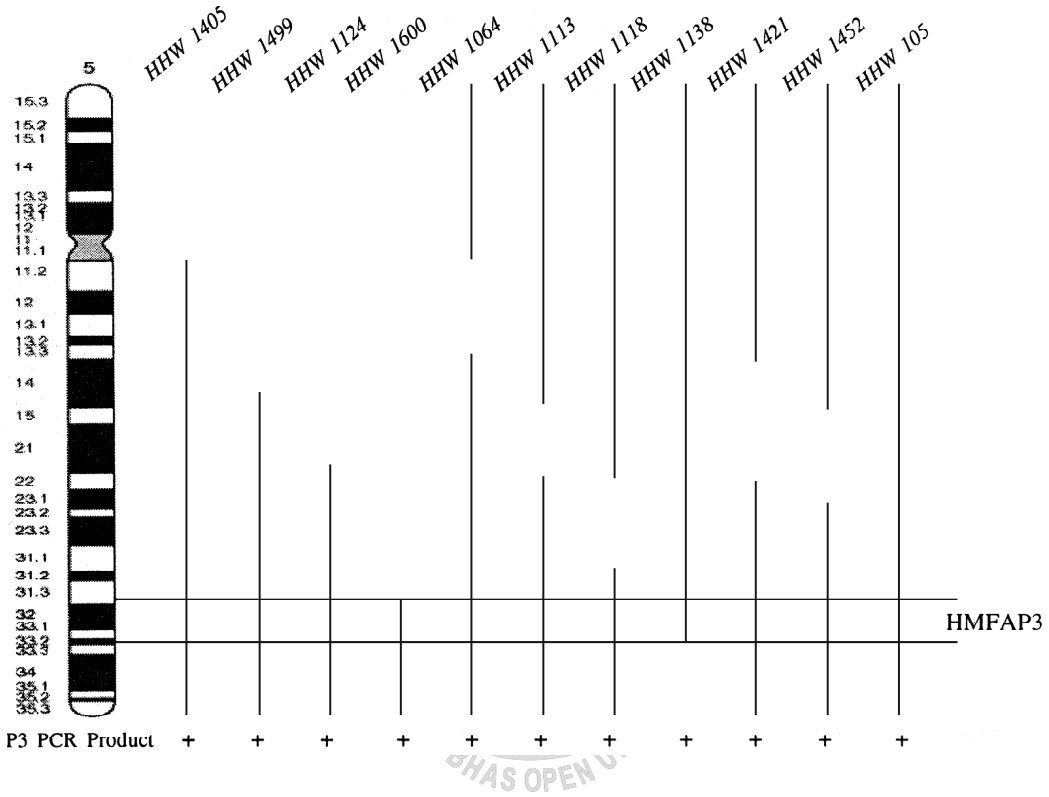


Fig. 4.2 Subchromosomal localization can be achieved by mapping against a panel of hybrid cells containing translocation or deletion chromosomes. The figure illustrates PCR-based mapping of the human microfibrillar protein MFAP3 using a panel of 5q translocation and deletion hybrids. Vertical black bars to the right indicate the extent of human chromosome 5 sequences, which are retained in the hybrids. Hybrids HHW1405, 1499, 1124 and 1600 contain translocation chromosomes with 5q breakpoints and retention of the segment distal to the breakpoint. By contrast, translocation hybrid HHW1138 retains material proximal to the 5q breakpoint. Hybrids HHW1064, 1113, 1118, 1421 and 1452 have different interstitial deletions of 5q. The solid blue vertical bar to the left indicates the inferred subchromosomal location as defined by breakpoints in hybrids HHW1600 and HHW1138 (blue horizontal lines near bottom). Reproduced from **Abrams et al. (1995)** Genomics, **26**, pp. 47-54, with permission from Academic Press, Inc

up some of the donor chromosome fragments. These cells are useful for mapping insofar as they have taken up a random set of other chromosome fragments from the donor, as well as the selected fragment. Stably incorporated donor fragments are either integrated into rodent chromosomes or are assembled into novel human minichromosomes formed around fragments containing a functional centromere. When a set of DNA markers from the human chromosome is assayed in a panel of such radiation hybrids, the patterns of cross-reactivity can be used to construct a map. (Fig. 4.3).

The principle is very similar to meiotic linkage analysis : the nearer together two DNA sequences are on a chromosome, the lower the probability that they

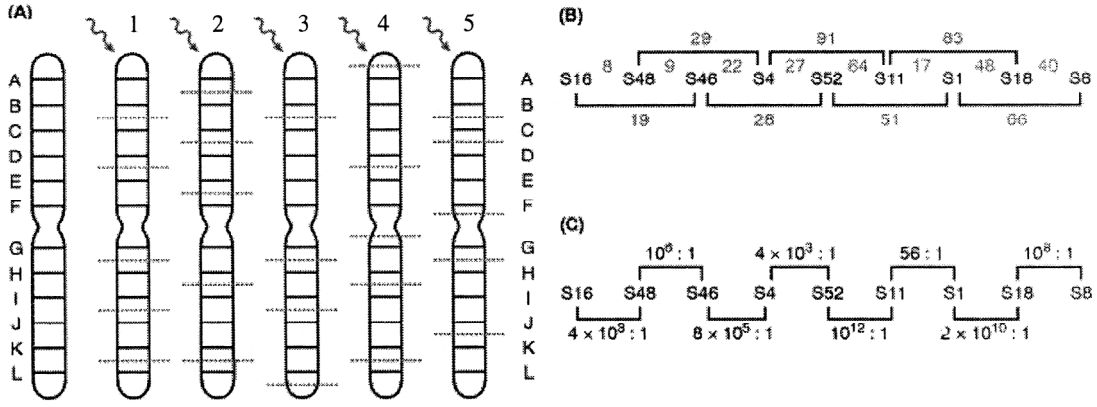


Fig. 4.3 Constructing radiation hybrid maps. (**A**) Breakpoints occur randomly. Five possible examples of breakpoints (dashed blue lines) on the same type of chromosome are shown. Markers close together will tend to occur on the same fragment, e.g. A and B in all cases other than example 2. Thus, if a radiation hybrid contains marker A it will frequently also contain marker B, but rarely a distant marker such as L. (B) Ordering of markers on human 21q. The order of markers *D21S16-D21S8* as inferred by *Cox et ai* (1990) from radiation hybrid mapping is shown. Figures on the top panel refer to distances between markers in centiRays8000. For example, the SO 6-S48 interval is 8 cR8000: at a radiation dose of 8000 rad, there is 8% frequency of breakage between them, and so a 92% chance they will occur together on one fragment. (C) Odds ratios refer to the likelihood of the indicated order for pairs of markers compared with that with the markers inverted. For example, the calculated likelihood for the order S16-S48-S46-S4 is 106 times greater than for the order S16-S46-S48-S4

will be separated by the chance occurrence of a breakpoint between them. The frequency of breakage between two markers can be defined by a value, analogous to the recombination frequency in meiotic mapping. The value,varies from 0 (the two markers are never separated) to 1.0 (the two markers are always broken apart). As in meiotic mapping, the value underestimates the distance between markers that are far apart on the same chromosome, in this case because a cell can take up two markers on separate fragments. A more accurate estimate is provided by a mapping function, D = -In (1 - value), which is analogous to the Haldane mapping function used in meiotic linkage analysis. D is measured in centiRays (**cR**). *D* is dependent on the dosage of radiation, so it is referenced against the number of rads. For example, a distance of 1 cR_{8ooo} between two markers represents a 1% frequency of breakage between them after exposure to 8000 rad of X-rays.

Radiation hybrids derived from monochromosomal hybrid donor cells have been superseded by whole-genome radiation hybrids where the donor is an irradiated normal human diploid cell. The first such panel consisted of 199 hybrids made by fusing an irradiated 46,XY human fibroblast cell line to TK

hamster cells (Walter *et al.*,>1994), Gyapay et al. (1996) used 404 microsatellite markers of known location to show that this hybrid panel could generate accurate maps, and then used it to map 374 unmapped ESTs. A subset of 93 of the hybrids has been made widely available as the Genebridge 4 panel. The 93 hybrids average 32% retention of any particular human sequence, with an average fragment size of 25 Mb. Laboratories can map any unknown STS by scoring the 93 Genebridge hybrids and comparing the pattern with patterns of previously mapped markers held on a central server.

This has turned into an extremely powerful and convenient tool for physically mapping any STS or EST. A second human-hamster panel, Stanford G3, was made using a higher dose of radiation, so that the average human fragment size is smaller. The 83 hybrids in G3 average 16% retention of the human genome, with an average fragment size of 2.4 Mb. Thus G3 can be used for finer mapping. The impressive results of large-scale use of these panels can be accessed at http://www.ncbi.nlm.nih.gov/genemap98/.

4.4 Suggested questions

1. What is heterokaryon? What are the methods of heterokaryon selection?
2. Explain the principle of mapping by radiation hybrids.
3. How can genes be assigned to specific regions on chromosomes by low-resolution mapping (radiation hybrids)?

4.5 Suggested books

1. Albert', B. et at. (2003) *Molecular Biology of the Cell,* Fourth edition; Garland Sciences, New York.
2. Lewin, B. (2004) *Genes* VIII, John Willey, New York.
3. Klug, W, and Cummings, M. (2003) *Concepts of Genetics;* Seventh Edition; Pearson Education, Singapore.
4. Karp, G. (2005) *Cell and Molecular Biology,* Fourth Edition; John Willey, New York.
5. Gilbert, S. (1997) *Developmental Biology,* Fifth Edition, Sinauer Associates Publishers, Massachusetts.
6. Russell, P. (1998) Genetics, Fifth Edition; Addison Wesley Longman, New York.
7. Strachan, T. and Read, A. P. (2004) *Human Molecular Genetics,* Third Edition; Oxford University Press, Oxford.

Unit 5 Human Cytogenetics

Structure

5.1 Techniques in human chromosome analysis

5.1.1 Introduction

The correct chromosome number for man was established only after the *application of tissue culture methods* to cytogenetics. Before 1956, the chromosome number of man was considered to be 48. Tjio and Levan (1956), analyzing cultures of human embryonic lung fibroblasts, found a consistent chromosomal number of 46. At the same time and independently Ford and Hamerton (1956), using meiotic cells obtained from testicular biopsy material, found only 23 pairs of chromosomes. Mitotic cells from the same specimens contained 46 chromosomes. These two papers mark the beginning of modern human cytogenetics. The 15-year period from 1956 to 1971 saw the development of a standardized system of nomenclature which became more refined as the identification of human chromosomes became more precise. The delineation of most of the syndromes associated with chromosomal abnormalities occurred during syndromes associated with chromosomal abnormalities occurred during this period.

5.1.2 Terminologies used in the identification of human chromosomes karyotype & idogram

Human chromosomes can be arranged in an orderly fashion to produce a *karyotype* (Rowley, 1969). A karyotype is comoposed of individual chromosomes from a particular cell; the chromosomes are aligned in pairs and identified according to the standard nomenclature accepted by cytogeneticists. Karyotypes of different cells will reflect the variations in chromosomal morphology present in these cells. An *idiogram* is the schematized drawing of a composite of many karyotypes and is not directly related to any particular cell. Idiograms generally are not prepared for clinical purposes. But are used for comparing chromosomal patterns of different species.

Metaphase chromosome

Metaphase chromosomes differ from one another in size and shape (Fig. 5.1) Each metaphase chromosome is identified by its size, shape, and specific banding pattern. The absolute size of any chromosome varies with the stage of mitosis. Chromosomes are longer and less coiled in prophase and shorter and more compact at the end of metaphase. The duration of treatment with mitotic blocking agents and the type of hypotonic solution also influence the absolute size of the chromosomes. In general, the longest human metaphase chromosome is about 7-8 urn in length, whereas the shortest is about 2μm long. Each metaphase chromosome is composed of two *chromatids* joined at the *centromere* (the site of attachment of the spindle fibre). The position of the centromere is specific for

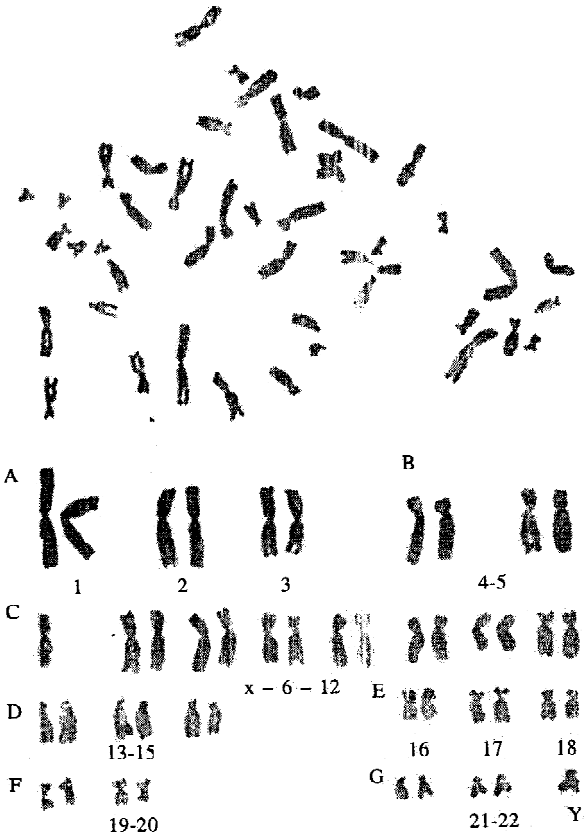


Fig. 5.1 Intact metaphase plate from a normal male with 46 chromosomes from bone marrow specimen (top) and karyotype of cell (bottom). Chromosomes are arranged in seven groups of morphologically similar chromosomes. The X chromosome is included with group C from which it cannot be distinguished. The Y chromosome in this cell is about the size of G-group chromosomes, but it can be differentiated from them by the size of the short arm

each chromosome and divides it into a long and short arm. The relative length of the two arms (arm ratio) is important for the identification of chromosomes.

A chromosome with a centreomere in the middle that divides it into two equal arms is called *metacentric.* When the centromere is somewhat nearer to one end of the chromosome, so that there is a distinct long and short arm, the chromosome is said to be submetacentric. If the centromere is very near to one end of the chromosome, which thus only a very short small arm and a relatively longer long arm, the chromosome is called *acrocentric* (Fig. 5.2). *A telocentric* chromosome, not normally found in human cells, has the centromere at the end.

5.1.3 Standardization of nomenclature of chromosome

Four conferences were held between 1960 and 1971 to revise the nomenclature used to identify human chromosomes along with the improvement of technologies. The last two conferences were held in Chicago and Paris. When the first reports on human chromosomal abnormalities appeared in 1956, each group of investigators used its own system of arranging and numbering the chromosomes. The necessity for adopting a uniform system was generally recognized, and a standard nomenclature of human mitotic chomosomes was adopted at the Denver Conference in 1960. Numbers were assigned to each pair of autosomes as nearly as possible in decending order of length. The sex chromosomes, XX pairs of autosomes (44) plus two sex chromosomes, making a total of 46 chromosomes. Seven groups of morphologically similar chromosomes could be distinguished (Table 5.1). It soon became evident, however, that not all pairs of chromosomes could be identified with certainty, even in preparations of the highest technical quality.

Patau (1960) proposed that the seven groups of morphologically similar chromosomes be indentified by capital letters A through G, an arabic number being added only when the individual chromosome could be identified with certainty. This system had the great advantage of flexibility, since it permitted general recognition of the group to which a chromosome belonged without implying identification of the specific chromosome involved in an abnormality. Patau's recommendations were accepted at the London Conference on the normal human karyotype (1963).

Whereas individual investigators were using the same systems (Denver and London) for identifyng individual chromosomes, they were using different systems for describing chromosomal abnormalities. This led to confusion when data from different laboratories were collated. A major achievement of the Chicago Conference in 1966 was the development of a uniform system of notation designed to facilitate coding for data retrieval. It was agreed that the analysis of the karyotype would be recorded with the total chromosomal number first, followed

by the sex chromosomes, and finally by any additional abnormalities. Thus the karyotype of a normal male was written 46, XY; a normal female was 46, XX. The recommended nomenclature symbols used in describing normal or abnormal chromosomes are summarized in Table 5.2.

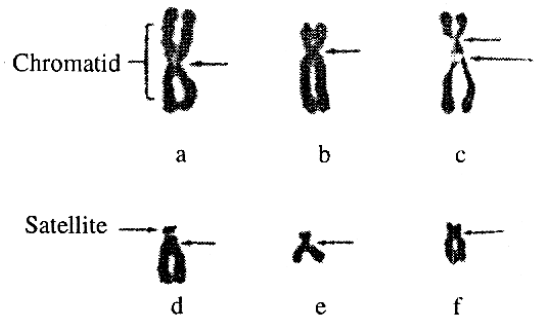


Fig. 5.2 Centromere position in typical chromosomes, a, Two chromatids joined at centromere or primary constriction. Centromere (arrow) is median and divides chromosomes in two equal arms. Chromosome is metacentric. b, Submedian centromere (arrow) divides the chromosome into short arm and long arm Chromosome is submetacentric. c, Submetacentric chromosome with secondary constriction (long arrow) in long arm. d, Large acrocentric chromosome with subteminal centromere (short arrow). Satellites (long arrow) are separated from short arm by secondary constriction, e and f, G-group chromosome and Y from same cell. Long arm chromatids of Y are close together and short arm is larger than G-group chromosome

Table 5.1 Systems of identification of human metaphase chromosomes*

Denver			Chromosome description	
Patau		Chromosome	Centromere	
group	Group	number	position	Morphology
A	1-3	1,3	Median	Metacentric
		2	Submedian	Less metacentric than above
B	4-5	4,5	Submedian	Submetacentric
C	X 6-12	X,6,7,9,ll	Submedian	Submetacentric but more metacentric than remainder
		8,10,12	Submedian	Less metacentric than above
D	13-15	13,14,15	Subtermina	Acrocentric—all may have satellites
E	16-18	16	Median	Metacentric
		17,18	Submedian	Submetacentric
F	19-20	19,20	Median	Metacentric
G	21-22	21,22	Subterminal	Acrocentric—all may have satellites
		Y	Subterminal	Acrocentric—similar in size to group G but usually morphologi cally distinct; long arms tend to be close together: satellites are not present

*Rowley (1969)

Table 5.1 Nomenclature symbols

Chicago Conference

A—G	the chromosome groups
1—22	the autosome numbers
X,Y	the sex chromosomes
diagonal(/)	separates cell lines in describing mosaicism
?	questionable identification of chromosome or chromosome structure
*	chromosome explained in text or footnote
ace	acentric
cen	centromere
dic	dicentric
end	endoreduplication
h	secondary constriction or negatively staining region
i	isochromosome
inv	inversion
mar	marker chromosome
mat	maternal origin
p	short arm chromosome
pat	paternal origin
q	long arm of chromosome
r	ring chromosome
s	satellite
t	translocadon
s t repeated symbols	duplication of chromosome structure

Paris Conference

A. Recommended changes in Chicago Conference nomenclature

\+ 1. The + and — signs should be placed *before* the appropriate symbol where they mean additional or missing whole chromosomes. They should be placed *after* a symbol where an increase or decrease in length is meant. Increases or decreases in the length of secondary constriction, or negatively staining regions, should be distinguished from increases or decreases in length owing to other structural alterations by placing the symbol h hetween the symbol for the arm and the + or — sign (e.g., 16qth +).

2. All symbols for rearrangements are to be placed before the designation of the chromosome (s) involved in the rearrangement, and the rearranged chromosome (s) always should be placed in parentheses, e.g., r(18), i (Xq), die (Y).

B. Recommended aditional nomenclature symbols

del	deletion
der	derivative chromosome
dup	duplication
ins	insertion
inv ins	inverted insertion
rep	reiprocal translocation*
rec	recombinant chromosome
rob	Robertsonian translocation* (“centric fusion”)

tan	tandem translocation*
ter	terminal or end ("p ter" for end of short arm; "q ter" for end of long arm)
:	break (no reunion, as in a terminal deletion)
: :	break and join
→	from-to

* Optional, where greater precision is desired than that provided by the use of t as recommended by the Chicago Conference.

5.1.4 Recommendations of the Paris Conference

The fluorescent karyotype, published by Caspersson et ah (1971), was accepted as the basis for the assigning of numbers to each chromosome.

A. Definitions

The bands seen with the fluorescent dyes (quinacrine) were called Q-bands and were accepted as the reference bands. Those bands of chromatin stained by methods that demonstrate "constitutive heterochromatin" were called C-bands and they are mainly confined to the centromeric region. The bands stained with basic dyes such as Giemsa were called G-bands and, except in one techique, they correspond quite well with Q-bands. One of the techniques using Giemsa, the exception just noted, gives patterns that are opposite in intensity to the G-bands; these were called R-bands.

A band was defined as a part of a chromosome which is clearly distinguishable from its adjacent segments by appearing darker or lighter with the Q, G, R, or C staining methods. By definition there were no "interbands." In the construction of the chromosome map, each band was refered to by its midline and not by its margins. A chromosome *landmark* was defined as a consistent and distinct morphological feature that is an important diagnostic aid in identifying a chromoosome. A region was defined as any area of a chromomsome lying between two adjacent landmarks. A chromosome arm lacking a prominent landmark consists of only one region.

B. Band numbering

Regions and bands are numbered consecutively from the centromere outward along each chromosome arm (Fig. 5.3). A band used as a landmark is considered as belonging entirely to the region distal to the landmark and is accorded the band number of "1" in the region. A band bisected by the centromere is considered as two bands, each being labeled as band 1, in region 1, of the appropriate chromosome arm.

For the designation of a particular band, four items are required: the chromosome number, the arm symbol, the region number and the band number within that region. These items are given in order without spacing or punctuation.

For example, Ip33 indicates chromosome No. 1, short arm, region 3, band 3. If a band defined in the present chromosome map has to be subdivided, the original band designation will be followed by a decimal point and the sub-bands will then be numbered sequentially from the centromere outward, e.g., Ip33.1; Ip33.2; Ip33.3, indicating that the orginal band 33 in the short arm of chromosome No. 1 has been divided into three sub-bands, 33.1 being proximal and 33.3 distal to the centromere. This system is thus relatively simple and yet sufficiently flexible to accommodate further refinements as the banding techniques are improved.

C. Characterization of chromosomes by the various banding techniques

The technical quality of the chromosomes is of the utmost importance for characterizing chromosomes by their banding patterns; in fact, it is much more important than previously adopted methods when morphology was the sole criterion for preparing the karyotype. The most distinct banding patterns are obtained when the chromosomes are relatively long and free of overlaps. It is important to note that the morphological characteristics of size and centromere position remain critical parameters used in the identification of chromosomes. Thus, the cytogencticist uses banding patterns as well as overall morphology to distinguish individual chromosomes. In general, the number of distinct bands increases with increasing length of the chromosome. It is thus, meaningless to mention the absolute number of bands in a chromosome arm, since the number varies with the state of contraction, the quality of the preparation and the type of treatment and stain. Once identification of the individual chromosomes by means of the major landmarks is mastered, careful observation of very good preparations can reveal a number of fine bands which can be used, among other things, for defining the site of chromosomal breakage and rejoining.

The intensity of fluorescence is influenced by the position of the chromosome in the metaphase. Chromosomes in the center of the cell frequently fluoresce more brightly than homologous chromosomes that are on the periphery. It is therefore necessary to consider the *pattern* of the bands in a particular chromosome and to appreaciate the fact that the overall intensity of fluorescence of homologs may be different. There may also be "spreading" of fluorescence from brighter to duller chromosomes, such that a 19 or a 22 adjacent to an X or the long arm of the Y may appear to be much brighter than normal.

The terms "distal" and "proximal" refer to the position of a band with respect to the centromere. The following terms were used in the Paris report to indicate the approximate intensity of fluorescence:

negative	no or almost no fluorescence
pale	as on distal lp
medium	as the two broad bands on 9q
intense	as the distal half of 13q
brilliant	as on distal Yq

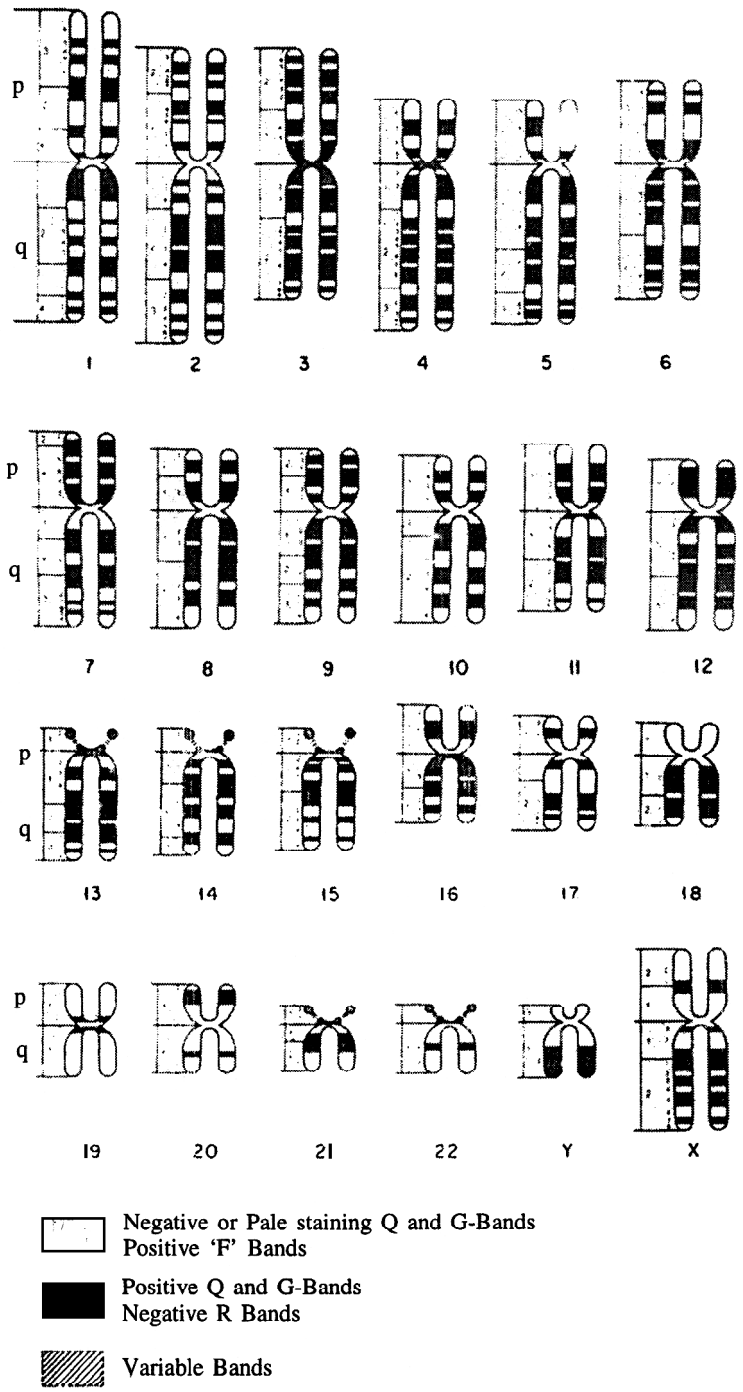


Fig. 5.3 Diagrammatic representation of chromosome bands as observed with the Q-, G-, and R-staining methods; centromere representative of Q-staining method only. Reproduced from the report of The Paris Conference (1972)

D. Characterization of chromosomes on the basis of different boarding techniques :

Chromosome 1 is the largest chromosome and is usually metacentric. The distal 40% of the short arm shows pale fluorescent bands (32 to 36), and the proximal segment shows two bands of medium fluorescence; the more proximal band divides region 1 from 2, and the more distal band divides region 2 from 3 (Fig. 5.4). The area of the secondary constriction, a poorly staining gap, which by definition is in the long arm, is adjacent to the centromere and shows negative fluorescence; it constitutes region 1. The long arm also contains a central intense band which divides region 2 from 3, with a less intense band distal to it which divides region 3 from 4.

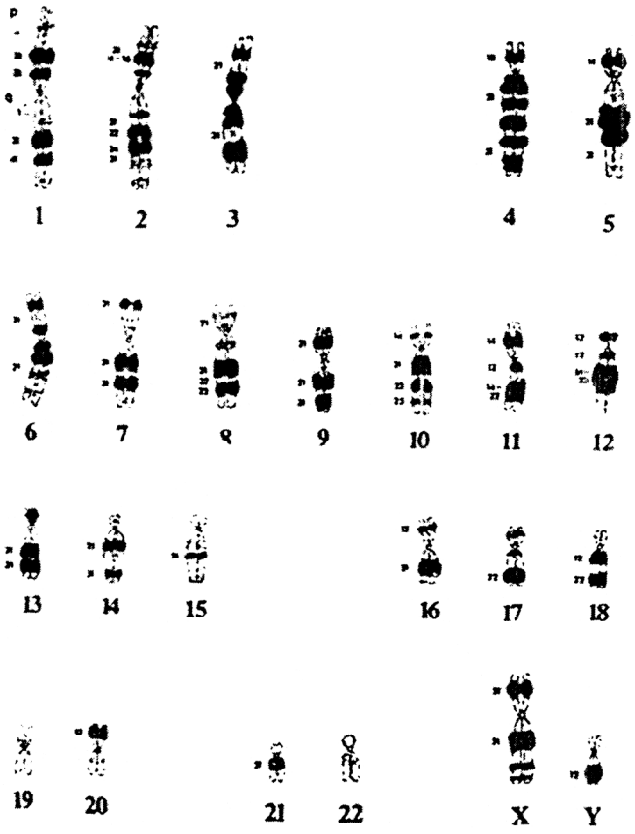


Fig. 5.4 Idiogram of fluorescent bands in human chromosomes (see text)

The large block of densely stained material in the long arm adjacent to theL centromere is the most prominent feature of No. 1 in cells treated to produce G-bands (Fig. 5.5). This area corresponds to the negatively quinacrine-stained secondary constriction. The two proximal bands in the short arm and the very

darkly stained central band and less darkly stained distal band in the long arm are also present. The end of the short arm is faintly stained.

The technique for staining C-bands revals the same densely staining region of the secondary constriction, generally called "constitutive heterchromatin." as do the G-band techniques. The R-banding technique (Dutrillaux and Lejeune, 1971) demostrates moderately staining material in this region; elsewhere along the chromosome, however, dark bands appear pale and vice versa. This reversal of staining intensity as compared with G-bands is particularly noticeable at the end of the short arm.

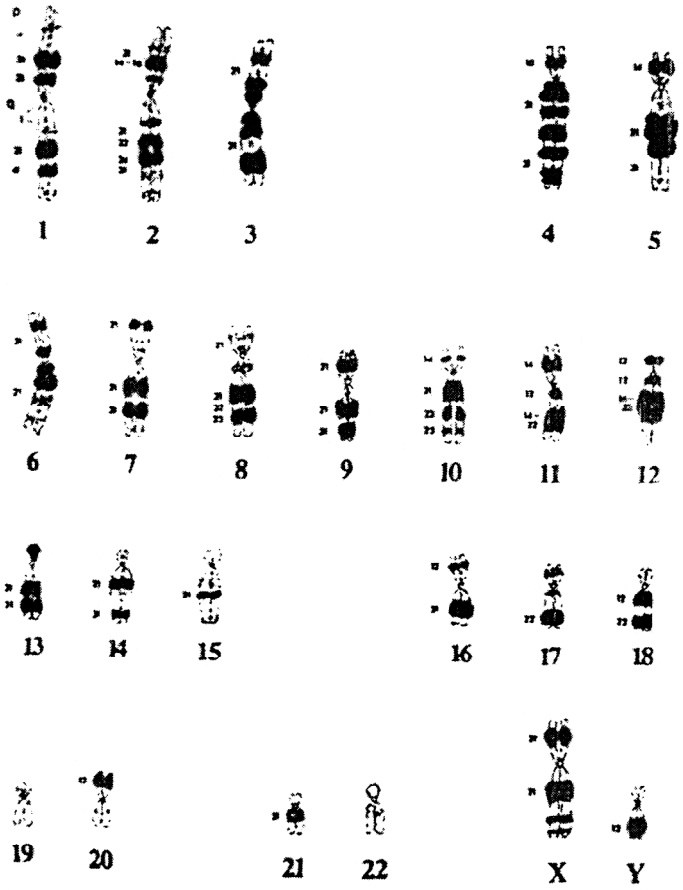


Fig. 5.5 Idiogram of Giemsa bands in human chromosomes (see text)

Chromosome 2 is a large chromosome, less metacentric than No. 1, and lacking conspicuous landmarks. It shows a number of bands of medium fluorescence along the entire length; the central band (14 to 16) in the short arm and the two cenral bands (22 to 24) of equal intensity in the long arm are most prominent. In the short arm, the dull band distal to the central band divides

region 1 from 2. In the long arm, the dull band proximal to the proximal central band divides region 1 from 2, whereas the dull band distal to distal central band divides region 2 from 3.

Chromosome 3 is a large metacentric chromosome which is smaller than No. 1. It has a nearly symmetrical banding pattern. There is a distinct band of pale fluorescence in the center of each arm which separates two broad bands of medium intensity. This pale band separates region 1 from 2 in each arm. The distal medium band in the short arm is narrower, but frequently appears more intrensely fluorescent than that in the long arm, whereas the terminal pale band is longer in the short than in the long arm.

Similar morpological features are observed in cells treated to produce G-bands. The centromeric regions stains darkly, but the variation in intensity of stain, seen with fluorescence, is not obsrved in G-bands.

Chromosome 4 is a long submetacentric chromosome that, similar to No. 2, lacks prominent landmarks. It has one band (15) of medium fluorescence in the short arm and four or five relatively evenly spaced bands of medium intensity in the long arm. The long arm is divided into three regions by a proximal dull band separating region 1 from 2 and a distal dull band separating 2 from 3.

Chromosome 5 has more distinctive characteristics than No. 4 and is frequently the first pair of B-group chromosomes indentified in the cell. The central band (14) of medium fluorescence in the short arm is brighter and frequently wider thanthat in No. 4. There is a broad central band of medium fluorescence which separates region 1 from 2 in the long arm, with a prominent distal pale band which separates region 2 from 3. Frequently, the terminal portion of the long arm of No. 5 is paler than No. 4.

Chromosome 6 is the largest and one of the last submetacentric C-group chromosomes. The most prominent feature is a distinct band of pale fluorescence in the middle of the short arm, which separates two bands of medium intensity and which also separates region 1 from 2. The long arm contains a number of bands of medium intensity; those bands near the centromere are frequently brighter and more idstinct. A dull band in the middle of the long arm separates region 1 form 2.

Chromosome 7 is slightly smaller than No. 6, but their centromere positions are similar. The two bands of intense fluorescence in the center of the long arm are the most prominent feature of this chromosome. The proximal band divides regions 1 and 2,whereas the distal band divides regions 2 and 3. There is a distinct band of medium fluorescence at the end of the short arm which divides region 1 from 2.

Chromosome 8 is one of the most submetacentric C group chromosomes, and it lacks distinctive landmarks. The short arm shows less intense fluorescence

than the long arm, and a cenrtral pale band which divides region 1 from 2 may be seen in good preparations. A distal pale band (22) separates two medium bands in the long arm, The more proximal of these medium bands separates region 1 from 2.

Chromosome 9 is the middle-sized C chromosome and is more metacentric than Nos. 8 or 10. The long arm shows a negatively staining centromeric region (12) with two distal evenly spaced bands of medium intensity. The proximal band of the pair may appear wider and separates region 1 from 2; the distal band separates 2 from 3. The short arm has a characteristic heart -shaped appearance with a central band of medium intensity which divides region 1 from 2, Both G and Q techniques show a prominent, faintly staining region near the centromere. The R-bands are the reverse, except for the centromeric region which is pale; the R-bands are similar in staining intensity to Q-and G-bands. The negatively fiuorescing and pale-staining region in the long arm near the centromere shows a large block of material that stains intensely with Giemsa after treatment to produce C-bands. This C-band material presumably represents one of the types of constitutive heterochromatin, the size of which may vary with individuals.

Chromosome 10 is one of the smaller, less matacentric C-group chromosomes which can be identified by the three evenly spaced bands in the long arm. The proximal band is most intense (and divides region 1 from 2) and the distal is the least intense. The short arm shows medium fluorescence.

Chromosome 11 is one of the least submetacentric C group chromosomes; and it may be very slightly larger than No. 10. It is somewhat similar to No. 9, but can be distinguished on the basis of the following features: There is a narrow medium band (12) in the middle of the negatively staining centromeric portion of the long arm. In poor preparations, this narrow band may be very faint or not apparent. A broad band of medium fluorescence is present in the middle of the long arm. A narrow dull band in the middle of this broad band separates region 1 from 2. This broad band usually appears as a *single* band that distinguishes it from the definite double band in the long arm of No. 9. The short arm shows medium fluorescence (14) and has a rather squarish appearance. In contrast, the short arm of No. 9 tends to taper near the centromere.

The C-banding technique reveals that the amount of centromeric stining material in No. LI, and the X is second only to that in No. 9.

Chromosome 12 is the most submetacentric of the C-group chromosomes; it is similar in size to No. 10. The short arm shows a band of mediumn fluorecence (12) which is smaller than that in No. 11, because the short arm is smaller..The band (12) of medium intensity in the long arm near the centromere is wider than that found in No. 11. A short band (13) of negative fluorescence separates band

ql2 from a distal long segment of medium intensity which divides region 1 from 2. This distal segment and the terminal dull band are both longer than the corresponding bands of No. 11.

The X *chromosome* is the third largest chromosome in the C group; it is, with 6 and 11, among the least submetacentric chromosomes in this group. The X chromosome is frequently the most fluorescent of the C-group chromosomes. Both Xs in the female show identical fluorescence patterns. In both arms, the region proximal to these medium bands shows pale fluorescence. Two other evenly spaced, less distinct and less intense bands of fluorescence are found in the distal long arm.

Chromosome 13 is the largest of the D-group chromosomes and shows intense fluorescence of the distal half of the long arm. In good preparations this segment is seen as two bands; the more proximal band divides regions 1 and 2. and the more distal divides 2 and 3. Flourescence of the satellites and short-arm region shows variable intensity and is an inherited variant; a proximal intense band in the long arm has also been observed as an inherited variant.

Chromosome 14 shows a broad medium to intense band in the proximal half of the long arm which divides regions 1 and 2 and a narrow medium and, dividing regions 2 and 3. close to the distal end of the long arm. Variable fluorescence of the satellites and short-arm region has been noted.

Chromosome 15 is the smallest D chromosome and shows the last intense fluorescence of its group. The proximal half of the long arm shows medium florescence which divides region 1 from 2, whereas the distal half is pale. This chromosome is distinguished from No. 14 by the absence of a distal *medium* band, although there may be a distal band of faint fluorescence in No. 15. In poor preparations, it is frequently difficult or impossible to differentiate No. 14 from 15. The satellites and short-arm region show variable fluorescence.

Chromosome 16 is the largest and most metacentric of the E-group chromosomes and is one of the few chromosomes that could be identified solely by morphology. The short arm shows a band (12) of medium fluorescence which is less intense than the central band which separates region 1 from 2 in the long arm. The long arm contains a proximal segment of negative fluorecence (11) which corresponds to the secondary constriction.

The G-band technique reveals the medium bands in the long and short arm and the densely staining secondary constriction. The R-bands are the reverse of Q-bands, except that the region of the secondary constriction is pale. Obanding reveals a large block of constitutive heterochromatin in the region of the secondary constriction which varies in length in different individuals.

Chromosome 17 is the palest staining of the E-group chromosomes. It has a

single distal band (22) of medium fluorescence in the long arm. The dull band proximal to this band divides region 1 from 2.

Chromosome 18 is the smallest E-group chromosome. The long arm contains two bands of medium intensity (12and 22), the proximal band being the brighter and wider of the two. The dull band between these two bands separates regions 1 and 2.

Chromosome 19 is the most weakly fluorescent chromosome in the karyotype and it is difficult to distinguish the long and short arms except in very good preparations. There is a fluorecent spot (12) in each arm adjustment to the centromere; the spot in the short arm is longer and brighter than that in the long arm.

G-bands are similar to Q-bands, and the chromosome shows the same very pale staining except for the centromeric region which is well stained.

Chromosome 20 also show weak fluorescence, but more than No. 19. The short arm is brighter than the long arm. The C-band is medium sized and smaller than in No. 19.

Chromosome 21 is the smaller of the G-group chromosomes and is much brighter than No. 22. The long arm shows a proximal intense segment. Which divides regions 1 and 2, and distal pale segment. There is variable fluorescence of the short-arm region and the satellites; this is an inherited variant. This variability has been useful in determining the source of the meiotic error that results in Down's syndrome (Robinson, 1973).

Chromosome 22 is the larger of the G-group chromosomes and shows very dull fluorescence, similar to No. 19. In fact, in some preparations it may be difficult to distinguish Nos. 19 and 22. A narrow pale band may be observed in the middle of the long arm. A bright fluorescent band in the short arm and variable fluorecence of the satellites are inherited variants.

G-bands are similar to Q-bands, although the centromere region stains darkly with Giemsa. The C-band is medium sized and larger than in No.21.

The *Y chromosome* was the first human chromosome identified with quinacrine fluorescence (Zech, 1969) because of the brilliance of the distal long arm in cells from all but a few males. The variability in the length of the Y chromosome is well established; it is now evident that this polymorphism is correlated with variation in the length of the brilliant segment.

The Y has a variable appearance with G-banding techniques, but may show two distal bands; it is relatively pale when stained with the R-banding technique. The size of the distal C-band in the long arm is directly related to the length of the brilliant segment on fluorescence.

5.2 Human karyotype—banding—nomenclature

5.2.1 Introduction

Metaphase chromosomes show little morphological differentiation in conventional preparations. The size, the position of the centromere, and occasional secondary constriction(s) are the only criteria that can be employed for recognizing chromosomes. In species with high diploid numbers, chromosome pairs with similar morphology become increasingly common, thus making the identification of individual pairs extremely difficult. The human karyotype is comparatively favorable because at least the chromosomes can be classified, according to morphology, into seven groups, and a few pairs can be identified unequivocally. In the karyotype of the laboratory mouse, all chromosomes are acrocentric and do not even allow grouping.

Cytologists have attempted a variety of ways, such as distribution of chemical-i induced breaks, unstained chromosome regions induced by low temperature, and diferential DNA replication time revealed by autoradiography, to further differentiate the chromosomes longitudinally, but all these methods are laborious and the results are ambiguous. The first break through was recorded not long ago when Caspersson I and his collaborators (1969a, b) found that certain fiuorochromes, e.g., quinacrine mustard, when applied to cytological preparations and observed with ultraviolet optics, produced characteristically bright and dark bands. Later Caspersson *et al.* (1970a.b) applied the technique to human chromosome prepsrations and found that the fluorescent banding pattern is likewise specific for each chromosome pair.

Credit to the second major advance must go to Joseph G. Gall and Mary Lou Pardue who perfected the in situ DNA/RNA hybridization technique. In their studies on the cytological locations of the satelite DNA of the laboratory mouse (Pardue and Gall. 1970), they treated the cytological preparations with a series of chemicals in order to achieve molecular hypridization. In these preparations, the centromeric areas of the mouse chromosomes stained more deeply with Giemsa than the Chromosome arms. They regarded the densely stained centromeric areas as heterochromatin. The discovery of a simple staining procedure led to an explosive activity in inventing new procedures, particularly regarding the chromosomes of man and other mammals.

5.2.2 Classification of banding patterns :

The Q-bands. Florescent banding with quinacrine mustrard or quinacrine dihydrochloride.

The C-bands. Constitutive heterochromatin revealed by the PardueGall *in situ* hybridization procedure or its modifications.

The G-bands. Crossbands of chromosomes revealed by a variety of procedures. These bands coincide well with Q-bands, i.e., deeply stained G-bands are brightly fluorescent in Q-band preparations.

The R-bands. The "reverse" banding pattern following the procedure of Dutrillaux and Lejeune (1971).

5.2.3 Cell harvest and slide preparations

Arresting

Harvesting cells is strictly conventional. Bone marrow cell culture or any cell population containing a high incidence of mitosis are suitable. Agents such as colchicine, colcemid, and vinca alkaloid can be used to accumulate mitosis. However, over condensed chromosomes yield very poor banding, so that prolonged mitotic arrest should not be done.

Fixation

The cell polulations should be treated with a hypotonic solution prior to fixation. It matters little which kind of hypotonic solution is used. The cells, after hypotonic solution treatment, are fixed according to the type of slide preparations to be made, viz,, "Carnoy" fixative (1 glacial acetic acid: 2 methanol) for air-dired slides. For cell populations which require squash technique (e.g., many solid tissues), the fixative to use is 45—50% acetic acid. However, the preparations are not suitable for G-bands, though they are excellent or C-bands.

It is suggested that the air-dried slides be incubated at 37°C for 1 hour (without covering) and thereafter be kept in air-tight slide boxes containing a drying agent such as silica gel. The slides may be used immediately after this incubation period or may be stored for a few weeks.

5.2.4 C-Banding

In good C-band preparations, the constitutive heterochromatin should stain deeply and the euchromatin should show only a faint outline of the chromosome. However, flame-dried preparations, when improperly treated, will show G-bands as well as C-bands, which is confusing.

The original procedure (Pardue and Gall, 1970) and the modifications thereof (Arrighi and Hsu, 1971; Yunis et al., 1971) are useful for distinguishing constitutive heterochromatin and euchromatin in mammalian chromosomes. The procedure to be described is a simplified version designed to reveal C-bands using air-dried and flame-dried preparations, although some comments will also be made for squash preparations.

A. Reagents

1. HCL : prepare 0.2 N solution.

2. NaOH : Prepare a 0.07 N solution.
3. SSC : Prepare a lOx concentrate (a solution of 0.15 M sodium citrate and 1.5 M NaCL) and dilute with distilled water to the desired concentrration.
4. Giemsa staining solution.
5. Phosphate buffer solution (0.01 M Sorensen's phosphate buffer, pH 7.0).

B.Procedure

1. Treat the slides with HCL at room temperature for 15 minutes. Rinse with distilled water three times.
2. Treate the slides with NaOH for 2 minutes. Rinse with 70%, then 95% ethanol three times for a period of 5 minutes each. Air dry the slides.
3. Place slide horizontally, with cell-side up, in a moist chamber, and add either 2x or 6x SSC to the cell area of the slide. Place a coverglass over the SSC solution.
4. Incubate the moist chambers containing the slides at 60°—65°C for 16-20 hours.
5. Rinse in either 2x or 6x SSC (three times, 5 minutes each), 70% ethanol (three times, 5 minutes each), 95% ethanol (three times, 5 minutes each), and air dry.
6. Stain in Giemsa solution.

C. Comments

1. After a considerable amount of experimentation, it is generally opined that the HCL treatment is an important step in eliminating the G-bands in C-band preparation, particularly when air-dried and flame-dried slides are used. In sqush preparations, HCL treatment is not a vital step and may be omitted for C-band preparations.

2. The concentration of NaOH and the duration of the NaOH treatment are important. As a standard, one may start with 0.07 N for 2 minutes. This combination may be too strong and the resulting euchromatic chromosomes may appear bloated and show an empty appearance. If such a result is obtained, one must experiment with reduced concentration of NaOH solution (0.02 N, 0.01 TV) and time of treatment (1 minute, 30 seconds, or even 15 seconds). If the chromosomes still appear distorted, one should then try the solution suggested by Stefos and Arrighi (1970): a 2 x SSC solution with pH adjusted to 12 by NaOH. This solution is particularly useful for small chromosomes such as the microchromosomes of the birds and the chromosomes of *Drosophila* (Hsu, 1971). Conversely, some C-bands require a prolonged NaOH treatment.

3. Many laboratories use Coplin jars filled with 2 x or 6 x SSC for the over night incubation. This is undersirable because the glass slides will stain heavily with Giemsa, thus interfering with the observations on the chromosome banding.

Incubating slides in moist chambers eliminates this defect. If, however, such defect is not observed, Coplin Jars are of course convenient.

A simple moist chamber can be constructed as follows. Use a Petridish of suitable size. If square (120 mm each side) Petri dishes are used, 10 ml of either 2 x or 6 x SSC is placed in the bottom (15 ml for overnight treatment at 65°C). Next a stand for the slides is placed in the bottom half. The stand should be as small as posible and of sufficient height so that the slide is above the salt solution in the bottom half of the Petri dish. The slide(s) is placed on the stand. A few drops of the solution are placed on the slide to cover the cells. A coverglass is then placed over the solution, the Petri dish is covered, and the entire chamber is placed in an oven set at the desired temperature.

4. The stock Giemsa-staining solution is diluted with phosphate buffer, and the concentration varies with each new lot of stain. Usually concentrations varying from 2 to 10% have been used and stained the slides from 5 to 30 minutes.

5. Slides should be of good quality and should be cleaned in some manner. Slides can be cleaned in 95% ethanol, soap, and dilute HCL. AH seemed to be acceptable. Coverglasses should also be cleaned.

D. Squash preparation

This section is for squash preparations only. If squash preparations are used for C-bands, the slides should be dipped into a solution of 0.1 % gelatin and 0.01 % chrome alum and dried prior to squashing. This thin coat prevents cellular loss during the treatments. However, if the HCL treatment is omitted, the slides should be treated with RNase (100 pg/ml. diluted in 2 x SSC) at 37°C for 1 hour using the moist chamber method. Rinse the slides for 5 minutes each in three changes of 2 x SSC, 70% ethanol, and 95% ethanol and air dry. Treat the slides with NaOH solution (0.07 N NaOH or 2 x SSC, pH 12). Try several treatment times, e.g., 1 minute, 2 minutes, 4 minutes, etc. if 0.07 //NaOH is used, rinse in 70% ethanol, 95% ethanol, dry,and incubate in 6 x SSC as usual. If 2 x SSC at pH 12 is used, rinse slides in three changes of 2 x SSC for 10 minutes each. Do not dry but drain and immediately place the slides at 65°C in a moist chamber, as suggested earlier.

5.2.5 G-Banding

G-bands are the crossbands of various width and shades stained with Giemsa, Leishman's, Writht's, or similar stains. They usually correspond to the Q-bands but do not always correespond to the C-bands. In some cases, the C-bands and the G-bands may be opposite in staining behavior. For example, the C-band of human chromosome 9 is relatively unstained with G-band techniques.

The Y chromosome of man shows a distinct C-band in the distal portion of the long arm, but the same chromosome is somewhat variable in G-band staining, usually deeply stained throughout. Thus, G-banding does nor replace C-banding in assessing information.

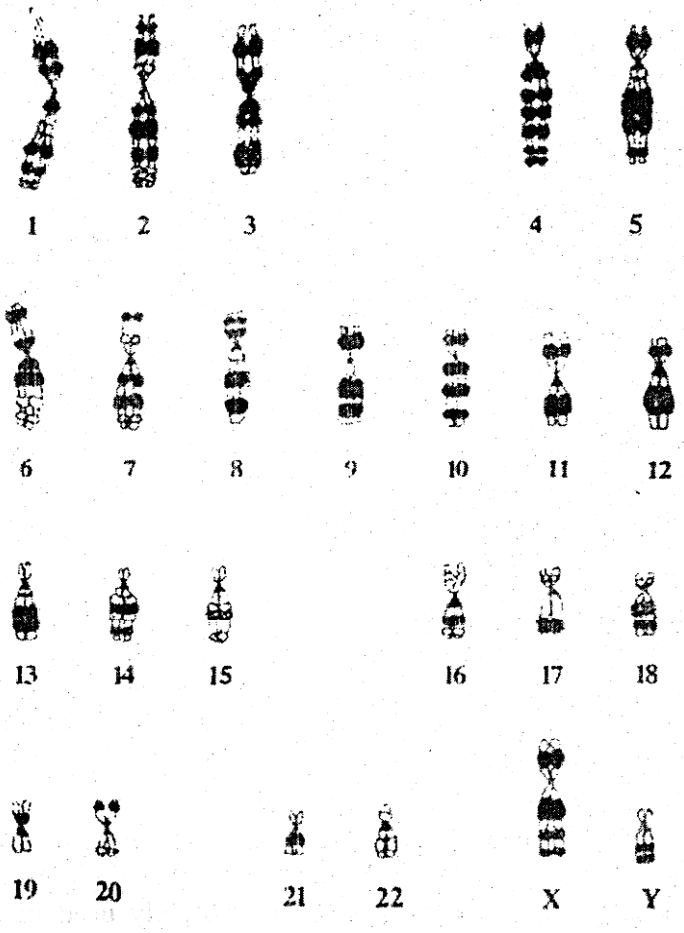


Fig. 5.6 Kayotype of human male showing C-banding pattern from a flame-dried preparation. Note the large amount of heterochromatin at the centromere areas of Al, C9, E16, and the distal portion of the Y. Variations in amounts exist in the two homologs of Al, Courtesy of Mrs.Ann Craig-Holmes and Dr. M. W. Shaw

There were many papers published in 1971 and 1972, each proposing a certain procedure to reveal crossbands in mammalian chromosomes (Summer et al., 1971; Drets and Shaw, 1971; Patil et al, 1971; Schned, 1971a,b; Seabright, 1971; Wang and Fedoroff, 1972; Kato and Yosida, 1972; Utakoji, 1972). Since the results of the various procedures are similar to one another, it is unnecessary to present the procedures for all of them. The trypsin procedure described here more or

less follows the one devised by Seabright (1971) with recommendations for individual laboratory modifications.

A. Reagents

1. Trypsin solution : Seabright uses Bacto trypsin (Difco Catalogue No. 0153) prepared by adding 10 ml of sterile distilled water or isotonic saline to each vial as the stock solution. This stock solution is diluted 1 : 10 with saline before use.

It is really not necessary to use the particular brand of trypsin recommended by Seabright. Most laboratories carry monolayer cell cultures which require trypsin to dislodge the cells for harveest or for subculturing. Usually it is a crude trypsin solution (0.20—0.25% dissolved in a balanced salt solution without Ca^{2+} and Mg^{2+}). In some laboratories, purified trypsin solution (0.01—-0.02%) is used. Whatever the kind and the concentration, the trypsin solution routinely used in the cell culture laboratory can be considered as the "stock solution". The trypsin solution used for G-bands is prepared by diluting the stock trypsin solution with saline, balanced salt solution, or "rinsing solution" (balanced salt solution without Ca^{2+} and Mg^{2+}). In our laboratory we use the rinsing solution.

For laboratories using trypsin solution for G-banding only, it is advisable to dispense the stock trypsin solution in small containers and store them in a freezer. Keep only a small amount in the refrigerator for immediate use.

2. Rinsing solution : Physiological saline or balanced salt solution without Ca2~ and Mg2'. This solution is used diluent of the stock trypsin solution as well as for rinsing the slides after the trypsin treatment.

3. **95% Ethanol**
4. **Giemsa staining solution :** See C-bands above.
5. **Phosphate buffer :** Like that of C-bans.

B. General principles

Many factors may influence the success of the G-band staining by trypsin treatment. It is, therefore, pointless to follow a set recipe without knowing these factors because the preparations may give excellent results if one knows how to modify the procedure. The success of G-banding depends primarily on the combination of the concentration of the trypsin solution and the duration of treatment, but the following factors dictate the correct combination :

1. **The method for preparing the slides :** The flame-dried preparations are more resistant to the trypsin treatment than air-dired preparations.
2. **The age of the slides :** The longer the slides are stored, the more resistant the cells are to the treatment. Cells of very old slides often give spotty, instead

of banded, chromosomes. Refixing of the slides in the Carnoy fixative sometimes helps.

3. **Heating the slides :** Air-dried slides be heated at 37°C for 1 hour (without covering). This procedure seems to give more consistent results.
4. **The salt composition of the trypsin solution (including the diluent) :** The presence of divalent cations in the solution slow the reaction but do not prevent it.
5. **The temperature of the trypsin solution (the higher the temperature, the faster the reaction) :** In laboratories with air conditioning, room temperature is suitable. The trypsin solution should be stabized at room temperature for approximately 30 minutes prior to use. For laboratories without room temperature control, it is probably a good practice to stabilize the trypsin solution at 4°C (refrigerator) or even in an ice bucket (Deaven and Petersen, 1973).
6. **Trypsin concentration :** We suggest a dilution of 1:5 or 1:10. If these are too strong, dilute further. If too weak (as in the case of flamelried preparations), the concentration may be raised to 1:2 or the undiluted trypsin used.
7. **Time of treatment :** The time of treatment of course depends on all the factors mentioned above. As a general principle, it should be adjusted to give good results in not more than 2 minutes but not less than 30 seconds.

C. Procedure

1. Prepare trypsin solution in a Coplin jar. Using a Coplin jar is somewhat more convenient than flooding the slides with the solution but either way is acceptable. Use two or three slides and vary the duration of trypsin treatments as the initial monitor. Since for best results monitoring is necessary, it is advisable to prepare at least 10 to 12 slides of good quality.
2. Rinse with physiological saline or rinsing solution. Seabright suggested, at this stage, inspection of the wet slides by phase-contrast microscopy to determine the effect of trypsin treatment. The chromosomes should appear slightly swollen. The preparations can be treated again with trypsin if necessary.
3. Rinse with 95% ethanol and let dry.
4. Stain with diluted Giemsa (2% Giemsa solution in phosphate buffer) for 4— 10 minutes. Overstaining may obliterate some of the lighter bands.
5. The slides can be pulled out of Giemsa, rinsed quickly with deionized water, and air dried. It is not necessary to mount the slides.

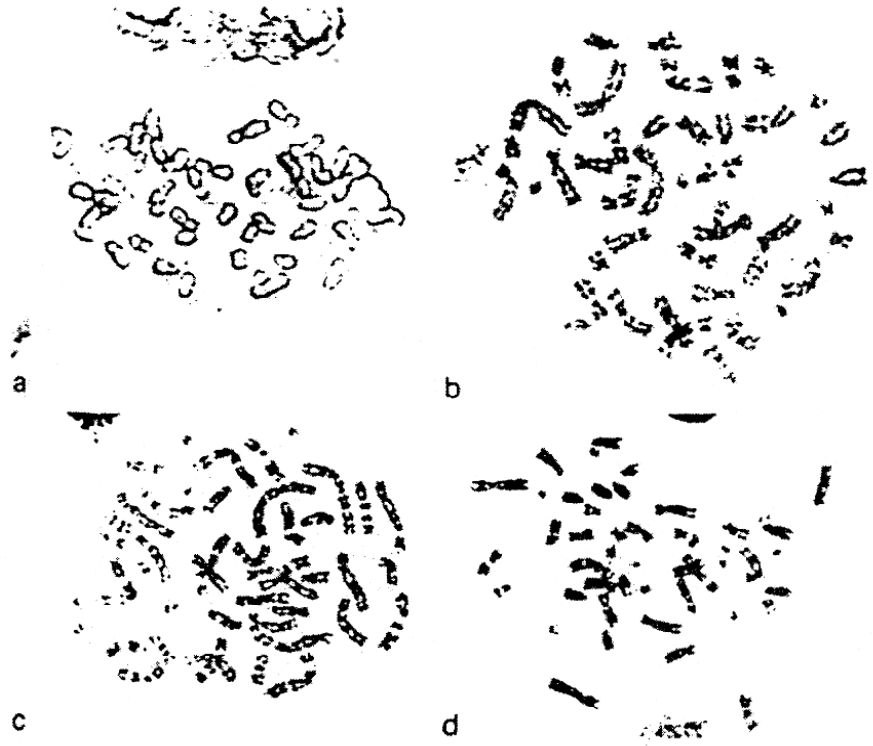


Fig. 5.7 Metaphase plates of human cells showing results of various durations of trypsin treatment from extreme over treatment to proper timing, (a) Extreme overtreatment; (b) overtreatment, highly unsatisfactory; (c) slight overtreatment; chromosomes are fuzzy but discreate bands can be seen; (d) proper treatment

Examine the stained preparations to determine the proper duration of the trypsin treatment. The chromosomes in undertreated preparations, in overtreated preparations, the chromosomes will show a series of appearances ranging from completely "ghost" chromosomes (Fig 5.7a) to those with poorly differentiated crossbands and fuzzy outlines (Fig. 5.7b). The appearance of these cells indicates that the treatment time or the concentration of trypsin solution should be reduced. Figure 5.7c shows a metaphase with reasonably good but slightly overtreated chromosomes, and Fig 5.7d shows proper G-bands.

6. Once the proper combination of trypsin concentration and the duration of treatment is determined by the preliminary monitor, treat the rest of the slides according to the best combination in the same day using the same solutions. The solutions in the Coplin jars should be discarded each day.

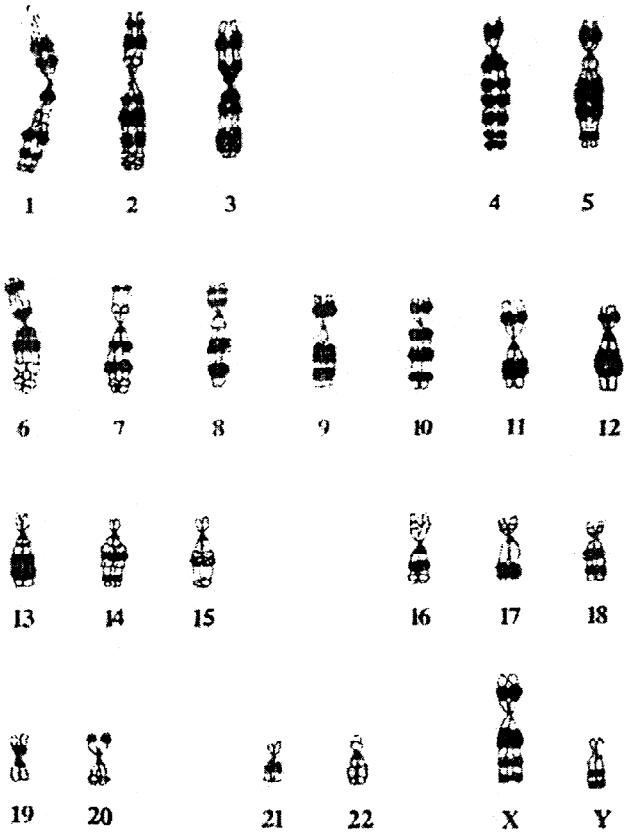


Fig. 5.8 Karyotype of human male showing G-banding pattern produced by the trypsin technique. Courtesy of Mrs. Marina Seabright

7. Figure 5.8 presents a male human karyotype showing the distribution of G-bands.

D. Comments

1. If the trypsin procedure does not give satisfactory G-banding, it may be worthwhile to try the urea procedure of Kato and Yosida (1972). Prepare a stock urea solution (*8 M* aqueous) and mix 3 parts of this stock solution with 1 part of Sorensen's phosphate buffer (0.15 M, pH 6.8), heat to 37UC. Treat the slides at 37°C for 10 minutes. Rinse in tap water and stain directly without drying. For old slides, elevate the temperature to 60°C.

2. For combination staining, e.g., Q-banding and C-banding of the same cell, it is advisable to perform Q-banding first. The same slide can be then used for C-banding. It is recommend, however, separate runs for C-banding and G-banding. From each sample, one can always prepare a sufficient number of slides for slides for al treatments as well as some for storage.

5.3 Numerical and structural abnormalities of human chromosomes—syndromes.

5.3.1 Introduction

Fifty years has elapsed since Tjio and Levan in 1956 established that human somatic cells have 46 chromosomes. There technique employing treatment of cells grown in culture with colchicine and hypotonic solution, was applied by numerous workers over the next half-dozen years, demonstrating the chromosomal findings in such disorders as trisomy 21, trisomy 13, trisomy 18, Turner's syndrome, and Klinefelter's syndrome. During this same period, the significance of the Barr body was realized, and lyonization of the X chromosome was hypothesized.

The next few years saw the application of autoradiography for better identification of chromosomes, the delineation of more subtle chromosome disorders such as the cri-du-chat (5p—) syndrome, Wolf-Hirschhorn (4p—) syndrome, 18p—, 18q—, 13q—and a host of other anomalies. Amniocentesis was employed for prenatal diagnosis.

Following 6 years have been marked by advances in our understanding of heterochromatin, chromosome banding, the identification of the Y body, and documentation of previously unrecognized alterations (small translocations, inapparent inversions etc.).

From the first conference held in Denver in 1958 to the one, convened in Paris in 1971, chromosome nomenclature has developed parri passu with advances in our understanding of these disorders.

Gross human chromosome abnormalities are not rare. Over 25% of human abortuses lost before the eighth week of pregnancy have abnormal karyotypes. Large surveys on newborns have consistently shown that about 0.5% exhibited aneuploidy (Sergovich et al., 1969; Lubs and Ruddle, 1970).

5.3.2 Autosomal abnormalities of human chromosome—syndromes

A. 4p— Syndrome (Wolf-Hirschhorn Syndrome)

The syndrome described independently by Wolf *et al.* (1965) and Hirschhorn *et al.* (1965), results from partial deletion of the short arm of one of the late replicating no. 4 chromosomes. It is much less common than the 5p— syndrome. About 25 cases have been described to date, all sporadic Translocation has not been demonstrated with the possible exception of the case described by Wilson *et al.* (1970). Parental age has been somewhat increased (Fryns et al., 1973). A 4r syndrome has been described (Carter *et al.* 1969).

The disorder is characterized by severe psychomotor and growth

retardation. Birth weight is usually about 2000 gm in spite of normal gestation time. Fetal activity is diminished. Most infants are hypotonic.

The skull is microcephalic and often there is cranial asymmetry. In a few cases, midline scalp defects have been noted (Hirschhorn et at., 1965; Wolf *et al,* 1965 Miller *et al,* 1970). Hemangioma on the brow is frequent. A prominent glabella and ocular hypertelorism are almost constant features. Divergent strabismus, antimongoloid obliquity of the palpebral fissures have been noted in about half the cases. Iris coloboma has been found occasionally.

The ears have narrow external canals and are low set and simplified in from. The nose is misshapen or beaked with a broad base. The philtrum is short with a down-turned Cleft lip or, especially, cleft palate and micregnathia have been noted in most cases.

Males comonly exhibt cryptorchidism and, especially, hypospadias. Absent uterus and streak gonad have been described. Congenital heart malformations, most often atria! or ventricular septal defects, have been noted in about 50% (Wolf *et al.*, 1965, Arias *et al.,* 1970, Guthrie *et al.,* 1971) and may result in death during the first year.

In several patients there has been dimpling of the skin over the sacrum and elsewhere, such as shoulders, elbows, or knuckles. The pelvic and carpal bones are late in ossification. Pseudoepiphyses are seen in the phalanges and at the base of each metacarpal.

B. 5p—Syndrome (Cri-du-Chat Syndrome)

Described initially by Lejeune *et al.* (1963), over 150 examples have been documented to date. The syndrome is present in about 1% of institutionalized individuals with intelligence quotients less than 35. The syndrome results from deletion of 35—55% of the short arm of one of the early replicatig B group chromosomes (German *et al,* 1964; Miller et al, 1969). Most decletion are thought to occur as a result of two breaks. If these occur in the short arm, an interstital deletion results. If deletion occurs in both arms, a ring chromosome is produced (Rohde and Tompkins. 1965). Maternal age is not elevated. About 70% of those identified at birth are females: however, most older patients have been male (Breg *et al,* 1970). The reason for this discrepancy is not evident. Mosaicism has also been described; patients having all the stigmata of the full-blown syndrome (Zellweger, 1966; Mennicken *et al,* 1968). About 10—15% result from translocation (Warburton and Miller, 1967). Pericentric inversion has also been described (Faed *et al,* 1972).

As the name implies, the syndrome is characterized by a catlike, weak, shrill cry in infancy caused by hypoplasia of the larynx (Ward *et al.*, 1968). However, the cry usually disappears with time, even within a few weeks of age

(Gordon and Cooke, 1968; Breg *et al,* 1970). The cry, almost one octave higher than normal, is quite monotone in quality (Schroeder *et al,* 1967).

The infant face is characterized by microcephaly, round form, hypetelorism, antimonogoloid obliquity of. palpebral fissures, epicanthus, bilateral alternating strabismus, broad nasal bones, and low-set ears. Most patients have mild micrognathia. However, the rouness of the face and the ocular hypertelorism disappear with age. The face becomes thin and the philtrum short. Permature graying of the hair has been noted in about 30%. Dental malocclusion is common (Breg *et al,* 1970, Gordon and Cooke, Niebuhr, 1971).

There is usually severe mental retardation (I. Less than 25), failure to thrive, and hypotonia in infancy. Birth weight is usually less than 2500 gm in spite of normal gestation time. Adult height usually ranges from 124 to 168 cm. (49 to 66 inches) Various musculoskeletal anomalies have included hypotonia, flat feet, mild scoliosis, large frontal sinuses, small ilia, syndactyly, and short metacarpals and metatarsals (Mennicken *et al,* 1968).

Dermatoglyphic alterations include simian creases in about 35%. Eight or more whorls have been noted in about 30%.

C. Group C deletion, trisomy, trisomy mosaicism, and partial trisomy

There are seven pairs of C-group chromosomes and hence many possible types of trisomy, partial trisomy, or deletion involving this group. These states, with few exceptions, have not been clinically recognized and, thus, are probably lethal. The use of newer banding techniques on the chromosomes of abortuses may shed light on this question (Hirschhorn *et al,* 1973).

There is a paucity of information concerning anomalies of chromosome no. 6 documented by banding, de Grouchy *et al.* (1968) described a child with bulbous nose, preauricular tubercle, hernias, hypospadias, undescended tested, deep acromial dimples, and psychomotor retardation.

Deletion of the long arms of chromosome no, 7 was reported by Shokeir *et al.* (1973). The child exhibited psychic and somatic retardation, urinary malformations, flexion contractures at the elbows, and low finger ridge count. The most striking aspects of trisomy 8 or trisomy 8 mosaicism syndrome are mental retardation, abnormally shaped skull, reduced joint mobility, various vertebral anomalies, supernumerary ribs, strabismus, absent patellae, short neck, long slender trunk, cleft palate, and marked palmar and plantar creases (Oikawa *et al.,* 1969; Lejeune *et al,* 1969; Emberger et al, 1970; Riccardi *et al*.)

Deletion of the short arms of chromosome no. 9 was noted by Alfi *et al.* (1973) and probably by Kistenmacher and Punnett (1970). The patients had trigonocepha, mental retardation, ocular hypertelorism, anteverted nostrils, malformed pinnas. long philtrum, short neck, hypertonia, congenital heart disease,

and an incrreased number of digital whorls.

Trisomy for the short arm of no. 9 was defined by Rethore *et al.*, (1973) who reviewed earlier cases. Clinical features included mental retardation, microcephaly, enophthalmos, hypertelorism, mongoloid palpebral fissures, bulbous nose, abnormal pinna, hypoplasia of the phalanges, and abnormal finger creases.

Partial trisomy for the long arm of chromosome no. 10 was reported by de Grouchy *et al*., (1972). Facial dysmorphia was evident with microcephaly, large forehead, flat round face, arched and wideset eyebrows. Antimongoloid palpebral fissures, microphthalmia, cleft palate, small nose with depressed bridge, malformed pinnas, short neck, micrognathia, various skeletal anomalies (osteoporosis, various rib abnormalities, scoliosis), congenital heart disease, and genitourinary defects. All patients had severe mental retardation.

Partial trisomy for the short arm of chromosome no. 11 was described by Falk *et al*. (1973) and Sanchez *et al*. (1974). Findings common to both cases were mental retardation, marked frontal bossing, nystagmus, antimongoloid palpebral fissures, strabismus, broad fingers or toes, and cleft lip and/or palate.

D. Trisomy 13 syndrome (Patau's syndrome, trisomy D_1)

Trisomy 13 was first recognized by Patau *et al.* (1960), although Bartholin in 1657 may have given the first description of the clinical features (Warburg, 1960). The phenotype is so striking that diagnosis is usually made on clinical ground before the karyotype has been made. The incidence has been estimated to be about 1 per 6000 births (Conen and Erkrnan, 1966).

Arhinencephaly, apneic spells, seizures, feeding difficulties, severe mental retardation, and deafness are common. Any of the holoprosencephalic states (cyclopia, ethmocephaly, cebocephaly, and premaxillary agenesis) may be associated with trisomy 13 (Conen *et al.*, 1966; Fujmoto *et al.*, 1973). Moderate microcephaly with sloping forehead and wide sagittal suture and fontanels have been noted in over 60%.

Microphthalmia or iris coloboma with retinal dysplasia, ocular hypertelorism, and malformed pinnas occur in about 80% (Cogan and Kuwabara, 1964). Capillary hemangiomas in the glabellar region and localized scalp defects in the parieto-occipital area have been described in about 75%. Cleft lip and/or cleft palate and micrognathia have been noted in 60-70% (Conen *et al.*, 1966; Taylor, 1968).

Musculoskeletal abnormalities include postaxial polydactyly of the hands or feet with overlapping flexed fingers (about 75%) with hyper convex narrow fingernails. The calcaneus is often prominent and frequently there are rockerbottom feet.

At least 80% have congenital heart defects, genital anomalies include cryptorchidism (over 90%) in males and bicornuate uterus (about 50%) and hypoplastic ovaries in females.

Polymorphonuclear neutrophils frequently (25—80%) have nuclear projections in cases of trisomy 13 owing to primary nondisjunction. Excellent ultrastructural study of the projections has been carried out (Waltzer *et al,* 1966, Lutzner and Hecht, 1966). Fetal hemoglobin, Hb-Gower and other hemoglobins have been elevated but there is good evidence that those changes disappear with age and merely represent general delayed maturity (Marden and Yunis, 1967).

DNA replication studies have demonstrated that the D-group chromosome involved is number 13, which is the longest and the latest of the pairs to replicate (Yunis and Hook, 1966).

1. Trisomy 13 Caused by Primary Nondisjunction

About 75% of cases of 13 trisomy are caused by primary nondisjunction. There is no sex predilection. The mean age for mothers of infants with 13 trisomy caused by this type is elevated (32.4 years), far higher than for cases caused by translocation or mosaicism (Magenis *et al.*, 1968; Taylor *et al.*, 1970).

There have been several examples of 13 trisomy occurring with other chromosomal abnormalities in the same sibship (Klinefelter's syndrome, Turner's syndrome, Down's syndrome and triploidy), but this may be chance association (Visfeldt, 1969).

2. *Translocation D_1*

About 20% of the cases of trisomy 13 are caused by translocation, far more common than occurs in Down's syndrome (Magenis *et al.*, 1968; Taylor *et al.*, 1970). In at least 85%, the translocation has occurred between two D chromosomes. Maternal age is not elevated (25.6 years). There appears to be definite male predilection. Fertility and intelligence in balanced carriers are quite variable (Wilson, 1971).

3. *Mosaicism*

About 5% of the cases of Dj trisomy are caused by mosaicism. About half of these examples are caused by an extra chromosome 13 in proportion of the cells. The remainder result from a complex assortment of chromosomal abnormalities (Magenis *et al.*, 1968; Taylor *et al.*, 1970).

As in translocation D_1 trisomy, the age of the mother of a D_1 trisomy mosaic is not elevated (25.4 years) in contrast to mothers of G_1 trisomy mosaics. The clinical stigmata, as expected, are less severe than in those of children with classic trisomy 13 (Bain *et al.*, 1965).

4. Partial Trisomy

Partial trisomy for the distal segment of the long arm of chromosome 13 was documented by banding techniques by Taysi *et al.* (1973) and by Escobar *et al.* (1974). The latter authors reviewed several case reports which had been documented prior to the advent of banding. Common clinical characteristics included psychomotor retardation, seizures, microcephaly, *frontal bossing,* open anterior fontanel, short neck, inguinal and umbilical hernias, polydactyly, rocker bottom feet, distal axial triradius, and elevated fetal hemoglobin. Life expectancy over a year was frequent. Absent were cleft lip and palate, sloping forehead, microphthalmia, and neutrophil drumsticks, common findings in trisomy 13.

E. Dq—and Dr syndromes

Over 60 case reports have been published in which the patient had deficiency of part of the long arm of a D-group chromosome (Dq—) or in which a D-group chromosome was replaced by a ring (Dr) (Lejeune *et al.*, 1968; Gilgenkrantz *et al.*, 1971; Niebuhr and Ottosen, 1973). Although these cases may represent a heterogeneity, there is good evidence to suggest that most involve No. 13 (wilson *et al.*, 1973).

Only a few examples of Dr have been described in which the chromosome has been identified as no. 15 (Jacobsen, 1966; Emberger *et al.*, 1971). The phenotype in these cases was not striking : short stature, mental retardation, and microcephaly. Mean survival has been 39 months for Dq— cases and 89 months for Dr examples (Taylor, 1970).

All patients have exhibited mental and somatic retardation and many have been hypotonic.

Musculoskeletal abnormalities have included bilateral hip dislocation, focal lumbar vartebral agenesis, inguinal hernia, coxa valga, and synostosis of the fourth and fifth metacarpals.

F. Trisomy 18 syndrome (Edwards syndrome)

In 1960, Edwards *et al.* and, almost simultaneously, Patau *et al.* 1961) described a new syndrome associated with the presence of an extra chromosome in the E group which was subsequently shown to be a no. 18 chromosome (Yunis *et al,* 1964).

The most constant features of this syndrome, noted in over 75% of the cases, include : deveelopmental retardation, failure to thrive, feeding difficulties, hypertonia, limited hip abduction, flexion deformities (usually ulnar deviation) of fingers, short sternum, congenital heart disease (ventricular septal defect—90%, patent ductus arteriosus—70%, and atrial septal defect—20%), short

dorsiflexed halluces, rockerbottom feet, calcaneovalgus deformity of feet, and cryptorchidism (Weber and Sparkes, 1970).

Craniofacial anomalies almost always present include prominent occiput, low-set malformed pinnas, and micrognathia. Severe anomalies found at autopsy, apart from the cardiac anomalies noted above include Meckel's diverticulum, heterotopic pancreatic tissue, thin diaphragm with eventration, and various renal anomalies.

Dermatoglyphic alterations are frequent. Over 85% of finger prints are simple arches. Over 30% have a simian palmar crease and over 40% have a single flexion crease in the fifth finger.

Trisomy 18 has an uncommon but yet definite association with aplasia of the radius and thrombocytopenia.

1. Trisommy 18

The incidence of trisomy 18 in trhe more recent surves has varied from 1 per 3500 to 1 per 7000 births (Taylor, 1968, Benady and Harris, 1969; Garfinkel and Porter, 1971). Mean maternal age is elevated, 32 years (Taylor, 1968).

There is a 3 : 1 female predilection caused, in large part, by a greater male fatality rate during the first few weeks of life (Weber, 1967).

The mother often exhibits small weight gain during pregnancy and indicates that fetal movements were feeble. Most examples are postmature. Mean birth witht is less than 2300 gm. The placenta is often small with umbilical artery, and hydramnios has been noted in over 50%.

Thrity percent fail to survive more than 1 month, 50% succemb by 2 months, and less than 10% live more than 1 year. Mean survival time is about 70 days (females—134 days, males—15 days).

2. Double Trisomies

Double primary nondisjunction has been observed in 5—10% of cases (Hamerton, 1971). Mean survival time for douuble trisomies has been 3 weeks. Maternal age is markedly increased in this group.

3. Trisomy 18 Caused by Translocation

Translocation is usually sporadic but examples of familial trnslocation have been recorded (Hamerton, 1971). Means maternal age is lower than for those with trisomy 18 caused by nndisjunction.

G. 18p—syndrome

Deletion of the short arms of chromosome 18 is associated with a variable phenotype. Maternal age is elevated. There is a 2 : 1 female sex predilection (Parker *et al,* 1973).

Mental retardation is a constant feature but of variable degree. Birth weight is low and somatic growth retarded. There is no characteristic facial dysmorphia. Frequently, however, hypertelorism, epicanthal folds, strabismus, and ptosis of lids are noted. The ears are low-set, large, floppy, and poorly formed.

H. Trisomy 21 (Down's syndrome)

Langdon Down (1866) first extensively described the syndrome which has received his name, calling it "Mongoloid idiocy" or "Mongolism." In 1959. Lejeune demonstrated that the disorder was associated with an extra chromosome in the G group. In 1960, Polani et al. described translocation Down's syndrome, and in 1961, clarke et al discovered rnosaicism for an extra G-group chromosome. Yunis et al (1965), by means of autoradiography, identified the chromosome as one of the no. 22 chromosomes, although by this time the term trisomy 21 had been so extensively employed that it has remained.

The incidence of trisomy 21 is between 1 and 2 per 1000 live births among various populations (Mikkelsen, 1971). Over 95% of the cases are caused by nondisjunction, the remainder resulting from translocation.

The skull is brachycephalic with shortening of the anteroposterior diameter and flattening of the occiput in about 75%. The cephalic index (normally 0.75—0.80) is usually greater than 0.80 and may exceed 1.00. (Roche et al., 1961). In infants with trisomy 21, the fontanels are larger than normal and closure is late. In those over 10 years of age, a patent metopic suture is found in 65% of males (normal—9%) and in 40% of females (normal—12%). An extremely common feature (over—90%) is absence of frontal and spehenoid sinuses and hypoplasia of the maxillary sinuses (Spitzer *et al.*, 1961, Betlejewski *et al.* 1964). There is poor development of the bones of the middle face, producing a relative prognathism and ocular hypotelorism (Gerald and Silverman, 1965).

The profile is flattened owing to hypoplasia of the nasal bones. The palpebral fissures are oblique, the outer canthus being slightly higher than the inner. Epicanthal folds are extremely common. Speckled iris (Brushfield's spots) and lens opacity are present in about 85% and 60% of patients, respectively.

Various other anomalies include missing or malformed teeth (especially maxillary lateral incisors and mandibular second premolars), delayed eruption, increased periodontal destruction, and malocclusion (Cohen and Cohen, 1971).

The hands are characteristically short and broad, the fifth finger usualy being abbreviated and clinodactylous, and having a single flexion crease in about 20% of the cases. There is usually greater space than normal between the hallux and the rest of the toes.

Hypotonia, especially marked in infancy, improves with age. loints are usually hyperextensible. The penis and scrotum are usually small and about

25% have cryptorchidism Public hair is straight. Congenital cardiac anomalies is present in about 40% Down Syndromic persons.

Diastasis recti, duodenal atresia, or umbilical hernia occur in about 10% (Butterworth *et al.*, 1964).

Radiographic changes include reduced iliac and acetabular angles in the young infant (Nicolis and Sacchetti, 1963) and hypoplastic middle phalanx of the fifth finger.

Intelligence quotients range from 25 to 70, most Down's syndrome patients 3 years of age or less having I.Q. s of 50—59 but slipping with increasing age to 25-49 (Penrose and Smith, 1966).

Dermatoglyphic anomalies include distal axial triradius in the palm (over 80%), bilateral simian creases (30%), single flexion crease in fifth finger (20%), 10 ulnar loops (30%), and hallucal arch tibial (70%) or small loop distal (30%) patterns (Preus and Fraser, 1972).

Because of susceptibility to respiratory infection, early mortality used to be great. With the introduction of antibiotics, the means survival age is almost 20 years. There is a twentyfold increased association with acute leukemia (Conand Erkman. 1966).

Numerous attempts have been made to established specific biochemical alterations. However et aL, (1965) found decreased blood serotonin and increased galactose phosphate uridyltransferance leukocyte alkaline with Down's syndrome.

1. Trisomy G_1 due to Primary Nondisjunction

As discussed above, about 95% of cases of Down's syndrome are sporadic primary trisomics, resulting from nondisjunction which is age dependent. This occurs at the first meiotic division in the mother (Robinson, 1973). If the mother is less than 20 years of age at time of conception, the risk of producing a child with trisomy 21 is about 1 per 2500 live births. This risk gradually increases until 35 years, after which there is a more marked increase in frequency such that a mother over 45 years has about 1 chance in 50 or less of having a child with Down's syndrome.

2. Association of Down's Syndrome with Other Primaiy Nondisjunctions

Individuals with trisomy 21 have been occasionally (about 1 per 200) found to have another extra chromosome (double primary nondisjunction), the most frequent type being 48,XXY, G+ (Hamerton *et al,* 1965; Taylor and Moores, 1967). Other forms such as 48,XXX,G+ and 48,XYY,G+ have also been described (Yunis *et al,* 1964, Uchida *et al.*, 1966). This association is much higher than might be expected by chance.

3. Translocation Down's Syndrome

Down's syndrome patients born to young mothers as well as those with affected relatives often have the extra G] chromosome attached to another chromosome. This has been designated translocation and comprises about 3.5% of cases of Down's syndrome. It may be sporadic or familial. Translocation Down's syndrome is not age dependent. About 8% of Down's syndrome patients born to mothers less than 30 years of age have exhibited translocations as opposed to 1.5% born to mothers over 30 years old. It is widely accepted that the short arms of acrocentric chromosomes have nucleolar organizers and that these points are likely to break, producing a high frequency of structural chromosome aberrations.

In familial translocation Down's syndrome, one of the parents has 45 chromosomes instead of the normal 46. One of the small G-group chromosomes is "missing" since it has been translocated to another chromosome. The parent carrying the translocation chromosome is phenotypically normal, since no significant amount of genetic material has been lost in the translocation process. In most cases, the extra G(chromosome is phenotypically normal, since no significant amount of genetic material has been lost in the translocation process.

4. Down's Syndrome Mosaicism

Patients having two different cell populations, one trisomic for chromosome G; and another normal, constitute about 2—3% of patients with Down's syndrome. This condition is usually suspected when the phenotypic expression of trisomy 21 is not fully expressed or when the intelligence of the patient is higher than expected. In addition, they may have children with Down's syndrome (Weinstcin and Warkany, 1963). Individuals having trisomy Gt mosaician may vary in phenotype from typical trisomy 21 to normal. There is no age dependency (Richards, 1969). One cannot correlate the percentage of trisomic blood cells with intelligence. Richards (1969) found about 20% more trisomic cells in fibroblasts than in lymphocytes.

If mosaicism is found in one of the parents of a child with Down's syndrome, meiotic study of ovary or testis should be carried out. There is evidence that if half the cells are abnormal, about 25% of the childern will have Down's syndrome (Mikkelsn,1971a).

I. Nonmongoloid "Trisomy G"

Several cases of nonmongoloid "trisomy G" have been published (Uchida et at, 1968; Al-Aish, 1969; Lozzio, 1969; mikkelsen, 1969). Some have been designated as having trisomy 22 to contrast with trisomy 21 (Down's syndrome). At this point in time, within this group, with two possible exceptions cited

below, there seems to be no characteristic phenotype and it would appear likely that some of these represent centric fragments that may come from several different chromosomes.

J. G deletion syndromes

There are at least two relatively distinct phenotypes presumably representing monosomy or deletion of a portion of the long arm of two different G-group chromosomes (Warren *et al*., 1973).

1. The G_1 Deletion Syndrome (Antimongolism)

This syndrome consists of mental and growth retardation, hypertonia, nail anomalies, skeletal malformations, cryptorchidism, hypospadias, inguinal hernia, pyloric stenosis, thrombocytopenia, eosinophilia, and hypogammaglobulinemia. Facial and oral manifestations include microcephaly, large low-set ears, antimongoloid obliquity of palpebral fissures, highly arched or cleft, and micrognathia. Dermatoglyphic analysis has shown a marked increase in radial loops (Schindeler and Warren, 1973).

2. The G_1 Deletion Syndrome

This syndrome has less distinctive features: severe mental retardation, hypotonia, soft tissue syndactyly of the second and third toes, and clinodactyly of the fifth finger. Facial and oral manifestation include large, low-set ears, epicanthal folds, ptosis of eyelids, highly arched palate, and bifid uvula 1971; Stoll *et al.*, 1973). Dermatoglyphc analysis has shown a marked increase in whorls, a decrease in ulnar and radial loops, a distal axial triradius, and hypothenar patterns (Schindeler and Warren, 1973).

K. Triploidy

Triploidy, as noted later in this chapter, is a frequent cause of fetal wastage prior to the eighth intrauterine week. Diploid/triploid mosaicism is occasionally compatible with survival and there have even been several examples of pure triploidy.

5.3.3 Sex chromosomal abnormalities

A. Klinefelter syndrome

In 1942, Klinefelter *et al*., described a syndrome in post pubertal males consisisting of small firm testes with tubulan by a linization but with a normal number of Leydigs cells, aerosposrmia, gynecomaztia, elevated urinary gonadotropies and low concentrations of urinary 17-ketosteroids. Several years later, Bradleury etal. (1956) and plunkeft and Barr (1956) noted Chromotin-positive nuclei in the tissues of such patients, and Jacob and strong (1959) described an

XXY sex chromosome complement in chromotin-positive klinefelters syndrome.

Chromatin-positve males have been found to comprise about 2 per 1000 live male births. These also contains XXY, XXYY, XY/XXY, and other rarer forms of Klinefelter syndrome. It has been estimated that about 80% are XXY, 10% are mosaics, and the rest are XXYY and the more unusual types.

1. XXY Klinefelter syndrome

The clinical features of XXY klinefelter syndrome do not become apparent until after puberty. Body proportions usually do not appear remarkably abnormal, however, the lower extremities tend to be and about 60% have a span that exceeds their height by 3 cm or more.

The Prepubertal testes are normal size and microscopic appearance but during adolescence they fail to enlarge and remain small and firm, averaging less than 2cm in length. The seminiferous tubules are usually shrinken, hyalinized, and irregularly arranged. Elastic fibres are absent around the tunica propria of the tubules. Ledig cells are clumped, Rarely spermatogesesis can be demostrated. In nearly all cases the testes descend. The penis is usually of normal size but may be somewhat shorter than normal. The prostate is smather than normal. Gynecomastia develops after puberty in about 50% and facial hair is sparse in about 60-75%. Axillary hair may also be less.

The prepubertal testes are of normal size and microscopic appearance but during adolescence they fail to enlarge and remain small and firm, averaging less than 2 cm in length. The seminiferous tubules are usually shrunken, hyalinized, and irregularly arranged. Those tubules which are not sclerotic are immature and lined exclusively with Sertoli cells. About 50% have a female pubic pattern. Libido and potency are usually decreased. There is some evidence of increased tendency to pulmonary disorders, varicose veins, and, possibly, breast cancer. There is the same frequency of color blindness among XXY patients as in normal females. Although intelligence may be reduced, at least 75% of XXY males have normal intelligence. Personality is usually passive.

The incidence of XXY Klinefelter's syndrome is about 1.3 per 1000 live male births. Meternal age is significantly increased for XXY but not for XXYY,XXXY, or XXXXY patients. About 60% of XXY males are X^MX^MY, while 40% are X^MX^PY. The X^MX^MY state arises from nondisjunction either during oogenesis or at an early postzygotie division. The X^MX^PY condition probably has its origin in nondisjunction during the first meiotic division.

2. XXYY Klinefelter syndrome

Patients with the XXYY variant tend to be about 4 cm taller than average height, more aggressive, and more mentally retarded than those with XXY Klinefelter's syndrome (Borgaonkar *et al,* 1970). Otherwise the phenotype is quite

similar: small firm testes, eunuchoid body build, sparse, body hair, gynecomastia, and elevated gonadotropins. Almost all XXYY males described to date have been mentally retarded and many have been aggressive (Schlegel *et al.*, 1965).

As mentioned above, there is no increase in parental age in contrast to XXY Klinefelter's syndrome. The disorder is most likely due to nondisjunction in both the first and second meiotie divisions during spermatogenesis with production of an XYY sperm. One cannot,however, rule out the less likely possibility of nondisjunction at the second meiotie division in both parents.

Dermatoglyphic studies have shown that digital arch patterns are more common in the XXYY patient than in the XXY individual who, in turn, has more than the normal male.

A child with an XXYYY sex chromosome complement was noted to have mental retardation, lordosis, flexed index and fifth fingers, pes planus, and aggressive personality (Gracey and Fitzgerald, 1967).

3. XXXY Klinefelter's syndrome

Over 25 cases of XXXY Klinefelter's syndromoe have been published (Vormittag and Weninger, 1972). All have been mentally retarded. The phenotype is similar to that of the XXY male but the size of the penis is small (McGann *et al.,* 1970).

Two late-labeling X chromosomes have been demmonstrated. However, two Barr bodies are seen in only a proportion of cells (Vormittag and Weninger, 1972). The condition may arise from successive nondisjunction in either the maternal or paternal meiotie divisions (Pfeiffer and Sanger, 1973). Dermatoglyphic findings have not been consistent. An XXXYY male has been described by Bray and Josephine (1963).

4. XXXXY Klinefelter's syndrome

There have been over 70 cases of 49,XXXXY males published since Fraccaro and Lindsten documented the first example in 1960. Nearly all have been severely mentally retarded intelligence quotients ranging from 20 to 60. A marked difference between the XXY and XXXXY male is thepoordevelopment of the external genitalia in the latter. The penis is always minute and the testes very small and undersended with hypoplastic Leydig cells and absence of term cells. The scrotum is. usually hypoplastic.

Mild nicrocephaly, ocular hypertelorism (90%), myopia (25%), strabismus (50%), mild nongoloid obliquity of palpebral fisures (35%), epicanthus (80%, and short neck with redundant skin on the nape have been noted. Skeletal anomalies present in over half the cases. Congenital heart disorders have been noted in about 20% of cases. Gonadotropins have not been elevated.

Autoradiographic evidence has shown three heavily labeled X chromosomes

(Hsu and Lockhart, 1965). Three Barr bodies may be found in a proportion of interphase nuclei (Miller and Warburton, 1968).
Parental age is not elevanted. Postzygotic nondisjunction in an XXY zygote appears to be the cause for the XXXXY state, all the X chromosomes coming from the mother (Murken and Scholz, 1967; Race and Sanger, 1969).

5. XX Klinefelter syndrome

Less than 30 cases have been published of males having 46,XX karyotypes. They exhibit many of the stigmata of Klinefelter's syndrome and, sence, will be considered here (Anderspn *et al.*, 1972).

All have small testes, are infertile, and rarely shave. About 70% have gynecomastia and elevated gonadotropin levels. Plasma testosterone levels are very low (Neuwirth *et al.*, 1972). The penis and scrotum have been small in about half the cases. All are of normal intelligence and have normal skeletal proportions.

6. Klinefelter's syndrome Mosaicism

About 15% of patients with Klinetelter's syndrome have been found to have two or more chromosomally distinct cell populations. In each of these individuals, one of the cell populatins generaly has an XXY, XXXY, or XXXXY sex chromosome constitution while the other is XX or XY. The clinical expression of mosaicism for Klinefelter's syndrome depends on the type of sex consitution present at a critical time of development. Thus, one can find, for example, and XY/XXY mosaic who is phenotypically normal, provided the XY cells exerted the predominant genetic effect.

In a study of XY/XXY mosaics, Gordon *et al.* (1972) found that only one-half exhibited azospermia and about one-third had gynecomastia and elevated gonadotropins. About one-quarter had germinal epithelium. Among 6 patients with XX/XXY mosaicism, Ferguson-Smith (1969) noted comparable findings.

B. XYY syndrome

Although the presence of an extra Y chromosome had been described as early as 1961 (Sandberg et al., 1961), interest was markedly aroused by a finding of a disproportionately high percent (usually 2—4%) of such individuals in prisons and mental hospitals (Casey *et al,* 1968; Jacobs *et al.,* 1968; Marinello *et al.*, 1969; Hook 1973). It was soon noted that most XYY patients are excessively tall and not uncommonly midly mentaly retarded (mean intelligence quotient —90) (Valentine et al., 1971). However, the frequencey of the condition among newborn male infants is about 1 per 700 births (Ratcliffe *et al.,* 1970) and few of these individuals lead other than quite routine lives. The adult height of an XYY individual is usually over 180 cm while XYY children are usually above the 90th

percentile in height by 6 years of age. Leg length and trunk length are increased but the leg/trunk ratio is normal (Keutel and Dauner, 1969). Muscle weakness (especially of the pectoralis major) and poor coordination are commonly noted. Phenotypical alterations are subtle: mild facial asymmetry, mild pectus excavatum, and mild scapular winging. The ears tend to be long and often there is a bony chin point. Most have exhibited normal sexual development (Court Brown, 1969). There are no characteristic dermatoglyphic aherations (Hubbell *et al,* 1973).

The disorder probably arises from paternal nondisjunction during the second meiosis. Retarded intelligence (I.Q.—70), impulsive aggressive behaviour, bilateral simian creases, clinodactyly of fifth fingers, retarded bone age with pseudoepiphyses at the bases of the metacarpals and metatarsals, and lack of patellar epiphyseal calcification were described by Schoepflin and Centerwall (1972).

Ridler *et al.* (1973) noted low normal intelligence, behavior problems with aggressive outbursts, repeated pulmonary infections, hypotrophic testes, sparse body hair, and acne in a 48,XYYY patient. Conversely, Hunter and Quaife (1973) decribed no stigmate other than sterility.

C. Turner's syndrome

In 1938, Turner described a syndrome in postpubertal female^consisting of sexual infantilism, short stature, webbed neck, and cubitus valgus. Albright *et al.* (1942) showed that these patients had an elevated urinary excretion of gonadotropins, and Wilkins and Fleischmann (1944) described "streak" gonads devoid of ovarian folicles in such cases. Polani *et al.*, (1954) and Wilkins *et al.*, (1954) demostrated that most cases are chromatin-negative, and Ford *et al.* (1959) first described the XO karyotype.

Turner's syndrome has been estimated to occur 1 per 2500 female births (Maclean *et al.*, 1964; Mikamo, 1968) and has been frequently noted in abortuses. Parental age is not increased.

Variation in phenotype has led to some confusion concerning nomenclature. Since the most common features are short stature, streak gonads, and X monosomy or short arm loss of X chromosomal material, all patients with these features are classified here as examples of Turner's syndrome (Yunis, 1965). Deletion of the long arm of the Y chromosome has occasionally been associated with the Turner syndrome. Cases with streak gonads and sexual infantilism but of normal or increased suture and normal female or male sex chromosome complement will be referred to as having "pure gonadal dysgenesis" or, more accurately, XY or XX gonadal dysgenesis.

Primary amenorrhea and sterility are almost constant features of the XO Turner syndrome although exceptions have been noted. Breast development is

poor, the chest is broad with seemingly widely spaced, hypoplastic, at times, inverted nipples. The external genitalia are infantile and public hair is sparse.

The histological pattern of the dysgenetic gonad found in Turner's syndrome consists of long streaks of white wavy connective tissue stroma without follicles. Follicles are present, however, in fetal and infantile ovaries of patients with Turner's syndrome (Gordon and O'Neill, 1969).

Adult height is usually less than 57 inches (144 cm). Various skeletal anomalies include cubitus valgus (about 75%), short fourth metacarpals (about 65%), deformity of medial fibial condyle (about 40%) osteoporosis (about 50%), hypoplasia of first cervical vertebra (about 40%), and small carpal angle.

Birth weight is below the 3rd percentile in about half the cases. In infants, excess skin on the nape and peripheral lymphedema have been noted in 15—50% of the cases. During embryonic life, neck blebs or cystic hygroma are common (Singh and Carr, 1966; Rushton *et al.,* 1969). Toenails are frequently hypoblastic. With age, the excess skin on the nape metamorphoses into pterygium colli and, with improvement in deep lymphatic circulation, the peripheral lymphedema gradualy disappears. Increased numbers of cutaneous nevi are found in about 60%.

Epicanthal folds, ptosis of upper eyelids, prominent ears, and micrognathia are common facial features. The hairline is low at the nape.

Thyroid antibodies are elevated in XO Turner's syndrome but less frequently than in the X-iso X mosaic and glucose intolerance occurs with grater frequency in patients with Turner's syndrome and in their parents than in the normal population (Rimoin, 1973).

1. X Monosomy

Patients with an XO sex complement appear to comprise about 60% of the cases of Turner's syndrome. Furthermore, they appear to be more severely affected clinically than other forms of the disorder.

2. XO/XX and XO/XXX Mosaicism

Patients with Turner's syndrome may have two different cell populations, one having an XO sex constitution, the other a normal XX sex complement. Such individuals are called XO/XX mosaics and constitute about 7% of the cases of Turner's syndrome. The two cell population types may appear in every tissue of the body or only in certain ones.

Presumptive evidence for mosaicism lies in a discrepancy between sex chromatin pattern and karyotype, or through observing a low percentage of chromatin-positive nuclei (5—15%) in phenotypic females. The clinical spectrum of XO/XX mosaicism is wide and may vary from cases quite typical of Turner'

syndrome with many associated anomalies to cases with normal gonads and normal stature. About 20% menstruate (Ferguson-Smith, 1969). In contrast to patients with XO Turner's syndrome who are prone to aortic coarctation, those with XO/XX karyotypes are likely to have pulmonic stenosis with or without atrial spetal defect (Nora *et al.*, 1970), being similar to patients with Noonan's syndrome.

The usually accepted explanation for XO/XX mosaicism is loss of an X chromosome during cleavage in the early embryo.

About 5% of the cases of Turner's syndrome are XO/XXX mosaics. Clinically they resemble the XO/XX mosaic. Patients having three stem lines XO/XX/XXX have been reported but are quite similar phenotypically to XO/XX mosaics.

3. Isochromosome X (XXqi)

About 20% of patients having Turner's syndrome have an X isochromosome, i.e., replication of the long arm of the late replicating X chromosome. They exhibit short stature, sexual infantilism, primary amenorrhea, and skeletal anomalies.

The Barr body and polymorphonuclear neutrophils drumsticks are larger than normal. Drumsticks are also more nuemerous (Taft *et al.*, 1965; Sparkes and Motulsky, 1967).

4. Short and Long Arm Deletion of an X chromosome

Deletion of the short arm of an X chromosome (XXp—) results in the Turner phenotype. They are as short as individuals with the XO Turner syndrome but are less likely to have associated malformations (Atkins *et al.*, 1965).

Deletion of the long arm of an X chromosome (XXq—) is far less likely to be asociated with short stature. The girl was 159 cm tall and exhibited no stigmata of Turner's syndrome. She never menstruated and her ovaries were not palpable. As expected, her Barr bodies were smaller than normal. Xg1 studies showed that the Xpi was of maternal origin. However, Turner's syndrome has been reported in association with XXq—.

5. Y deletion

At least a dozen cases of Turner's syndrome associated with a dicentric Y Chromosome have been published (Armendares *et al.*, 1972; Cohen *et al.*, 1973). All have short stature, female phenotype, and most have associated anomalies. A patient with long arm isochrorriosome Y had drumstiicks have been noted in polymorphonuclear leukocytes.

5.4 Human genome

5.4.1 The Human Genome Project

As the recombinant DNA, gene cloning and DNA sequencing technologies improved in the 1970s and early 1980s, scientists began discussing the possibility of sequencing all 3 x 109 nueleotide pairs in the human genome. These discussions led to the lauching of the **Human Genome Project** in 1990. The initial goals of the Human genome project was to construct a detailed physical map of the entire human genome, and to determine the nueleotide sequences of all 24 human chromosomes by the year 2005. Scientists soon realized that this huge undertaking should be a worldwide effort. Therefore, an international **Human Genome Organization** (HUGO) was organized to coordinate the efforts of human geneticists around the world.

James Watson, who with Francis Crick, discovered the double-helix structure of DNA, was the first director of this ambitious project, which was expected to take nearly two decades to completle and to cost in excess of $3 billion. In 1993. Francis Collins, led the research teams that identified the cystic fibrosis gene, replaced Watson as director of the Human Genome Project. In addition to work on the human genome, the Human Genome Project has served as an umbrella for similar mapping and sequencing projects on the genomes of several other organisms, including the bacterium *E. coli*, the yeast *S. cerevistae*, the fruit fly D. *melanogaster*, the plant A. *thaliana*, and the worm *C. elegans.*

5.4.2 Bacterial Genomes

In 1995, *Haemophilus influerzae* was the first bacterium to have its genome sequenced in its entirety. By mid-2001, the complete sequences of 32 bacterial genomes were abailable in the public databases (collections of the sequences of genes, chromosomes and genomes). The genomes range in size from 580,070 bp for *Mycoplasma genitalium*, which is thought to have the smallest geneme of any self-replicating organism, to 4,411,529 bp for *Mycabacterum tuberculosis*, which causes more human deaths than any other infections bacterium, to 4,539,221 bp for *Escherichia coli*, the best-known cellulalr microorganism. The genome of *M. genitalium* is of special interest because it may aproximate the "minimal gene set" for a self-replicating organism—the smallest set of genes that will alow an organism to reproduce itself. Of course, the genome of *M. tuberculosis* is of great interest becauseof the pathogenicity of this organism and the hope that a complete understanding of its metabolism will suggest ways to prevent tuberculosis in humans: The need for new ways to combat this pathogen has been enhanced by the recent evolution of antibiotic-resistant strains of *M. tuberculosis*.

Of the bacterial genomes sequenced to date, the genome of *E. coli* (Fig. 5.9)

has undoubtedly caused the most excitement among geneticists. *E. coli* is the most planet.Geneticists, biochemists, and molecular biologists have utilized *E. coli* as the prefered model organism for decades. Most of what is known about bacterial genetics was learned from research on *E. coli.*

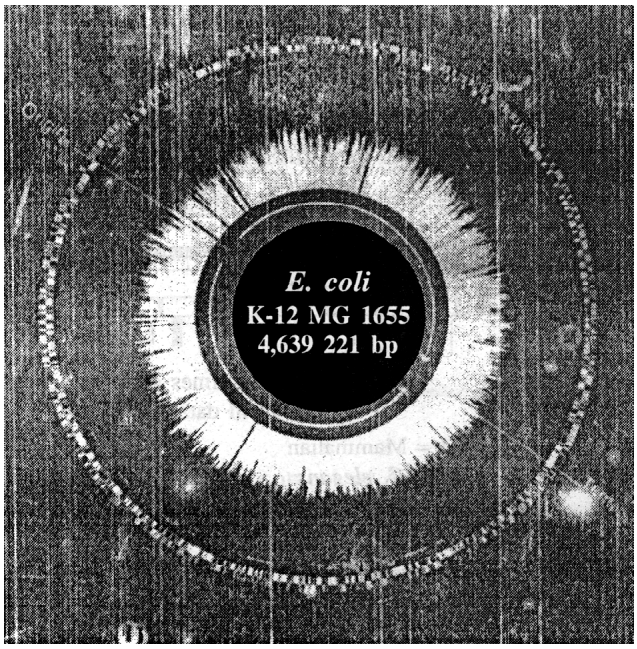


Fig. 5.9 Sequence-based map of the chromosome of *Escherichia coll.* The blue arrows mark the halves of the chromosome travered by the two replication forks. The outer concentric circle gives the position of genes encoding proteins similar to bacteriophagc proteins. The second concentration circle shows the location of genes that are transcribed clock-wise (gold) from one strand or counterclockwise (yellow) from the complementary strand. The sunburst in the center is a histogram in which the length of each ray is proportional to the randomness of codon usage within each coding sequence

The *E. coli* genome contains 4288 putative proteincoding sequences or genes. About one-third of these are well-studied genes encoding known products, whereas 38 percent are of unknown function, Putative protein-coding sequences that are not knwn to encode proteins are called **open reading frames** or **ORFs**. An **ORF** is a nucleotide sequence that begins with a translation-initiation signal (usually ATG). continues with a sequenced of base triplets specifying amino acids, and ends with one of the three translation-termination singnals.

The average distance between genes (size of intergenic regions) in the E. coli genome is 118 bp. Known and putative genes specifying proteins and stable RNAs make up 87.8 percent and 0.8 percent of the genome,respectively, and noncoding repetitive elements account for 0.7 percent of the genome. Thus, 10.7 percent of the genome must involve regulatory sequences and sequences with unknown functions.

Once the complete sequence of a bacterial genome is available, it can be searched using computers for similarities with other sequenced genomes. Such sequence comparisons can often be used to gain inferences about gene function. Because so much is known about gene function in *E. coli*, comparisons with other secuenced bacterial genomes are often very informative. For example, a comparison of the genomes of *Treponema pallidum*, the parasitic spirochete that causes syphilis, and *E. coli* shows that *T. pallidum* contains the genes that encode proteins involved in DNA replicartion and repair, transcription, and translation, but carries few genes encoding biosynthetic enzymes and trtanspoort protiens.

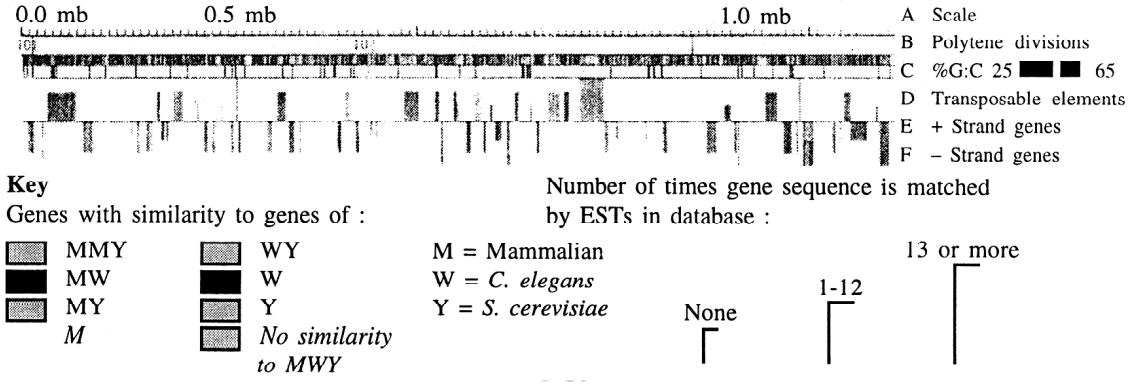


Fig. 5,10 Sequence-based map of chromosome 4 of *Drosophila.* The top line (A) gives map position in megabase paris (1 mb = 1 million hase pairs). The second panel (B) gives the polytene chromosome band number. The third panel (C) shows the percetage of G:C base pairs, and the fourth (D) shows the positions of transposable genetic elements. The bottom two panels show the positions of genes where transcription occurs with the plus strad as template (F) or with the minus strand as template (F). The color of each gene box in panels F and F indicates its similarity to genes of mammals, *C. elegans,* or *S. cerevisiae,* as shown by the key below the map, and the height of each gene box in panels E and F indicates the frequency of the sequence in the expressed-sequenced tag (EST) database for *Drosophila*

Transposable genetic elements make up about 10 percent of the *Arabidopsis* genome, far less than the 50 to 80 percent of the corn genome estimated to be derived from transposable elements.

The next challenge is to determine the functions of the *Arabidopsis* gene products. There is where this little weed really shines as a model system. It is idealy suited for genetic dissections of biological processes. The goal or the *Arabidopsis* research community is to determine the function of all 25,498 genes in the nect 10 years.

5.4.3 The Human Genome

Recall that the initial goals of the Human Genome project were (10 to map all the human genes, to construct detailed physical maps of all 24 chromosomes, and to sequence the entire genome by 2005. All of these goals are likely to be achieved ahead of schedule. With two first-draft sequences of the human genome already published in February 2001 a complete sequence of the euchromatic portion of the human genome will certainly be available long before 2005.

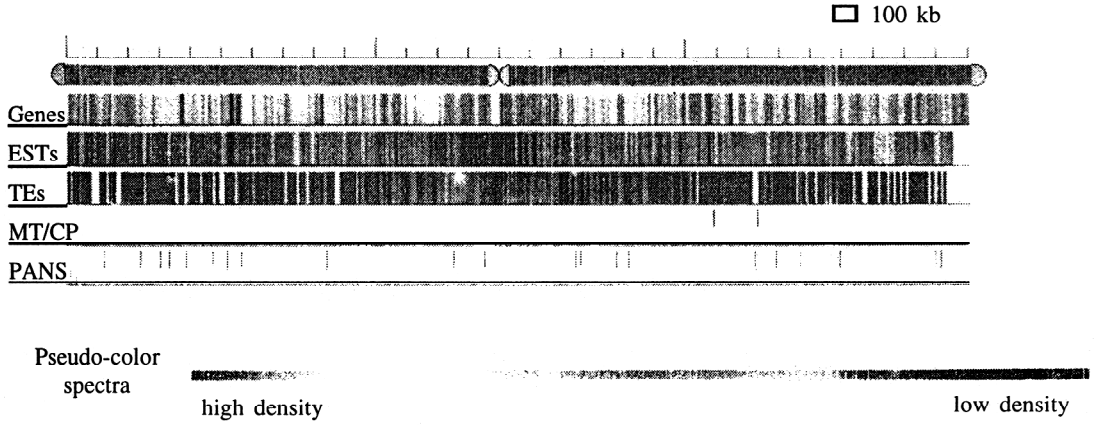


Fig. 5.11 Sequence-based map of chromosome 1 of *Arabidopsis.* Distance in 100-kb units is shown at the top. The chromosome is represented by the top bar, with sequenced regionsin red and the unsequenced centromere and telomerees in blue. The densities of genes (top panel below chromosome), matches to expressed sequence tags (ESTs; second panel), transposable small nuclear RNAs (bottom panel) are color coded, with red representing the highest density and dark blue the lowest density

Progress in mapping the human genome has been excellent. Complete physical maps of chromosomes Y and 21 and detailed RFLP maps of the X chromosome and al 22 autosomes were published in 1992. By 1995, the genetic map contained markers separated by, on average, 200 kb. A detailed micro satellite map of the human genome was published in 1996, and a comprehensive map of 16,354 distinct loci was released in 1997. All of these maps have proven invaluable to researchers cloning genes based on their locations in the genome.

Unfortunately, the resolution of genetic mapping in humans is quite low—in the range of 1-10 mb. The resolution of fluorescent *in situ* hybridization (FISH) is also approximately 1 mb. Higher resolution mapping (down to 50 kb) can be achieved by **radiation hybrid maping,** which is a modification of the somatic-cell hybridization mapping procedure. Standard somatic-cell hybridization involves the fusion of human cellls and rodent cells growing in culture and the correction of human gene products with human chromosomes retained in the hybrid cells.

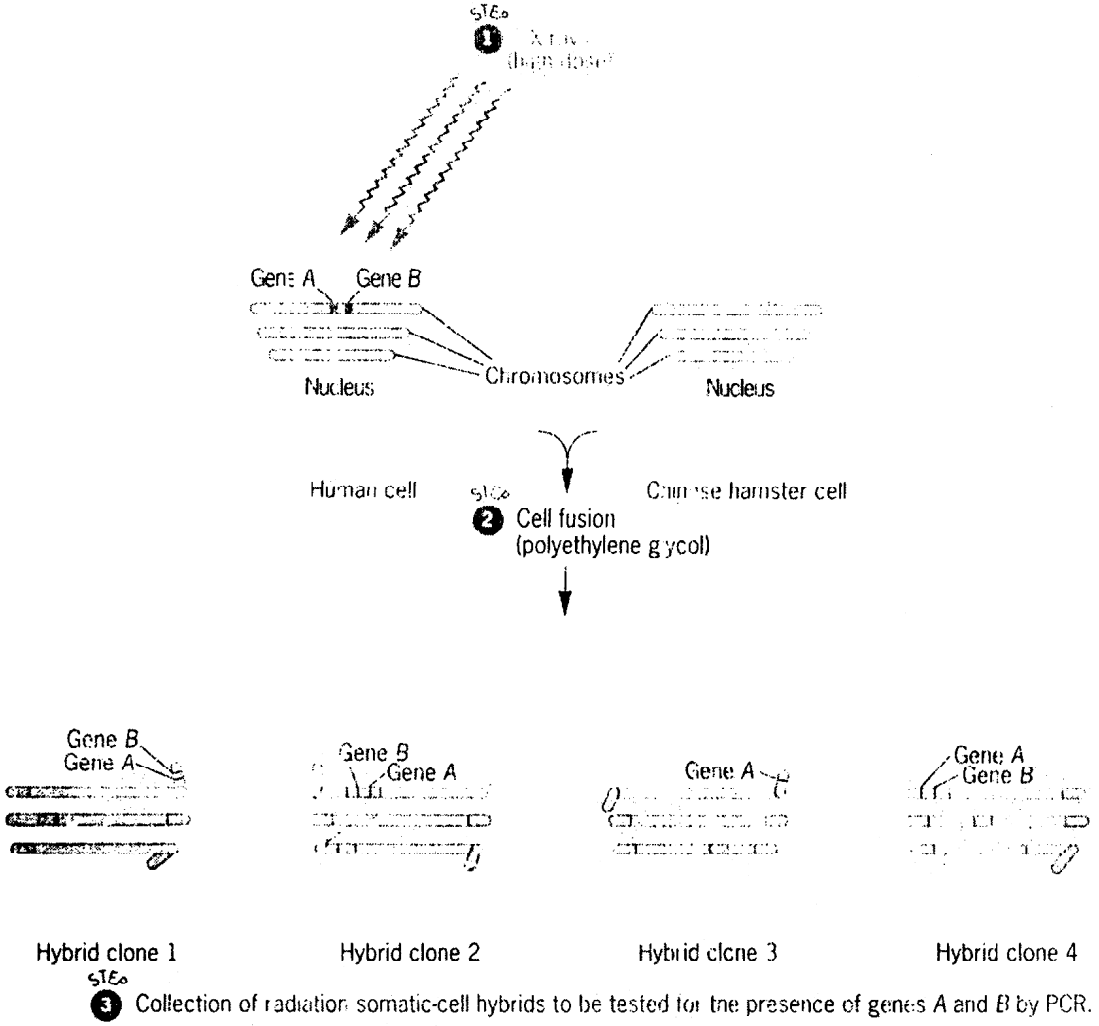


Fig. 5.12 The use of radiation somatic-cell hybrids for high-resolution mapping of the human genome. The rationale behind radiation hybrid maps in that the probability of an X ray-induced break between genes *A* and B is directly proportional to the distance between them on the DNA molecule. Note that genes *A* and *B* have remained together in hybrid clones 1, 2, and but were separated by an X ray-enduced break during the formation of hybrid clone 3

Radiation hybrid mapping is done by fragmenting the prior to cell usion (Fig. 5.12). The irradiated human eels are then fused with Chinese hamster (or other rodent) cells growing in culture,usually in the presence of a chemical such as polyethylene glycol to increase the efficiency of cell fusion. The human—Chinese hamster somatic-cell hybrids are then identified by growth in an appropriate selection medium.

Many of the human chromosome Chromosome 1 fragments become integrated into the Chinese hamster chromosomes during this process and are transmited to progeny cells just like the normal genes in the Chinese hamster chromosomes The polymerase chain reaction is the used to screen a large panesl of the seleted hybrid cells for the presence of human genetic markers. Chromosome maps are constructed based on the assumption that the probability of an X ray-induced break between two markers is directly proportional to the distances separating them in chromosomal DNA.

Several groups have used the radiation hybrid mapping procedure to construct high-density maps of the human genome. In 1997, Elizabeth Stewart and coworkers published a map of 10,478 STSs based on radiation hybrid mapping; their map of human chromosome 1 is shown in Fig. 5.13.

Whereas the gene maping work advanced quickly, progress towards sequencing the human genome initially lagged behind schedule. However, that all changed rapidly beginning inl998. During May of 1998, J. Craig Venter announced that he had formed a private company, Celera Genomics, with the goal or sequencing the human genome in just three

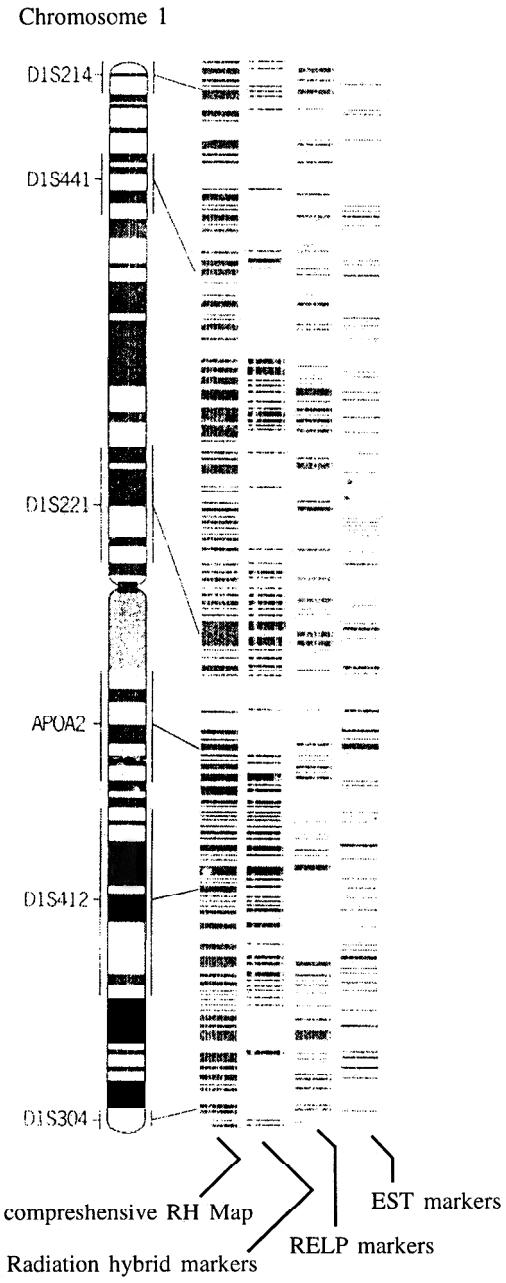


Fig. 5.13 A high-resolution radiat on hybrid map of human chromosome 1. The cytogenetic map of chromosome is shown on the locations of the comprehensive radiatipon hybrid map showing all the markers (red lines), the high confidence radiation hybrid markers (blue lines), the high confidence radiation hybrid markers (blue lines), the highconfidence radiation hybrid markers (blue lines), the high confidence radiation hybrid markers (blue lines), the RFLP markers (green lines), and the ESTs (parple lines)

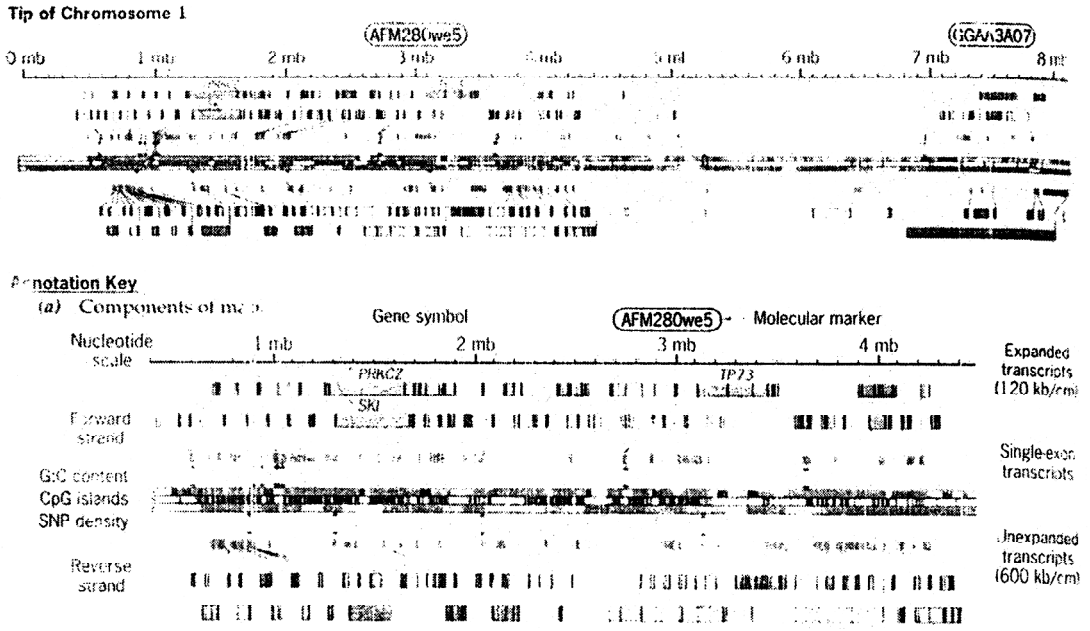


Fig. 5.14 A nnortated, sequence-based map of an 8 mb segment of DNA at the tip of human chromosome 1, assembled by rescarel ers at Celera Genomics. The top line gives distances in mb. The next three panels show predicted transcripts from one strand of DNA (the "forward strand"), whereas the bottom three panels show predicted transcripts specified by the other strand of DNA (the "reverse strand"). The middle three panels give the G:C content, the positions of CpG islands, which occur upstream of genes, and the density of single nucleotide polymorphisms (SNPs), respectively. The annotation key below the map of chromosome 1 shows the components of the map, the color code for gene product functions, and the color codes for G:C content and SNP density

years. Shortly thereafter, the leaders of the public Human Genome Propject's sequencing laboratories announced that they had revised their schedule and planned to complete the sequence of the human genome by 2003—two years earlier than originally poroposed. From this point in time, everything accelelrated.

The complete sequence of the first human chromosome—small chromosome 22—was published in December 1999. The complete sequence of human chromosome 21 followed in May of 2000. Then, with the interventionof the White house, Venter, of Celera Genomics, and Francis Collins, Director of thepublic Human Genome Project, agreed to publish first drafts of the sequence of the human genome at the same time. The Celera and public sequences were both published in February 2001. Figure 5.14 shows an annotated, sequence-based map of an 8mb segment at the tip of the short arm of human chromosome 1. This map ilustrates the positions and orientations of known and predicted genis in one small portion of the human genome. For summar maps of the entire human genome, see the February 15,2001, issue of *Nature* and the February 16, 2001, issue of *Science.*

The amount of information in these first drafts of the human genome was quite overwhelming including the sequence of over 2650 megabase pairs of DNA (over 2,650,000,000 bp). The human genome is more than 25 times the size of the previously sequenced Drosopbila and Arabidopsis genomes, and it is eight times the sum of al previously sequenced genomes.

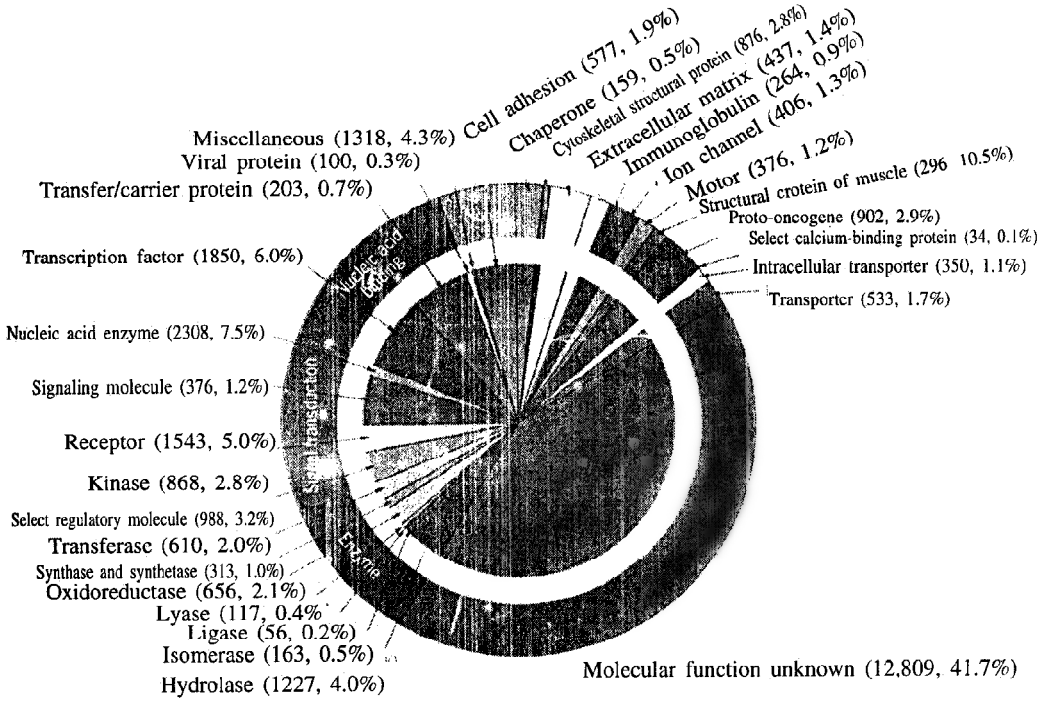


Fig. 5.15 Functional classification of the 26,283 genes predicted by Celera Genomics' first draft of the sequence of the human genome, Each sector gives the number and percentage of gene products in each functional class in parentheses. Note that some classes overlap: a proto-encogene, for example, may encode a signaling molecule

The sequence of the human genome provided one surprise: there appear to be only about 30,000 to 35,000 genes rather than the estimated 50,000 to 120,000 genes suggested by earlier studies. The distribution of functions for the 26,383 genes predicted by the Celera sequence is shown in Figure 5.15. About 60 perecent of the predicted proteins have similarities with proteins of other species whose genomes have been sequenced (Figure 5.16). Over 40 percent of the predicted human proteins share similarities with *Drosopbila* and *C. elegans* Proteins. The picture is quite different for mamilies of closely related proteins, which tend to perform important basic cellular functions. Only 94 of 1278 protein families predicted by the sequence of the human genome are specific to vertebrates. The rest have evolved from domains of proteins in distant ancestors, including prokaryotes and unicellular eukaryotes.

On average, there is one gene per 60 to 85 kh in the human genome, although there is some clustering of highly expressed genes in euchromatic regions of specific chromosomes. Exons make up only 1.1 percent of the genome, whereas inttrons make up 24 percent, with 75 percent of the genome being intergenic DNA. Of the intergenic DNA, at least 44 percent is derived from transposable genetic elements. The initial drafts of the human genome are far from complete, and the immediate goal will be to fill in the gaps in the genome ad produce a finished sequence in the next year or so. The other major goal is to determine the structure and function of the human proteome (all of the proteins encoded by the human genome).

Knowledge of the nucleotide sequences of entire genomes has spawned the development of a new scientific discipline, *bioinformatics*, a fusion of computer science and biology, with the goal of developing new tools with which to analyze

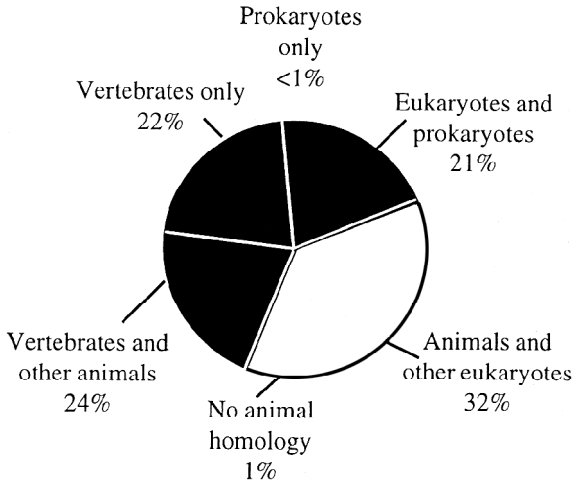


Fig. 5.16 Pie chart showing homology of predicted human proteins to protems of other species for those where homologues were deterced by comoputer searches of the public databases

the welth of data that genomics in providing. These new tools alow scientists to search genome databases for specific sequences or structrural features, to compare various features of different genomes, and to male inferences about the evolution of genomes. Indeed, with the sequence of the human genome aproaching completion, the question being asked is which genomes to sequence next—the mouse genome, the chimpanzee genome, and so on. One point is very clear: comparative genomics is providing unprecedented information about the evolution of species.

Unit 6 Cytogenetic Implications and Consequences of Structural Changes and Numerical Alterations of Chromosomes

Structure

6.1 Introduction

Between 20—50% of human abortuses have been shown to have a chromosome abnormality (Rashad and Kerr, 1965; Thiede and Metcalfe, 1966; Carr, 1967; Larson and Titus, 1970; Kajii *et al.*, 1973). An even wider range has been reported (8—50%), but the studies have not been comparable (Carr, 1971a).

Analysis of over 350 cases compiled from several published series has demonstrated that 45,XO is the most frequent single anomaly, constituting about 20% of the cases. Triploidy is only slightly less frequent, i.e., about 15%, while tetrapoidy has been demonstrated in 5%. Trisomies, as a group, have been found in about 50% of abortuses (E, 15%; G, 15%; D, 10%; C, 5%; A, B, F, 5%). The remainder are mosaics, or translocations (Carr, 1965, 1967; Inhorn 1967; Larson and Titus, 1970). only rarely is autosomal monosomy found (Kajii *et al.*, 1973).

It should be pointed out that within the E group, trisomy 16 and not trisomy 18 largely comprises this number (Carr, 1967; Waxman et a/., 1967). Use of recently developed banding techniques have shown trisomies for chromosomes 2, 3, 4, 6, 7, 8, 9, 10, 14, 15, 16, 18, 21 and 22 (Lauritsen et ah, 1972 Kajii *et al.*, 1973).

It is likely that findings in spontaneous abortuses do not necessarily reflect the frequency of clinical anomalies at conception since the more lethal ones probably never survived past a few cell divisions. This may explain the absence of viable trisomic states for A, B, most C, 14, 15, 16, 17, and F chromosomes being a more likely explanation than low frequency of meiotic or mitotic error for these chromosomes. The apparent high lethality of the 45,XO embryo cannot be explained. There is suggestions that embryos with chromosomal abnormalities are more likely to be aborted earlier than those with normal karyotypes (Carr, 1965, 1967; Szulman, 1965; Dhadial *et al.*, 1970).

The gross appearance of abortuses have some correlation with their chromosome status. A recognizable fetus is most frequent in the 45, XO group. Not uncommonly they can be recognized by cystic hygromas (neck blebs) in older fetuses. Triploid abortuses characteristicaly exhibit hydatidiform degeneration of villi and only rarely contain an embryo (Szulman, 1965; Carr, 1965, 1971b; Singh and Carr, 1967; Boue *et al.*, 1967). Trisomic abortuses do not have any specific phenotype with the possible exception of those having D-group trisomy,which not uncommonly have facial clefts. Presumably most of these are 13 trisomics (Roux, 1970). Kajii *et al.* (1973) found no example, however, of trisomy 13 in their large series.

When analyzed for mean maternal age, polyploid and XO abortuses have been found to be from younger mothers while trisomics have been from older mothers (Carr, 1965, 1971a; Szulman, 1965; Kerr *et al.*, 1966). However, the mean maternal age for 45,XO abortuses has been higher than that of survivors (Dhadial etal., 1970). Arakaki and Waxman (1970) found an increase in mean maternal age in cases of 16 trisomy abortuses.

In those couples having a history of two or more spontaneous abortions. 1 in 26 couples was found to have a translocation. This contrasts with the 0.4% found in the general population (Lucas *et al.*, 1972).

6.2 Chromosomes and cancer

Cancer cells may have bizarre karyotypes which may be hypodiploid, hyperdiploid, triploid, hypertriploid, hypotriploid, etc. Many unusual structural abnormalities have been described. Cells having over 1000 chromosomes have been documented. On the other hand, cancer cells have been described with normal karyotype and no evidence of structural abnormalities. While one may conclude that not all neoplasia is associated with gross chromosomal anomalies, one cannot exclude point mutations, gene deletions or duplications, or hidden rearrangements.

Review of chromosomal alterations concerning specific tumors is be yond the scope of this review and the reader is referred to Cervenka and Koulischer (1973).

A. Acute leukemias

In no form of acute leukemias have any specific chromosome abnormalities been described and, in at least half the cases, normal karyotypes have been found (Sandberg *et al.*, 1968; Krogh-Hensen, 1969; Whang Peng *et al.*, 1969). Furthermore, ostensibly identical clinical types of acute leukemia may minifest different chromosmoe patterns. Karyotypic changes, when present, are confined to the leukemic cells of the marrow or other organs. Long-term culture of

leukocytes from the blood of patients with acute leukemias is rarely successful. When an abnormal karotype is discerned, it seems to exhibit more hypoploid cell lines while acute lymphoblastic leukemias have more hyperploid lines. Karyotype analysis alone cannot be employed either for diagnosis or for prognosis concerning survival (Cervenka and Koulischer, 1973). During remission, aneuploid cells may disappear from the marrow only to reappear on relapse (Sandberg and Hossfeld, 1970).

B. Chronic myelogenous leukemia

In 1960, Nowell and Hungerford found delection of part of the long arm of a G chromosome associated with chronic myelogenous leukomia (CML). This unique structural abnormality, termed the Philadelphia chromosome (Ph^1), has been noted in over 90% of patients with CML (De Nava, 1969). Chronic myelogenous leukemia without the ph^1 chromosome and the ph^i chromosome without chronic myelogenous leukemia have been thoroughly reviewed by Cervenka and Koulische 91973).

The Ph^1 chromosome represents deletion with translocation to the long arm of a chromosome no. 9 (see Chapter 3). With.the use of quinacrine mustard fluorescent technique, it has been shown to be a G_{22} chromosome (Caspersson *el al.*, 1971; O'Riordan *et al.*, 1971). Its occurrence is limited to hematopoietic cells of all the granulocytic, erythocytic, and megakaryocytic types (Tough *et al.*, 1963; Clein and Flemans, 1966). Other tissues, such as skin fibroblasts, do not contain the ph1 chromosome.

The best technique for demonstration of the ph! chromosome is by direct study of bone marrow (Sandberg and Hossfeld, 1970).

It is an acquired, not an inherited, characteristic as demonstrated by its presence in only one of monozygotic twins with chronic myelogenous leukemia and not in the healthy co-twin (Jacobs *et al.*, 1966, Kosenow and pfeiffer, 1969). An unusual subgroup of ph'-positive CML patients are males who are missing the Y chromosome in all or in a portion of their marrow cells. However, fibroblasts and blood lymphocytes contain the Y chromosome (Lawler and Galton, 1966; Pedersen, 1968). Two or more ph1 chromosomes appearing in marrow cells either heralds or accompanies the transformation of CML to a blastic phase (Smalley, 1966).

A ph^1-like chromosome has been found in a small porportion of marrow cells of patients with acute myeloblastic leukemia, polycythema, thrombocytopenia, myeloid metaplasia with myelofiborsis, and erythroleukemia (Sandberg and Hossfeld, 1970). Khan (1973) reported two ph'-like chromosomes in acute myeloid leukemia.

C. Solid tumors

Most malignant tumors have aneuploid karyotypes ranging from hypodiploidy to extreme hyperdiploidy. Human tumor cell populations are clonal in nature, some tumors having but a single clone, others of two or more. No consistent cytogenetic findings have been described, but various markers have been noted, for example, a missing no. 22 chromosome in meningiomas (Zaftg and Singer, 1967; Mark *et al.*, 1972) and microchromosome in various neurogenic tumors (Cox *et al.*, 1965; Levan *et al,* 1968, Kucheria, 1968). Metastatic cells tend to have a higher ploidy and more variability in chromosome number (Sandberg *et al.,* 1967). In about 50% of the cases, abnormal ("marker") chromosome have been found in metastatic cancer cells. In general, there is a tendency toward relatively few chromosomes with distally placed centromeres, i.e., fewer B-, D-, and G-group chromosome and more A-3, C-group, and E 16 chromosomes (Atkin, 1970). Manolov and Manolova (1972) described a market band in a chromosome 14 in Burkitt's lymphoma.

Precancerous lesions, largely of the uterine cervia, have shown that dysplastic lesions exhibit chiefly pseudo-or near diploid karyotypes while carcinoma in situ shows an increase in ploidy and aberrations. Invasive carcinomas exhibit near diploid patterns, showing that progression does not depend on high chromosome counts (Atkin *et al,* 1967). Benign tumors have normal diploid karyotypes.

D. Waldenstrom's macroglobulinemia

Waldenstrom (1944) described a disorder characterized by intractable anemia and increased amounts of macroglobulin in serum, accompanied by fatigue, epistaxis, gingival hemorrhage, disturbances in vision, moderate lymphadenopathy, high sedimentation rate, and bone marrow lymphocytosis (Kok *et al.*, 1963). It is presently classified in the group of gammopathies. The disease usually appears after age 40 and is more frequent in males. Its relationship to lymphosarcoma and leukemia is not clear but patients diagnosed as having the disease sometimes develop chroniclymphatic leukemia or lymphoma.

Bottura *et al.* (1961) first described the presence of 47 chromosomes in about 50% of the cells, the supernumerary being about the size of an A-group chromosome. This finding was son confirmed by German *et al*. (1961) andBenirschke *et al*. (1962), who employed the term "W" chromosome.

The morphology of the marker chromosome is not constant. It usualy has been large with the centromere varying from metacentric to subterminal, but in some cases it has been as small as an F-group chromosome (Spengler *et al.,* 1966). The marker has been noted in both marrow and in peripheral cells in form 0—50% of cells (De Nava, 1969).

The abnormality is apparently acquired. Spengler *et al.* (1966) demonstrated the marker in one monozygotic twin who had Waldenstrom's macroglobulinemia but not in his normal co-twin. Interesting also are the findings of Lustman *et al.* (1968), who described an affected female with the marker whose otherwise healthy son had a normal karyotype but had an eleyated γ-globulin peak. Elves and Brown (1968) described the marker in 4 of 6 relatives of a patient with the disorder. Only one of the individuals had an elevated γ_1 fraction.

6.3 Diseases associated with spontaneous chromosome aberrations

At least seven inherited diseases have been found to be associated with spontaneous chromosome aberrations and increased frequency of leukemia or other neoplasias. The chromosome aberrations consist of gaps (achromatic regions), chromatid and chromosome breaks, fragments, reunion or translocation figures, ring chromosomes, and dicentric chromosomes. It should be emphasized that chromosome breakage may be very rarely seen in cells of ostensibly normal people. The enzyme deficiencies in the inherited disorders may result either in increased frequency in which openings appear in the DNA strands or in decreased speed with which such breaks are healed. Higurashi and Conen (1973) demonstrated greater *in vitro* chromosomal sensitivity in several of these disorders.

Fanconi's anemia, inherited as an autosomal recessive trait, is characterized by generalized skin pigmentation, pancytopenia with marrow hypoplasia, thumb and radius anomalies, hypogenitalism, and microcephaly (Fanconi, 1967). In 1964, Schroeder *et al*. noted that more than 40% of analyzed metaphases from peripheral blood cultures of patients with Fanconi's anemia exhibited chromatid gaps and breaks and chromosomal rearrangements. Direct bone marrow preparations have shown about 10% aberrant metaphases, usually involving B and C group chromosomes (Hirschman *et al.*, 1969; Shahid *et al.*, 1972). Of 41 cases subsquently studied, 36 were found to have similar findings. Among 170 known cases, four have terminated in leukemia and one had skin cancer (Swit and Hirschhorn, 1966; Swift, 1971). Heterozygotes have an increased frequency of leukemia (Gmyrek *et al.*, 1967; Swift, 1971). Occasionaly quadriradials and dicentric forms are noted but far less frequently than in Bloom's syndrome (vide infra).

Bloom's syndrome, consisting of growth retardation, sensitivity to.sunlight, and telagiectatic erythema, was reported by German (1969) to have chromosome breaks. Of 35 cases, four were found to have subsequently developed leukemia

or cancer, especially gastrointestinal. Cell lines with an abnormal karyotype have been described in clutured fibroblasts from a patient with Bioom's syndrome (Rauh and Soukup, 1968).

Quadriradial figures, i.e., a four-armed figure derived from two chromosomes, each arm consisting of sister chromatids of one of two homologous chromosomes. The autosomes most often involved are No. 1 and either No. 19 or 20. Asymmetric dicentric chromosomes, triradials, and abnormal new monocentric chromosomes can also be found. Heterozygotes may have the same types of figures, but less frequently than the homozygote.

Atxia-telangiectasia inherited as an autosomal recessive trait is characterized by retarded growth, progressive cerebellar ataxia, telangiectasia especialy about the face and bulbar conjunctiva, increased sonopulmonary infections, and decreased immunoglobulins (especially IgA and IgE). Approximately 10% develop lymphomas (pferiffer, 1970). Heche *et al.* (1966) reported a high frequency on in vitro chromosome breakage, a finding suported by Groop and Flatz (1967), pfeiffer (1970), and German (1972).

Lesser well-documented associations are wth glutathione reductase deficiency anemia, pernicious anemia, Kostmann's agranulocytosis (Schroeder and kurth, 1971), and possibly xeroderma pigmentosum (German *et al.*, 1970).

Matsaniotis *et al*. (1966) found approximately 20% aberrant cells from direct bone marrow preparations of a baby with Kostmann's agranulocytosis. Krogh-Jensen and Friis-Moller (1967) and Bottura and Continho (1988) described in vivo demonstration of chromosomal aberration in untreated pernicious anemia. The evidence for dominatntly inherited glutathione reductase deficiency is less solid and seems to depend on the stage of the disease (Hampel *et al.*, 1969).

German *et al*. (1970) detected a tendency toward the formation of pseudodiploid clones in cultured fibroblasts from a patient with xerox deama pigmentosum, an autosomal recessively inherited disorder, in which there is a proclivity toward development of skin cancer. Failure DNA repair following untraviolet light exposure has been demonstrated. Repair failure results from deficiency of ultraviolet-specific endonuclease.

Unit 7 Microbial Genetics

Structure

7.1 Introduction

Main constituent in microbial genetics are the bacteria. Bacteria reproduce asexually. However **parasexual** reproduction is also found among bacteria. In fact, genetic information is transferred from one bacterium to another by three totally distinct processes—**transformation, conjugation and transduction.**

Gene transfer in Bacteria :

Transformation (Definition) : This is the process by which a donor DNA molecule is taken up from the external environment and incorporated into the genome of a recipient cell.

Conjugation (Definition): This is the process by which bacterial cells make direct contact with each other, and DNA is transferred from one cell (the donor) to the other (the recipient cell).

Transduction (Definition) : This is the process by which DNA is transferred from one bacterial cell to another by a bacterial virus, or bacteriophage.

7.2 Bacterial mutation

Infection by the bacteriophage leads to the reproduction of the virus at the expense of the bacterial cell, which is lysed or destroyed. If a plate of *E. coli* is homogeneously sprayed with Tl, almost all cells are lysed. Rare *E. coli* cells, however, survive infection and are not lysed. If these cells are isolated and established in pure culture, all their descendants are *resistant* to Tl infection. This might be argued that the mutations responsible for Tl resistance were "induced" by the presence of the Tl viruses, and that, in the absence of the Tl viruses, the mutations would not have occurred. In 1943 Salvador Luria and Max Delbruck elegantly proved that such Tl-resistant cells result from **spontaneous mutation**.

Bacterial cells that bear spontaneous mutations, such as Tl resistance, can be isolated and established independently from the parent strain by means of various selection techniques. As a result, one can now induce and isolate mutations for almost any desired characteristic. Because bacteria and the viruses that infect them are haploid, all mutations are expressed directly in the descendants of mutant cells, adding to the case with which these microorganisms can be studied.

Bacteria are grown in either a liquid culture medium or in a petri dish on a semisolid agar surface. If the nutrient components of the growth medium are very simple and consist only of a organic carbon source (such as a glucose or lactose) and a variety of inorganic ions, including Na^+, K^+, Mg^{++}, and NH_4^+, present as inorganic salts, it is called **minimal medium.** To grow on such a medium, a bacterium must be able to synthesize all essential organic compounds (e.g., amino acids, purines, pyrimidines, sugars, vitamins, and fatty acids). A bacterium that can accomplish this remarkable biosynthetic feat—one that we ourselves cannot duplicate—is termed a **prototroph.** It is said to be wild type for all growth requirements and can grow on minimal medium. On the other hand, if a bacterium loses, through mutation, the ability to synthesize one or more organic components, it is said to be an **auxotroph.** For example, a bacterium that loses the ability to make histidine is designated as a *bis*$^-$ auxotroph, as opposed to its prototrophic bis$^+$ counterpart. For the bis bacterium to grow, this amino acid must be added as a supplement to the minimal medium. The medium that has been extensively supplemented is referred to as complete medium.

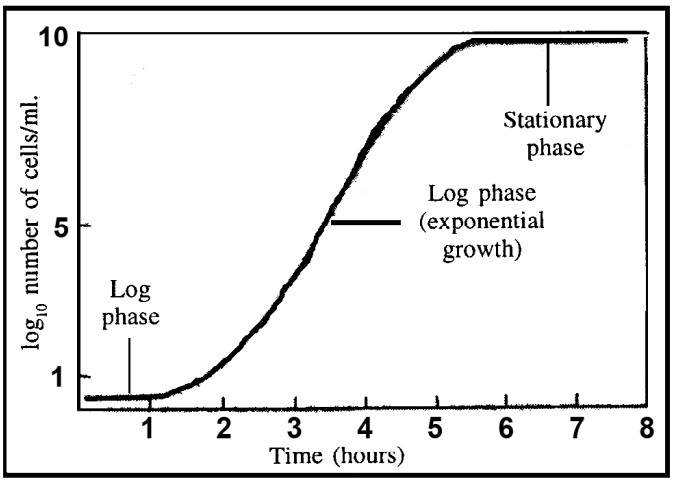


Fig. 7.1 A typical bacterial population growth curve illustrating the initial lag phase, the subsequent log phase where exponential growth occurs, and the stationary phase that occurs when nutrients are exhausted

To study mutant bacteria in a quantitative fashion, an inoculum (e.g., 0.1 ml, 1.0 ml) of bacteria is placed in liquid culture medium. The bacteria exhibit a characteristic growth pattern, as illustrated in Figure 7.1. Initially, during the lag phase, growth is slow. Then, a period of rapid growth follows called the **log phase,** during which cells divide many times with a fixed time interval between cell divisions, resulting in logarithmic growth. When the bacteria reach a cell density of about 10^9 cells per milliliter, nutrients and oxygen become limiting and cells enter the **stationary phase.** As the doubling time during the log phase may be as short as 20 minutes, an initial inoculum of a few thousand cells added to the culture can easily achieve a maximum cell density overnight.

Once cells are grown in liquid medium they can be quantitated. First, the bacteria are plated on (transferred to) semi-solid medium in a petri dish where, following incubation and many divisions, each cell gives rise to a colony visible on the surface of the medium. From the number of colonies that subsequently grow, it is possible to estimate the number of bacteria present in the original culture. If the number of colonies is too great to count, then serial dilutions of the original liquid culture can be made and plated, until the colony number is reduced to the point where it can be counted (Figure 7.2).

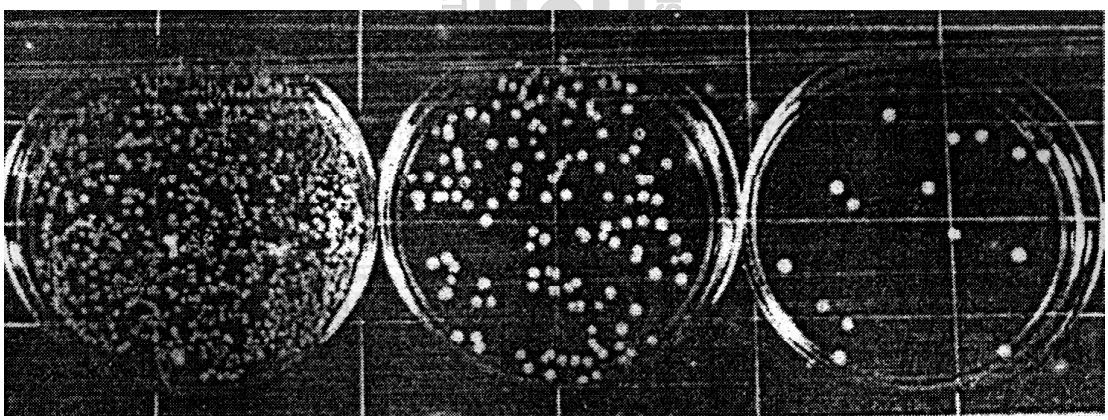


Fig. 7.2 Results of the serial dilution technique and subsequent culture of bacteria. Each of the dilutions varies by a factor of 10. Each colony was derived from a single bacterial cell

For example, assume that the three petri dishes in Figure 7.2 represent dilutions of 10^{-3}, 10^{-4}, and 10^{-5}, respectively (left to right). We select the dish were there are few enough colonies to be accurately counted. Because each colony presumably arose from a single bacterium, the number of a colonies times the dilution factor represents the number of bacteria in the initial milliliter. In this case, the dish farthest to the right contains 15 colonies. Since it represents a dilution of 10^{-5}, we can estimate the initial number of bacteria to be 15×10^5 per milliliter. Calculations such as these are useful in a number of studies.

7.3 Conjugation is one of the methods of genetic recombination in bacteira

7.3.1 Introduction

Development of techniques that allowed the identification and study of bacterial mutations led to detailed investigations of the arrangement of genes on the bacterial chromosome. Such studies began in 1946 when Joshua Lederberg and Edward Tatum showed that bacteria undergo conjugation, a parasexual process in which the genetic information from one bacterium is transferred to and recombined with that of another bacterium. Like meiotic crossing over in eukaryotes, genetic recombination in bacteria provided the basis for the development of methodology for chromosome mapping. Note that the term **genetic recombination,** as applied to bacteria and bacteriophages, leads to the *replacement* of one or more genes present in one strain with those from a genetically distinct strain. While this is somewhat different from our use of genetic recombination in eukaryotes, where the term describes crossing over that results in *reciprocal exchange events,* the overall effect is the same : Genetic informatin is transferred from one chromosome to another, resulting in an altered genotype. Two other phenomena that result in the transfer of genetic information from one bacterium to another, **transformation** and **transduction,** have also served as a basis for determining the arrangement of genes on the bacterial chromosome.

Lederberg and Tatum's initial experiments were performed with two multiple auxotroph strains (nutritional mutants) of *E. coli* K12. As shown in Figure 7.3, Strain A required methionine (met) and biotin (bio) in order to grow, whereas strain B required theonine (thr), leucine (leu), and thiamine (thi). Neither strain would grow on minimal medium. The two strains were first grown separately in supplemented media, and then cells from both were mixed and grown together for several more generations. They were then plated on minimal medium. Any bacterial cells that grew on minimal medium are **prototrophs** (wild-type bacteria that did not need nutritional supplements). It was highly improbable that any of the cells that contained two or three mutant genes would undergo spontaneous mutation simultaneously at two or three gene sites. Therefore, any prototrophs recovered must have arisen as a result of some form of genetic exchange and recombination.

In this experiment, prototrophs were recovered at a rate of $1/10^7$ (10^{-7}) cells plated. The controls for this experiment involved separate plating of cells from strains A and B on minimal medium. No prototrophs were recovered. On the basis of these observations, Lederberg and Tatum proposed that, while the events were indeed quite rare, genetic recombination had occurred.

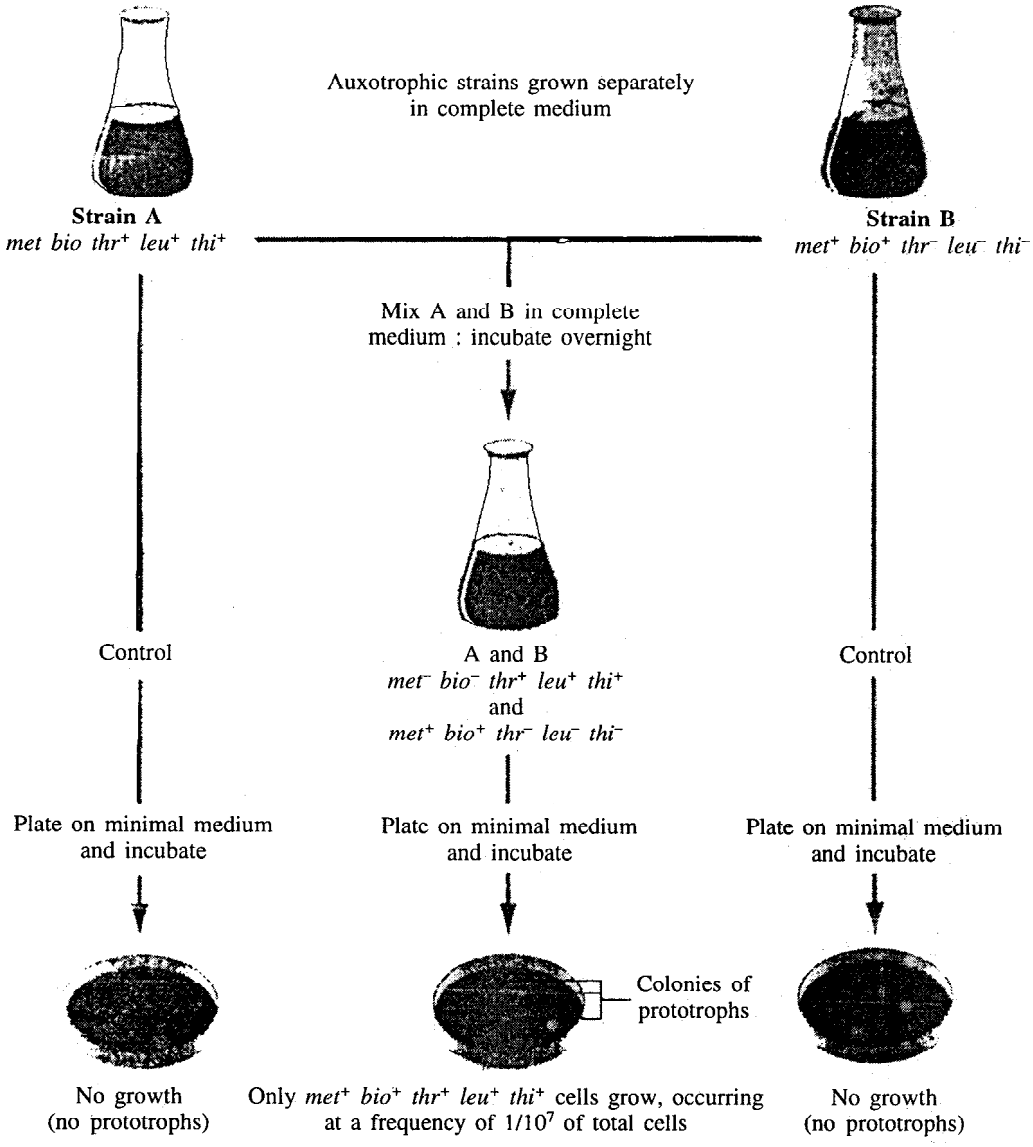


Fig. 7.3 Genetic recombination involving two auxotrophic strains producing prototrophs. Neither auxotrophh will grow on minimal medium, but prototrophs will, suggesting that genetic recombination has occurred

7.3.2 F^+ and F^- Bacterial strains

Lederberg and Tatum's findings were soon followed by numerous experiments designed to elucidate the genetic basis of conjugation. It quickly became evident that different strains of bacteria were involved in a unidirectional

transfer of genetic material. When cells serve as donors of parts of their chromosomes, they are designated as **F^+ cells** (F for "fertility"). Recipient bacteria receive the donor chromosome material (now known to be DNA), and recombine it with part of their own chromosome. They are designated as **F^- cells.**

7.3.3 Conjugation in bacteria

It was established subsequently that cell contact is essential to chromosome transfer. Support for this concept was provided by Bernard Davis, who designed a U-tube in which to grow F^+ and F^- cells (Figure 7.4). At the base of the tube was a glass filter with pores large enough to allow passage of the liquid medium, but too small to allow the passage of bacteria. The F^+ cells were placed on one side of the filter and F" cells on the other side. The medium was moved back and forth across the **filter** so that the bacterial cells essentially shared a common medium during incubation. Samples from both sides of the tube were then plated independently on minimal medium, but no prototrophs were found. Davis concluded that *physical contact is essential to genetic recombination.* Such physical interaction is the initial stage of the process of conjugation and is mediated through a conjugation tube called the F, or sex, pilus. Bacteria often have many pili, which are microscopic tube-like extensions of the cell. Different types of pili perform different cellular functions, but all piii are involved in some way with adhesion. After contact has been initiated between mating pairs via the F pili (Figure 7.5), transfer of DNA begins.

Later evidence established that F^+ cells contained a fertility factor (called the F factor) that confers the ability to donate part of their chromosome during conjugation. Experiments by Joshua and Esther Lederberg and by William Hayes and Luca Cavalli-Sforza showed that certain conditions could eliminate the F factor in otherwise fertile cells. However, if these "infertile" cells were then grown with fertile donor cells, the F factor was regained.

The conclusion that the F factor is a mobile element was further supported by the observation that, following conjugation and genetic recombination, recipient cells always become F^+. Thus, in addition to the *rare* cases of transfer of genes from the bacterial chromosome (genetic recombination), the F factor itself is passed to *all* recipient cells. On this basis, the initial crosses of Lederberg and Tatum (Figure 7.3) may be designated.

STRAIN A		STRAIN B
F^+	X	F^-
DONOR		RECIPIENT

Isolation of the F factor confirmed these conclusions. Like the bacterial chromosome, though distinct from it the F factor has been shown to consist of

a circular, double-stranded DNA molecule, equivalent to about 2 percent of the bacteria] chromosome (about 100,000 nucleotide pairs). Contained in the F factor, among others, are 19 genes, the products of which are involved in the transfer of genetic information (*tra* genes). These include those essential to the formation of the sex pilus.

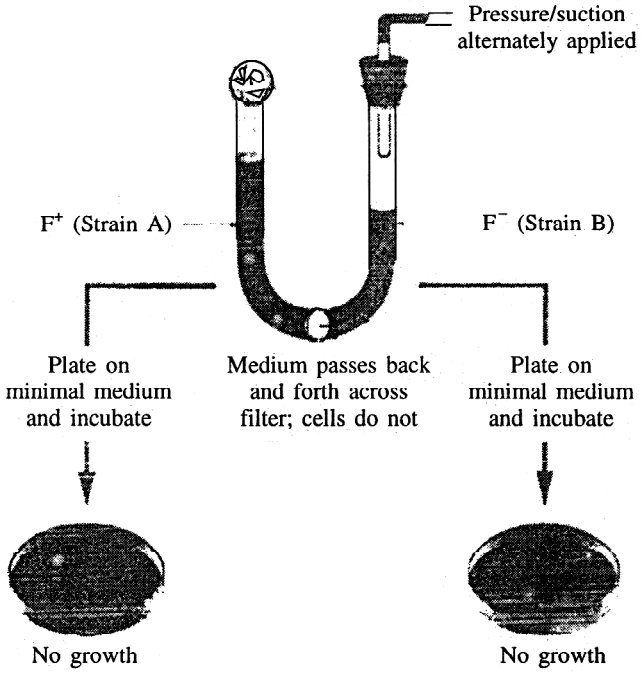


Fig. 7.4 When strain A and B auxotrophs are grown in a common medium but separated by a no genet.c recombination occurs and no prototrophs are produced. This apparatus is called a Davis U-tube

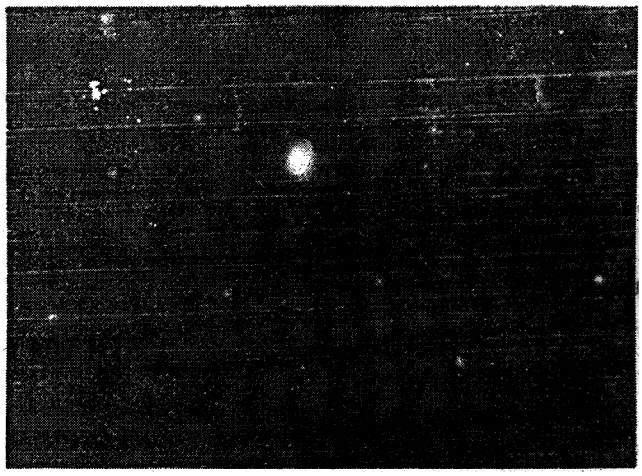


Fig. 7.5 An electron micrograph of conjugation between an F^+ *E coli* cell. The sex pllus linking them is clearly visible

As we soon shall see, the F^- factor is in reality an autonomous genetic unit referred to as a plasmid. However, in our historical coverage of its discovery, we will continue in this chapter to refer to it as a "factor."

It is believed that the transfer of the F factor during conjugation involves separation of two strands of the F factor double helical DNA and the movement of one of the two strands into the recipient. The other strand remains in the donor cell. Both of these parental strands serve as templates for DNA replication, resulting in two intact F factors, one in each of the two cells. Both cells are now F^+ (See in Figure 7.6).

To summarize, *E. coli* cells may or may not contain the F factor. When it is present, the cell is able to form a sex pilus and potentially serve as a donor of genetic information. During conjugation, a copy of the F factor is almost always transferred from the F^+ cell to the F^- recipient, converting it to the F^+ state. The question remains as to exactly how a very low percentage of F^- cells undergo genetic recombina-tion. The answer awaited further experimentation. Subsequent
discoveries not only clarified how genetic recombination occurs but also defined a mechanism by which the *E. coli* chromosome could be mapped.

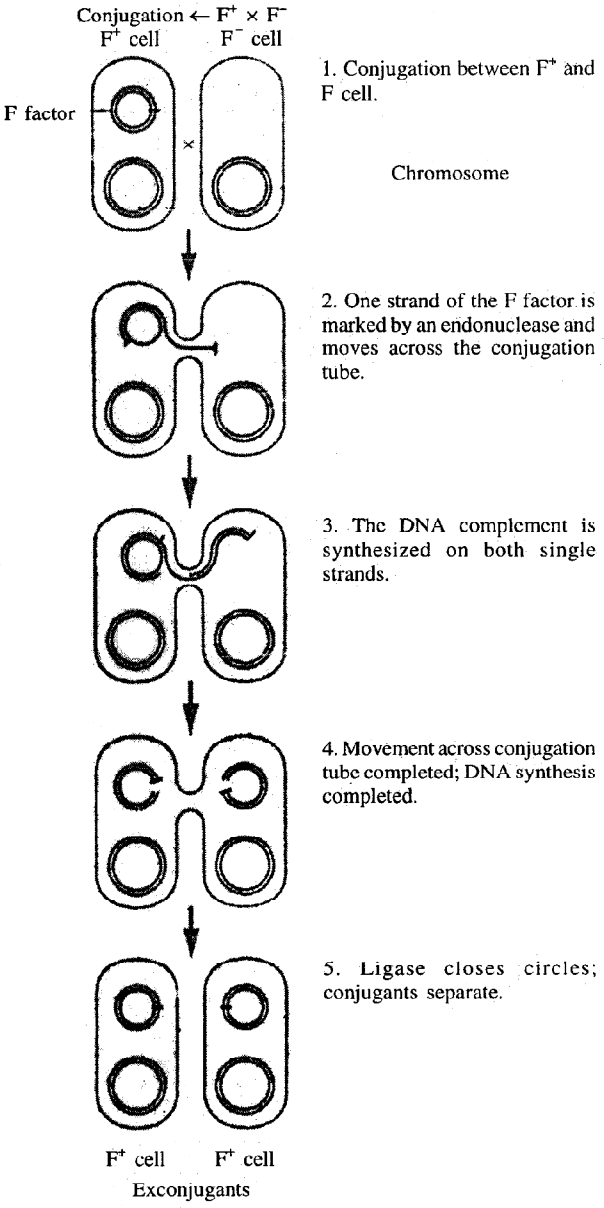


Fig. 7.6 An $F^+ \times F$ mating demonstrating how the recipient F^- cell is converted to F^+. During conjugation, the DNA of the F factor is replicated with one new copy entering the recipient cell converting it to F^+. The black bar has been added to the F factors to follow their clockwise rotation during replication

7.3.4 Hfr bacteria and chromosome mapping

In 1950, Cavalli-Sforza treated an F^+ strain *of E. coli* K12 with nitrogen mustard, a potent chemical known to induce mutations. From these treated cells, he recovered a strain of donor bacteria that underwent recombination at a rate of $1/10^4$ (10^{-4}), 1000 times more frequently than the original F^+ strains. In 1953, William Hayes isolated another strain demonstrating a similar elevated frequency. Both strains were designated **Hfr,** or **high-frequency recombination.** Because Hfr cells behave as chromosome donors, they are a special class of F^+ cells.

Another important difference was noted between Hfr strains and the original F^+ strains. If the donor is an Hfr strain, recipient cells, while sometimes displaying genetic recombination, never become Hfr; that is, they remain F^- In comparison, then,

$F^+ \quad x\ F^- \rightarrow F^+$ (low rate of recombination)
Hfr $x\ F^- \rightarrow F^-$ (higher rate of recombination)

Perhaps the most significant characteristic of Hfr strains is the nature of recombination. In any given strain, certain genes are more frequently recombined

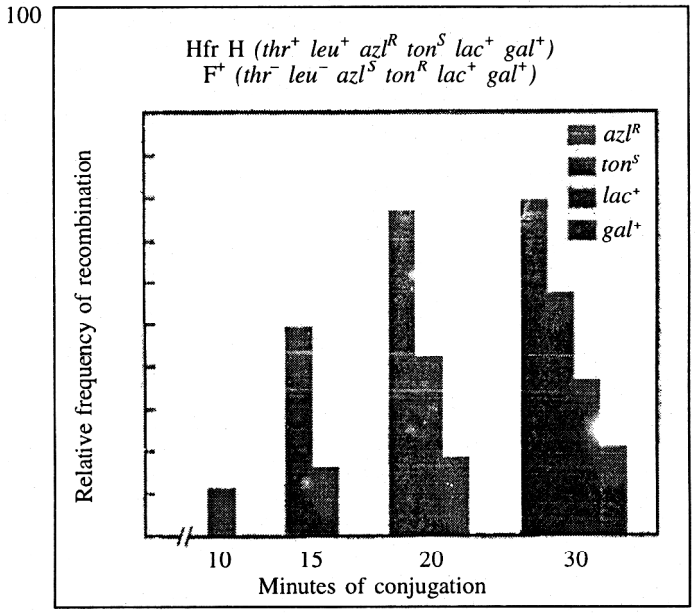


Fig. 7.7 The progressive transfer during conjugation of various genes from a specific Hfr strain of *E, coli* to an F strain. Certain genes *{azi* and *ton)* are transferred sooner than others and recombine more frequently. Others *(lac* and *gal)* take longer to be transferred and recombine with a lower frequency. Others *(thr* and *leu)* are always transferred and are used in the initial screen for recombinants

than others, and some do not recombine at all. This *nonrandom pattern of gene transfer* was shown to vary from Hfr strain to Hfr strain. While these results were puzzling. Hayes interpreted them to mean that some physiological alteration of the F factor had occurred, resulting in the production Hfr strains of *E. coli.*

In the mid-1950s, experimentation by Ellie Wollman and Francois Jacob explained the differences between cells that are Hfr and those that are F^+ and showed how Hfr strains allow genetic mapping of the *E, coli* chromosome. Wollman and Jacob first incubated a culture containing a mixture of an Hfr strain and an F strain. To facilitate the recovery of only recombinants, the Hrf strain was sensitive to an antibiotic while the recipient strain was resistant. At various intervals, the researchers removed samples and placed them in a blender. The shear forces created in the blender seperated conjugating bacteria so that the transfer of the chromosome was effectively terminated. To assay the cells for genetic recombination following the blender treatment, they were grown on medium *containing* the antibiotic in order to ensure the recovery of only recipient cells.

7.3.5 Interrupted mating technique

This process, called the interrupted mating technique, demonstrated that specific genes of a given Hfr strain were transferred and recombined sooner than others. Figure 7.7 illustrates this point. During the first 8 minutes after the two strains were initially mixed, no genetic recombination could be detected. In cells assayed at about 10 minutes, recombination of the azi^R gene could be detected, but no transfer of the ton^S, lac^+, or gat^+ genes was noted. By 15 minutes, 70 percent of the recombinants were azl^R; 30 percent were now also ton^s; but none was lac^+ or *gat.* Within 20 minutes, the lac^+ gene was found among the recombinants; and within 30 minutes, gal^+ was also being transferred. Wollman and Jacob had demonstrated an *oriented transfer of genes* that was correlated with the length of time conjugation was allowed to proceed.

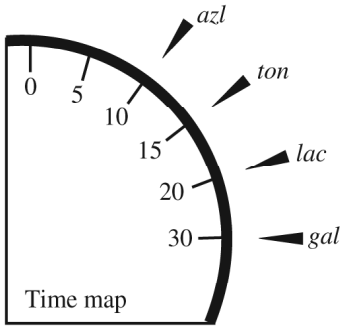


Fig. 7.8 A time map ot the genes studied in the experiment depicted in Figure 7.7

It appeared that the chromosome of the Hfr bacterium was transferred linearly and that the gene order and distance between genes, as measured in minutes, could be predicted from such experiments (Figure 7.8). This information served as the basis for the first genetic map of the *E. coli* chromosome. "Minutes" in bacterial mapping are equivalent to "map units" in eukaryotes.

Wollman and Jacob repeated the same type of experimentation with other Hfr strains, obtaining similar results with one important difference. Although genes were always transferred linearly with time, as in their original experiment, which genes entered first and which followed later seemed to vary from Hfr strain to Hfr strain [Figure 7.9(a)].

When they reexamined the rate of entry of genes, and thus the different genetic maps for each strain, a definite pattern emerged. The major difference between each strain was simply the point of the origin and the direction in which entry proceeded from that point [Figure 7.9(b)].

(a)

Hfr strain	(Ear liest)														(latest)
H	*thr*	–	*leu*	–	*azi*	–	*ton*	–	*pro*	–	*lac*	–	*gal*	–	*thi*
1	*lac*	–	*pro*	–	*ton*	–	*azi*	–	*leu*	–	*thr*	–	*thi*	–	*gal*
2	*pro*	–	*ton*	–	*azi*	–	*leu*	–	*thr*	–	*thi*	–	*gal*	–	*lac*
7	*ton*	–	*azi*	–	*leu*	–	*thr*	–	*thi*	–	*gal*	–	*lac*	–	*pro*

(b)

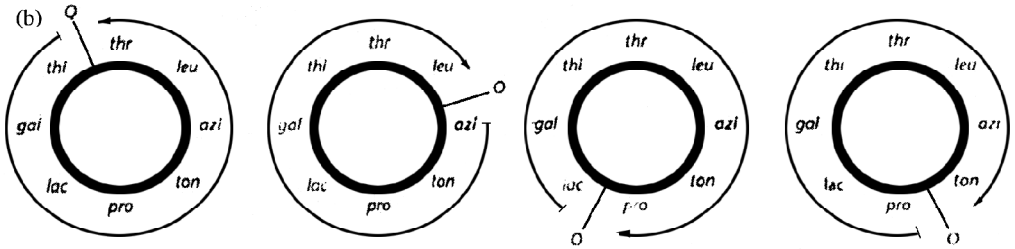


Fig. 7.9 (a) The order of gene transfer in four Hfr strains, suggesting that the *E. coli* chromosome is circular, (b) The point where transfer originates *(O)* is identified in each strain. Note that transfer can proceed in either direction, depending on the strain. The origin is determined by the point of integration into the chromosome of the F factor, and the direction of transfer is determined by the orientation of the F factor as it integrates

To explain these results, Wollman and Jacob postulated that the *E. coli* chromosome is circular. If the point of origin *(O)* varied from strain to strain, a different sequence of genes would be transferred in each case. But what determines *O*? They proposed that in various Hfr strains, the F factor integrates into the chromosome at different points and its position determines the *O* site. One such case of integration is shown in Figure 7.10 (Step 1). During conjugation between this Hfr and an F cell, the position of the F factor determines the initial point of transfer (Step 2 and 3). Those genes adjacent to *O* are transferred first. *The F factor becomes the last part to be transferred* (Step 4). Apparently, conjugation rarely, if ever, lasts long enough to allow the entire chromosome to pass across the conjugation tube (Step 5). This proposal explains why recipient cells, when mated with Hfr cells, remain F^-.

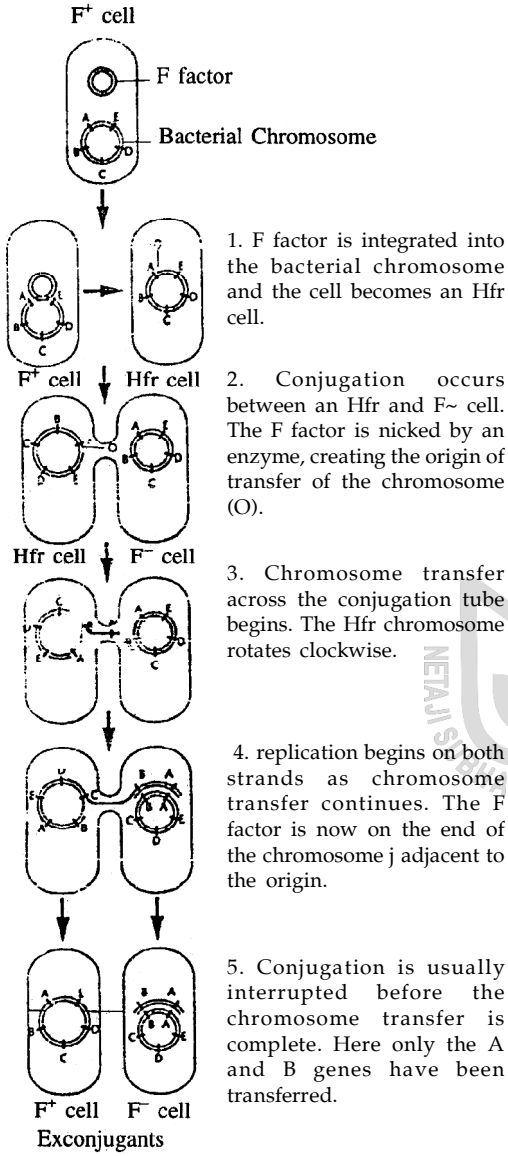


Fig. 7.10 Conversion of $F^{4"}$ to an Hfr state occurs by the integration of the F factor intot he bacterial chromosome. The point of integration determines the origin *(O)* of transfer, during conjugation, the F factor, now integrated into the host chromosome, is nicked by an enzyme, initiating transfer of the chromosome at that point. Conjugation is usually interrupted prior to complete transfer. Above, only the A and B genes are transferred to the F cell, which may recombine with the host chromosome

Figure 7.10 also depicts the way in which the two strands making up a DNA molecule unwind during transfer, allowing for the entry of one of the strands of DNA into the recipient (Step 3). Following replication, the entering DNA now has the potential to recombine with its homologous region of the host chromoso'me. The DNA strand that remains in the donor also undergoes replication.

The use of the interrupted mating technique with different Hfr strains has provided the basis for mapping the entire *E. coli* chromosome, Mapped in time units, strain K12 (or *E. coli* K12) is 100 minutes long. Over 900 genes have now been placed on the map. In most instances, only a single copy of each gene exists.

7.3.6 Recombination in $F^+ \times F^-$ matings : A review

The above model has helped geneticists to better understand how genetic recombination occurs during the F^+xF^- matings. Recall that recombination occurs much less frequently in them than in $Hfr \times F^-$ matings, and that random gene transfer is involved. The current belief is that when F^+ and F^- cells are mixed, conjugation occurs readily and that each F^- cell involved in conjugation with an F^+ cell receives a copy of the F factor, *but that no genetic recombination occurs.* However, at an extremely low frequency in a population of F^+ cells, the F factor integrates spontaneously from the cytoplasm to a random point in the bacterial chromosome, converting the F^+

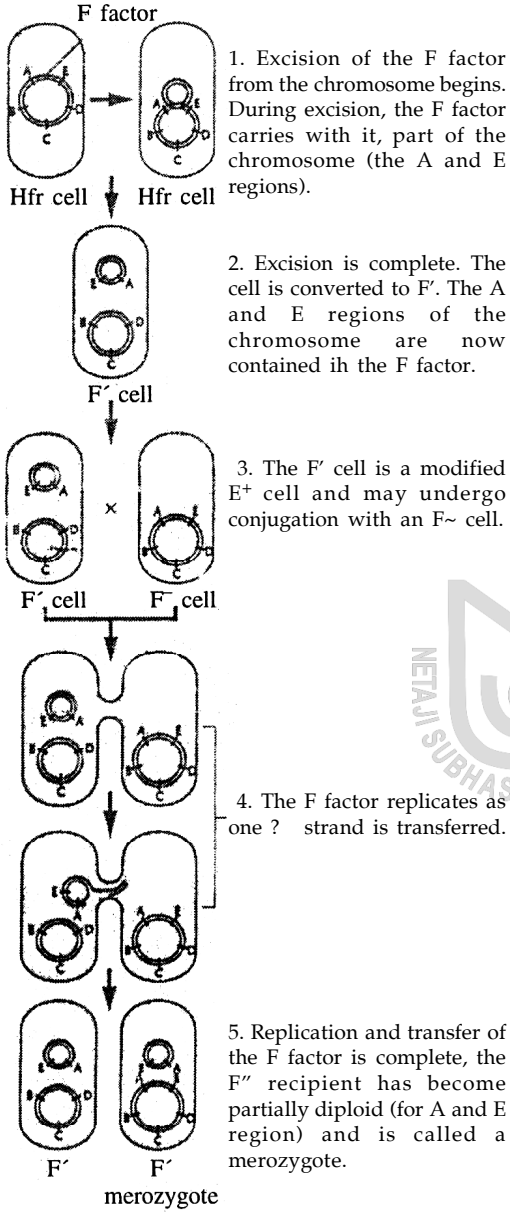


Fig. 7.11 Conversion of an Hfr bacterium to F[1'" and is subsequent mating with an F cell. The conversion occurs when the F factor loses its integrated status. During excision from the chromosome, it carries with it one or more chromosomal genes (A and E) F lowing conjugation with an F~ cell, the recipient cell becomes partially diploid and is called a merozygote. It also behaves as an F~ donor ones

cell to the state as we saw in Figure 7.10 Therefore, in F^+ x F^- crosses, the extremely low frequency of genetic recombination (10^{-7}) is attributed to the rare, newly formed Hfr cells, which then undergo conjugation with F^- cells. Because the point of integration of the F factor is random, a nonspecific gene transfer ensues, leading to the low-frequency, random genetic recombination observed in the F^+ x F^- experiment. Unless the recipient cell simultaneously or subsequently undergoes conjugation with a separate F^+ cell, it will remain F^-. Most often, the recombinants become F^+.

7.3.7 The F′ State and merozygotes

In 1959, during experiments with Hfr strains of *E. coli.* Edward Adelberg discovered that the F^+ factor could lose its integrated status, causing the cell to revert to the F^+ state (Figure 7.11 Step 1). When this occurs, the F factor frequently carries several adjacent bacterial genes along with it (Step 2). Adelberg labeled this condition F′ to distinguish it from F^+ and Hfr. F^+ like Hfr, is thus another special case of F^+. This conversion is described as one from Hfr to F^+.

The presence of bacterial genes within a cytoplasmic F factor creates an interesting situation. An F′ bacterium behaves like an F^+ cell, initiating conjugation with F^- cells (Figure 7.11-Step 3). When this occurs, the F factor, containing chromosomal genes, is transferred to the F^+ cell (Step 4). As a result, whatever chromosomal genes are part of the" F factor are now present in

duplicate in the recipient cell (Step 5), because the recipient still has a complete chromosome. This creates a partially diploid cell called a merozygote. Pure cultures of F′ merozygotes can be established. They have been extremely useful in the study of bacterial genetics, particularly in genetic regulation.

7.3.8 Bacterial recombination is dependent on rec poteins

Once researchers established that a unidirectional transfer of DNA occurs between bacteria, they were interested in determining how the actual recombination event occurs in the recipient cell. Just how does the donor DNA replace the comparable region in the recipient chromosome? As with many systems, the biochemical mechanism by which recombination occurs was deciphered through genetic stuies. Major insights were gained as a result of the isolation of a group of mutations representing genes called *rec.*

The first relevant observation in this case involved a series of mutant genes labeled *recA, recB, recC,* and *recD.* The first mutant gene, *recA,* was found to diminish genetic recombination in bacteria 1000-fold, nearly eliminating it altogether. The other *rec* mutations reduced recombination by about 100 times. Clearly, the normal wild-type products of these genes play some essential role in the process of recombination.

By looking for a functional gene product present in normal cells but missing in mutant cells, researchers subsequently isolated several gene products and showed that they played a role in genetic recombination. The first is called the RecA protein.* The second is a more complex protein called the RecBCD product, an enzyme consisting of polypeptide subunits encoded by three other *rec* genes. The roles of these proteins have now been elucidated *in vitro.* As a result of this genetic research, our knowledge of the process of recombination has been extended considerably. These discoveries underscore the value of isolating mutations, establishing their phenotypes, and determining the biological role of the normal, wildtype gene as a result of subsequent investigation.

7.3.9 F factors are plasmids

In the preceding sections we have examinined the extra-chromosomal heredity unit called the F factor. When it exists autonomously in the bacterial cytoplasm, the F factor is composed of a double-stranded closed circle of DNA [Figure 7.12(a)]. These characteristics place the F factor in the more general category of genetic .structures called plasmids. These structures contain one or more genes—often, quite a few. Their replication depends on the same enzymes that replicate the chromosome of the host cell, and they are distributed to daughter cells along with the host chromosome during cell division.

Plasmids are generally classified according to the genetic information specified by their DNA. The F factor confers fertility and contains genes essential

*Note that the names of bacterial genes begin with lowercase letters and are italicized. The names of the corresponding gene products (proteins) begin with an uppercase letter and are not italicized. For example, the m-4 gene encodes the RecA protein.

for sex pilus formation, upon which genetic recombination depends. Other examples of plasmids include the R and the Col plasmids.

Most R plasmids consist of two components : the RTF (resistance transfer factor) and one or those r-de-terminants {Figure 7.12(b)]. The RTF encodes genetic information essential to transfer of the plasmid between bacteria, and the r-determinants are genes conferring resistance to entibiotics.

The **Col plasmid,** ColEl, derived from *E. coli,* is clearly distinct from R plasmids. It encodes one or more proteins that are highly toxic to bacterial strains that do not harbor the same plasmid. These proteins, called **colicins,** may kill neighboring bacteria. Bacteria that carry the plasmid are said to be colicinogenic. Present in 10 to 20 copies per cell, the plasmid also contains a gene encoding an immunity protein that protects the host cell from the toxin. Unlike an R plasmid, the Col plasmid is not usually transmissible to other cells.

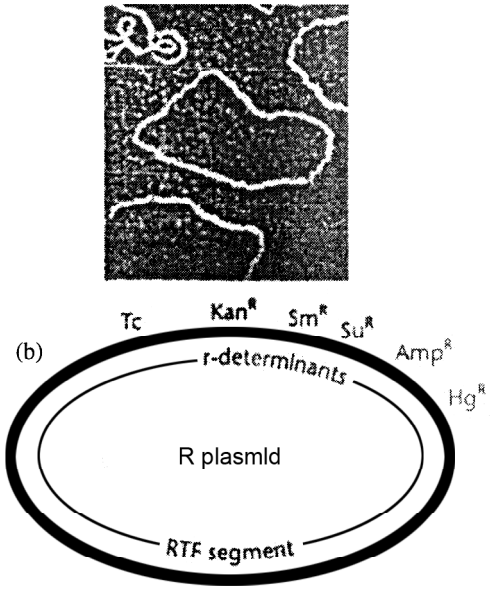


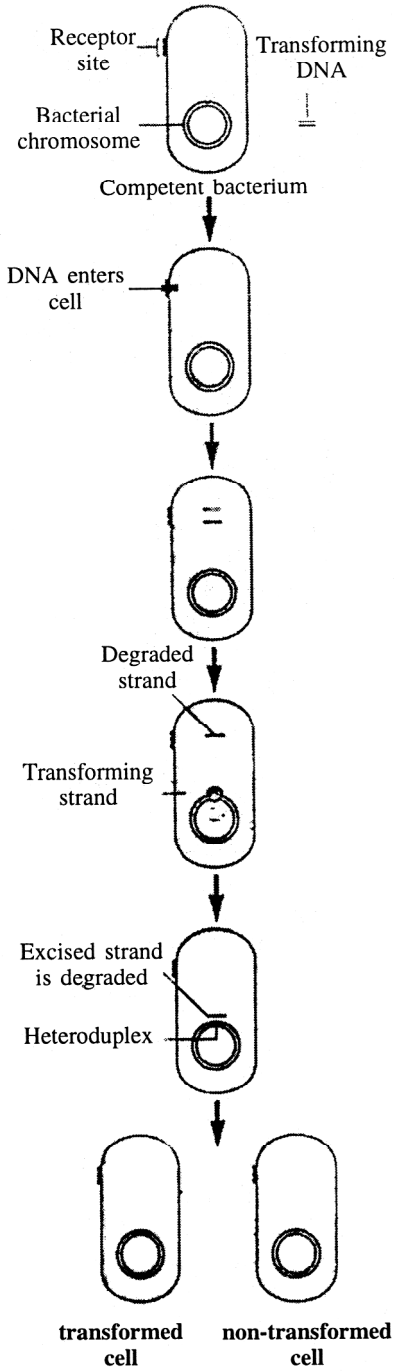
Fig. 7.12 (a) Electron micrograph of a plasmid isolated from *E. coll;* (b) diagramatic representation of an R plasmid containing resistance transfer factors (RTFs) and multiple r-determinance (Tc, tetracycline; Kan, kanamycin; Sm, streptomycin; Su, sulfonamide; Amp, ampicillian; and Hg, mercury)

7.4 Transformation is another process leading to genetic recombination in bacteria

Transformation is another process that provides a mechanism for the recombination of genetic information in some bacteria. In transformation, small pieces of extracel Iular DNA are taken up by a living bacterium, ultimately leading to a stable genetic change in the recipient cell.

7.4.1 The Transformation Process

Transformation (Figure 7.13) consists of numerous steps that can be divided into two main categories: (1) entry of DNA into a recipient cell, and (2) recombination of the donor DNA with its homologous region in the recipient chromosome. In a population of cells, only those in a particular physiological state, referred to as competence, take up DNA. Entry is thought to occur at a limited number of receptor sites on the surface of the bacterial cell. Passage



1. Extracellular DNA binds to the competent cell at a receptor site.

2. DNA enters the cell and strands separate.

3. One strand of transforming DNA is degraded. The other and pairs homologously with the host cell DNA.

4. The transforming DNA is integrated, forming a heteroduplex.

5. Following one round of cell division, a transformed and a non-transformed cell are produced.

Fig. 7.13 Proposed steps leading to transformation of a bacterial cell by exogenous DNA. Only one of the two strands of the entering DNA is involved in the transformation event, which is completed following cell division

across the cell wall and membrane is an active process requiring energy and specific transport molecules. This concept is supported by the fact that substances that inhibit energy production or protein synthesis in the recipient cell also inhibit the transformation process.

During the process of entry, one of the two strands of the invading DNA molecule is digested by nucleases, leaving only a single strand to participate in transformation (Step 2 and 3). The surviving DNA strand aligns with its complementary region of the bacterial chromosome. In a process involving several enzymes, this segment of DNA replaces its counterpart in the chromosome, which is excised and degraded (Step 4).

For recombination to be detected, the transforming DNA must be derived from a different strain of bacteria, bearing some genetic variation. Once integrated into the chromosome, the recombinant region contains one DNA strand from the bacterial chromosome and one from the transforming DNA. Because these strands are not genetically identical, this helical region is referred to as a **heteroduplex.** Following one round of replication, one chromosome is restored to its original configuration, identical to that of the recipient cell, and the other contains the transformed gene. Cell division produces one host cell and one transformed cell (Step 5).

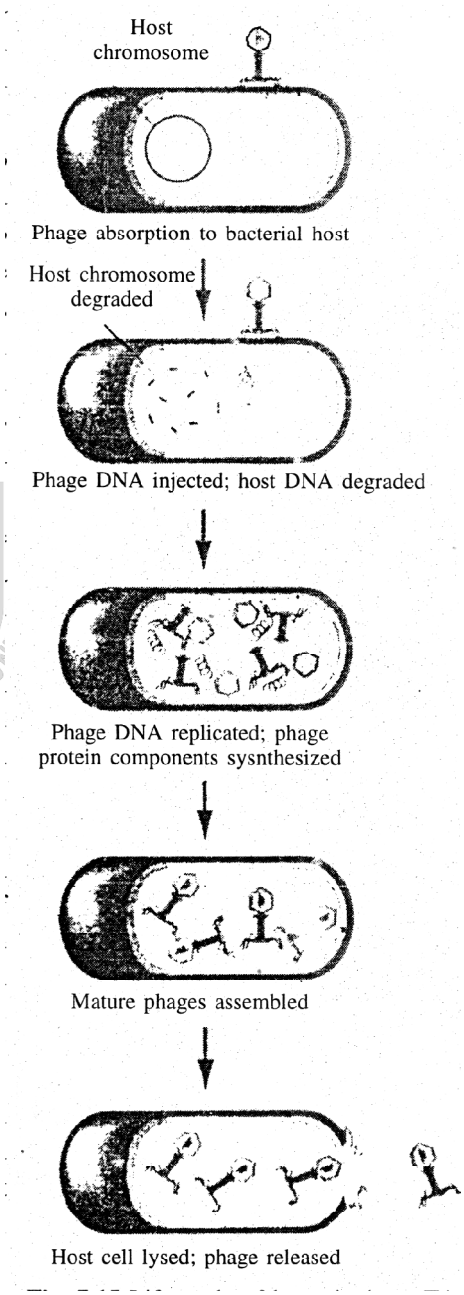


Fig. 7.15 Life cycle of bacteriophage T4

7.4.2 Lysogeny

The relationship between virus and bacterium does not always result in viral reproduction and lysis. As early as the 1920s, it was known that some bacteriophages could enter a bacterial cell and establish a symbiotic relationship with it. The precise molecular basis of this symbiosis is now well understood.

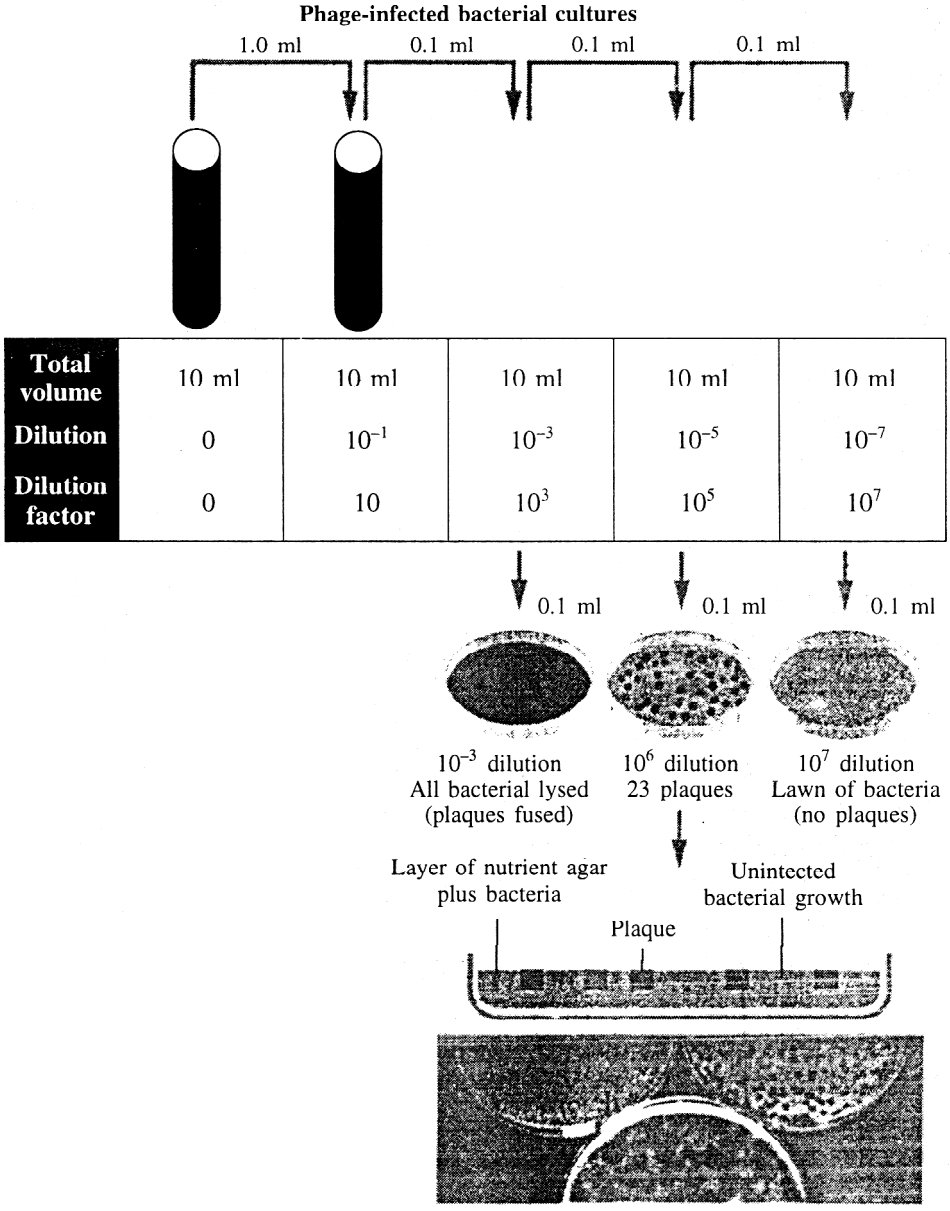


Fig. 7.16 Diagramatic illustration of the plaque assay for bacteriophage analysis. Serial dilutions of a bacterial culture infected with bacteriophages are first made. Then three of the dilutions (10^3, 10^6 and 10^7) are analyzed using the plaque assay technique in each case 0.1 ml of the diluted culture is used. Each plaque represents the initial infection of one bacterial cell by one bacteriophage. In the 10^3 dilution, so many phaes are present that all bacterial are lysed. In the 10^{-5} dilution 23 plaques are produced. In the 10^{-7} dilution, the dilution factor is so great that no phages are present in the 0.1 ml sample, and thus no plaques form. From the 0.1 ml sample of the 10^{-3} dilution, the original bacteriophage desnity can be calculated as 23, 10, 10^5 phages/ml (23.10^4 or 23.10). The photograph illustrates phage T2 plaques on lawns of *E. coli*

Upon entry, the viral DNA, instead of replicating in the bacterial cytoplasm, is integrated into the bacterial chromosome, step that characterizes the developmental stage referred to as lysogeny. Subsequenly, each time the bacterial chromosome is replicated, the viral DNA is also replicated and passed to daughter bacterial cells following division. No new viruses are produced and no lysis of the bacterial cell occurs. However, in response to certain stimuli, such as chemical or ultraviolet-light treatment, the viral DNA may lose its integrated status and initiate replication, phage reproduction, and lysis of the bacterium. (Fig, 7.15)

Several terms are used to describe this relationship. The viral DNA integrated into the bacterial chromosome is called a **prophage.** Viruses that can either lyse the cell or behave as a prophage are called **temperate.** Those that can only lyse the cell are referred to as virulent. A bacterium harboring a prophage has been lysogenized and said to be lysogenic; that is, it is capable of being lysed as a result of induced viral reproduction. The viral DNA, which can replicate either in the bacterial cytoplasm or as part of the bacterial chromosome, is sometimes classified as an **episome.**

7.5 Transduction is virus-mediated bacterial DNA transfer

In 1952, Norton Zinder and Joshua Lederberg were investigating possible recombination in the bacterium *Salmonella typhimurium.* Although they recovered prototrophs from mixed cultures of two different auxotrophic strains, subsequent investigations revealed that recombination was occurring in a manner different from that attributable to the presence of an F factor, as in *E. coll* What they discovered was a process of bacterial recombination mediated by bacteriophages and now called transduction.

7.5.1 The Lederberg-Zinder experiment

Lederberg and Zinder mixed the *Salmonella* auxotrophic strains LA-22 and LA-2 together and, when the mixture was plated on minimal medium, they recovered prototroph cells. LA-22 was unable to synthesize the amino acids phenylalanine and tryptophan *(phe^- trp^-),* and LA-2 could not synthesize the amino acids methionine and histidine (met^- his^-). Prototrophs (*phe' trp' met' his')* were recovered at a rate of about $1/10^5$ (10^{-5}) cells.

Although these observations at first suggested that the recombination involved was the type observed earlier in conjugative strains of *E. coli,* experiments using the Davis U-tube soon showed otherwise (Figure 7.17). The two auxotrophic strains were separated by a glass-smtered filter, thus preventing cell contact but allowing growth to occur in a common medium. Surprisingly, when samples were removed-foam both sides of the filter and plated independently on minimal medium, prototrophs were recovered only from the side of the tube containing LA-22 bacteria.

Since LA-2 cells appeared to be the source of the new genetic information (*phe*$^+$ and *trp*$^+$), how that information crossed the filter from the LA-2 cells to the LA-22 cells allowing recombination to occur, was a mystery. The unknown source was designated simply as a **filterable agent (FA).**

Three subsequent observations were useful in identifying the FA :

1. The FA was produced by the LA-2 cells only when they were grown in association with LA-22 cells. If LA-2 cells were grown independently and that culture medium was then added to LA-22 cells, recombination did not occur. Therefore, LA-22 cells play some role in the production of FA by LA-2 cells and do so only when the two share common growth medium.
2. The presence of DNase, which enzymatically digests DNA, did not render the FA ineffective. Therefore, the FA is not naked DNA, ruling out transformation.
3. The FA could not pass across the filter of the Davis U-tube when the pore size was reduced below the size of bacteriophages.

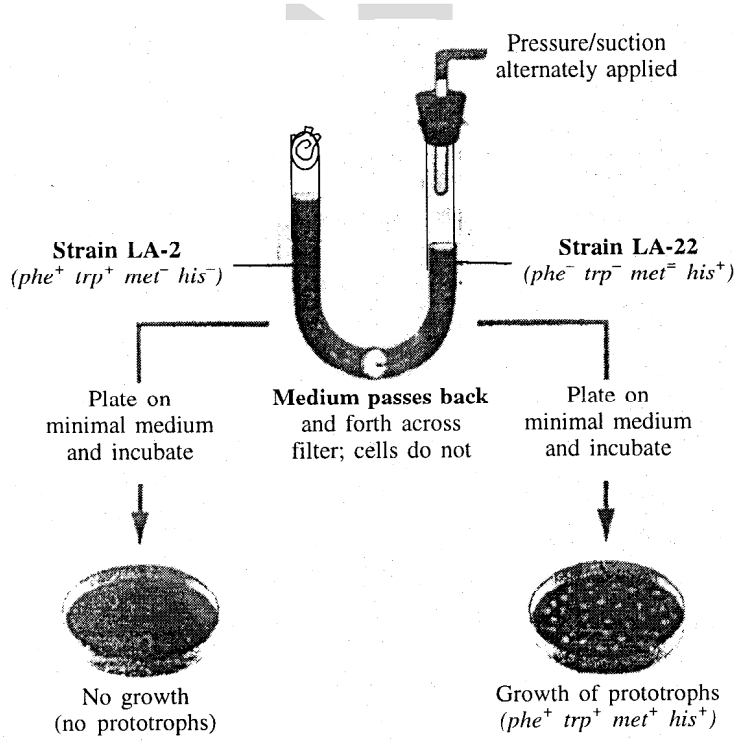


Fig. 7.17 The Lederberg-Zinder experiment using *Solmonelid*. After placing two auxotrophic strains on opposite sides of a Davis U-tube. Lederberg and Zinder recovered prototrophs from the side containing the LA-22 strain but not from the side containing the LA-2 strain. These initial observations led to the discovery of the phenomenon called transduction

Added by these observations and awareof temperate phages that could lysogenize *Salmonella,* researchers proposed that the genetic recombination event was mediated by bacteriophage P22, present initially as a prophase in the chromosome of the LA-22 *Salmonella* cells. It was hypothesized that rarely P22 prophages migt enter the vegetative or Iytic phase, reproduce, and be released by the LA-22 cells. Such phages, being much smaller than a bacterium, were then able to cross the filter of the U-tube and subsequently infect and lyse some of the LA-2 cells. In the process of lysis of LA-2, the P22 phages occasionally packaged in their heads a region of the LA-2 chromosome. If this region contained the *phe*$^+$ and *trp*$^+$ genes, and if the phages subsequenly passed back across the filter and infected LA-22 cells, these newly lysogenized cells would behave as prototrophs. This process of transduction, whereby bacterial recombination is mediated by bacteriophage P22, is diagrammed in Figure 7.18.

7.5.2 The nature of transduction

Further studies revealed the existence of transducing phages in other species of bacteria. For example, *E. coli* can be transduced by phages PI and X. *Bacillus* ***subtilis*** and *Pseudomonas aeruginosa* can be transduced by the phages SPO1 and Fl 16, respectively. The details of several different modes of transduction have also been established. Even though the initial discovery of transduction involved a temperate phage and a lysogenized bacterium, the same process can occur during the normal Iytic cycle. Sometimes a small piece of bacterial DNA is packaged

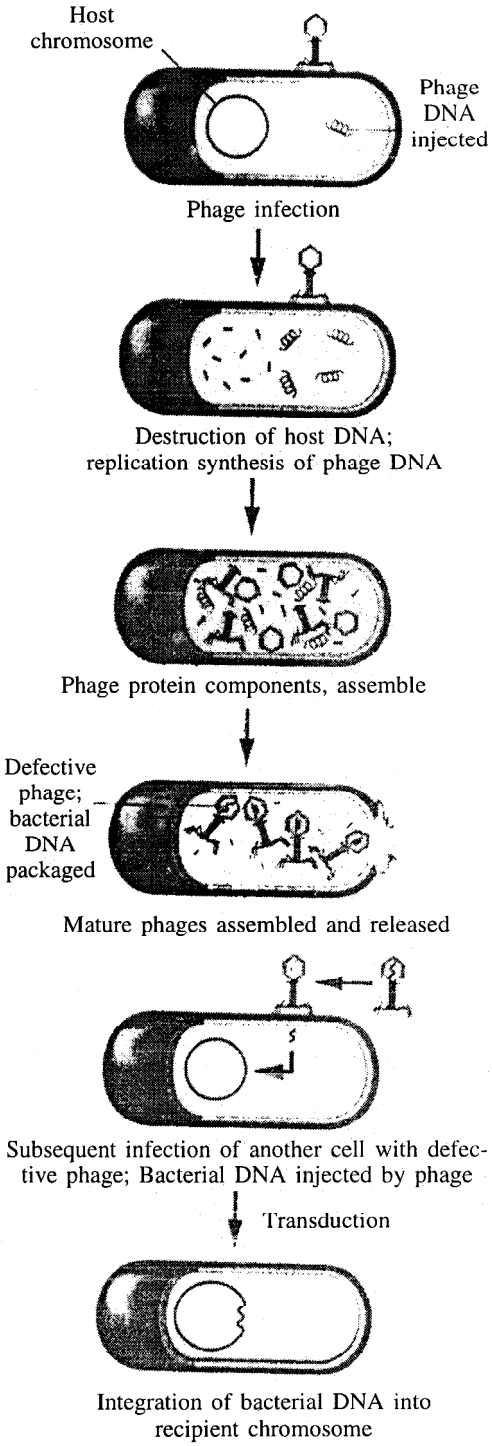


Fig. 7.18 Generalized transduction

along with the viral chromosome so that the transducing phage contains both viral and bacterial DNA. In such cases, only a few bacterial genes are present in the transducing phage. However, when *only* bacterial DNA is packaged, regions as large as 1 percent of the bacterial chromosome may become enclosed in the viral head. In either case, the ability to infect is unrelated to the type of DNA in the phage head, making transduction possible.

When bacterial rather than viral DNA is injected into the bacterium, it can either remain in the cytoplasm of recombine with the homologous region of the bacterial chromosome. If the bacterial DNA remains in the cytoplasm in the cytoplasm, it does not replicate but is transmitted to one of the progeny cells following each division. When this happens, only a single cell, partially diploid for the transduced genes, is produced—a phenomenon called **abortive transduction.** If the bacterial DNA recombines with its homologous region of

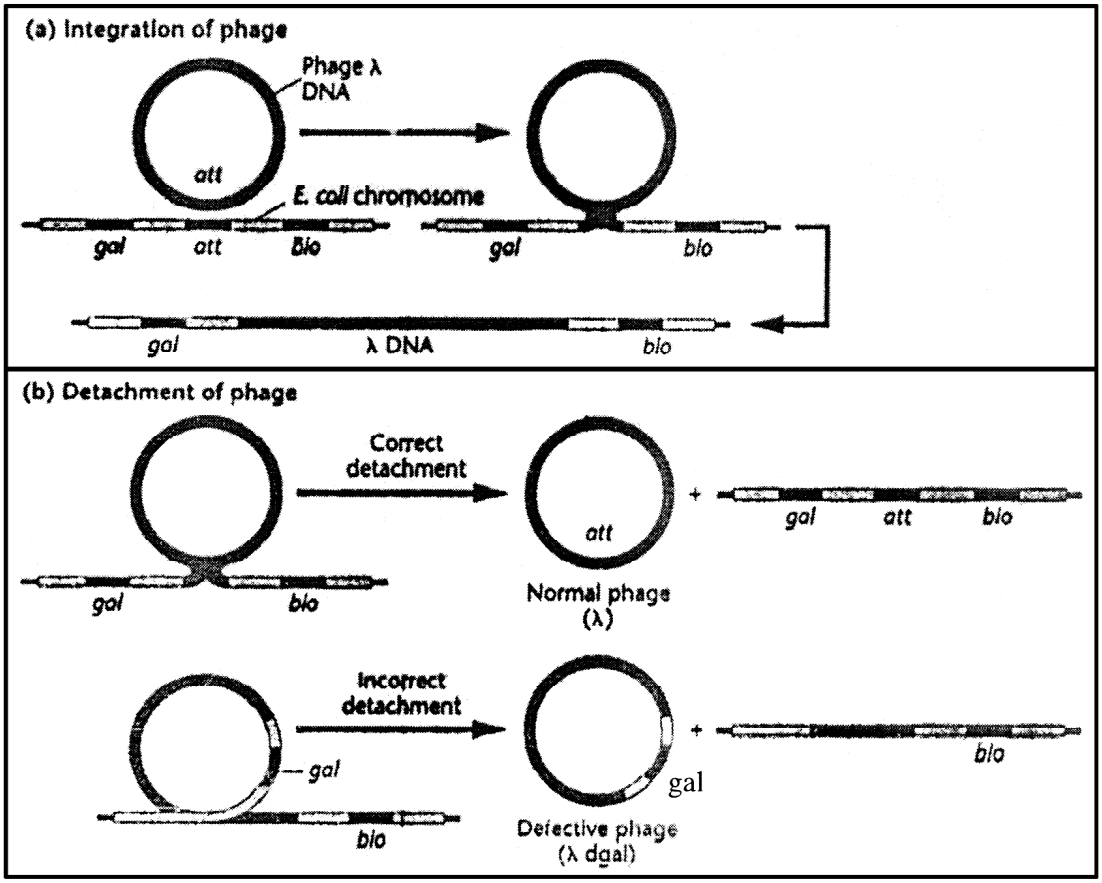


Fig. 7.19 The production of defective phage λ*dgal,* which can result in specialized transduction following another round of infection of *E. coli.* If detachment occurs correctly, no transduction will result

the bacterial chromosome, the transduced genes are replicated as part of the chromosome and passed to all daughter cells. This process is called **complete transduction.**

7.5.3 Transductionand mapping

Like transformation, generalized transduction has been used in linkage and mapping studies of the bacterial chromosome. The fragment of bacterial DNA involved in a transduction event is large enough to include numerous genes. As a result, two genes that are closely aligned (linked) on the bacterial chromosome may be simultaneously transduced, a process called contrasduction. Two genes that are not close enough to one another along the chromosome to be included on a single DNA fragment require two independent events in order to be transdaced into a single cell. Since this occurs with a much lower probability than cotransdoction, linkage can be determined.

By concentrating on two or three linked genes, transduction studies can also determine the precise order of these genes. The closer to each other linked genes are, the greater the frequency of contransduction. Mapping studies involving three closely aligned genes can be executed. The analysis of such an experiment is predicated on the same rationale underlying other mapping techniques.

7.5.9 Specialized transduction

In some instances, only certain genes are recombined, a situation called **specialized transduction.** This is in contrast to generalized transduction described above, where all genes have an equal probability of being recombined. One of the best examples involves transduction of *E. coli* by the temperate phage λ.

In this case, transduction is restricted to the *gal* (galactose) or *bio* **(biotin)** genes. The reason why transduction in volves only these genes became clear when it was learned that the λ DNA'always integrates into the region of the *E. coli* chromosome between these two genes at a site called *att* [Figure 7.19(a)]. Phage λ DNA that is integrated in the bacterial chromosome can subsequenly detach from it, reproduce, and lyse the host cell [Figure 7.19(b)]. Sometimes the excision process occurs incorrectly and carries either the *gal* or *bio E, coli* genes in place of part of the viral DNA [Figure **7.19(b)].** The resulting phage chromosome is defective because it has lost some of its own genetic information, but it is nevertheless replicated and packaged during the formation of mature phage particles. The virus can subsequently inject the defective chromosome into another bacterial cell.

In this case of specialized transduction, the defective phage chromosome

is integrated into the bacterial, chromosome and is replicated along with it. Such bacterial cells contain the defective phage DNA, making them diploid for the *gal* or *bio* genes. The presence of this transducing *gal′* or *bio′* DNA causes these auxotrophs to revert to a *gal′* or *bio′* phenotype.

7.5.5 Bacteriophages undergo intergenic recombination

Around 1947, several research teams demonstrated that genetic recombination also occurs in bacteriophages. These studies relied on the discovery of numerous phage mutations that could be visualized or assayed. Before considering recombination in these bacterial viruses, we will briefly introduce several of the mutations that were studied.

Phage mutations often affect the morphology of the plaques formed following lysis of bacterial cells. For example, in 1946 Alfred Hershey observed unusual T2 plaques on plates of *E. coli* strain B. Where the normal T2 plaques are small and have a clear center surrounded by a diffuse (nearly invisible) halo, the unusual plaques were larger and possessed a more distinctive outer perimeter (Figure 7.20). When the viruses were isolated from these plaques and replated on *E. coli* B cells, the resulting plaque appearance was identical. Thus, the plaque phenotype was an inherited trait resulting from the reproduction of mutant phages. Hershey named the mutant *raid lysis* (r) because the plaques were larger, apparently resulting from a more rapid or more efficient life cycle of the phage. It is now known that in wild-type phages, reproduction is inhibited once a particular-sized plaque has been formed. The mutant T2 phages are able to overcome this inhibition, producing larger plaques.

Unit 8 Cytogenetic effects of Ionizing and Non-ionizing Radiations

Structure

8.1 Introduction

Various agents shown to cause chromosome breaks have been termed "clastogens" by Shaw (1970). These include physical agents (X-rays, ultraviolet light, cold shock, magnetic fields, and sound waves), biological agents (certain genes, viruses and protozoa), and a host of chemical agents. It should be emphasized that most of these clastogens produced these effects *in vitro* by the addition of the agent to cultured lymphocytes and/or fibroblasts for varying times and concentration. In but few cases is there evidence for *in vivo* chromosome breakage.

8.2 Radiation

Survivors of the atomic bomb blasts in Japan have developed leukemia in proportion to the amount of radiation received. Furthermore, increased numbers of chromosome breaks and rearrangements have been found in lymphocytes of nonleukemic survivors (Bloom *et al.,* 1967). Similar anomalies (translocations and inversions) have been demonstrated in lymphocytes of individuals who have received X-ray therapy to the spine or injections of thorotrast (Buckton *et al.,* 1982; Court Brown *et al.,* 1967).

Although ultrasound can effect chromosomal breaks *in vitro,* there is no evidence that it does *so in vivo* (Macintosh and Davey, 1972).

Fibroblasts cultured from skin in the path of X-radiation have manifested chromosome abnormalities (Engel *et al.,* 1964; Visfeldt, 1966). Marrow cells may exhibit abnormalities even after many years following primary exposure (Goh, 1971). Leukemia is also more likely to develop in individuals who have received chronic exposure to radiation (Lewis, 1970). Maternal irradiation before and during the reproductive period increases the incidence of chromosomally abnormal conceptuses. However, most are nonviable and lost early in pregnancy (Alberman *et al,* 1972).

8.3 Viruses

There is insufficient evidence currently available to directly implicate viruses in effecting human chromosome abnormalities. However, a number of investigators have studied the effect of S V40 virus on cultured human fibroblasts and have observed altered growth patterns to the haphazard growth and alterations, within several months an emergence of a few stable heteropliod cells becomes evident (Moorhead, 1970).

Numerous other viruses (Rous sarcoma, vaccinia, rubella, herpes zoster, poliomyelities, influenza, polyoma, etc.) have been shown to produce chromosome breaks in infected cells *in vitro.* Furthermore, several viruses can produce abnormalities in metaphase chromosome in circulating lymphocytes during natural human infections (measles, chicken pox, mumps, and hepatities) (Moorhead, 1970). The effects have been of at least three types: single breaks, pulverization, and fusion and spindle abnormalities. The mechanism is unknown but may be related to addition of the virus like particles (Epstein-Barr virus) have been demonstrated, often exhibit a long submetacentric marker (Gripenberg *et al.,* 1969).

8.4 Chemical clastogens

Over 200 drugs or chemical shown to cause chromosome breaks *in vitro* (shaw, 1970) can be grouped into several categories with a few illustrations in each group: (a) nucleic acid related compounds (6-mercaptopurine and 5-fluorodeoxyuridine), (b) antibiotics (mitomycin C, streptomycin, actinomycin D and daunomycin), (c) central nervous saystem drugs (meprobamate, chlopromazine, mesealine, lysergic acid diethylamide, and scopolamine), (d) food derivatives and additives (caffeine, cyclamate, theobromine, and theophylline), (e) air and water pollutants (chloramine T and ozone), (f) pesticides (captan and thioTEPA), (g) alkylating agents (nitrogen mustards and cytoxan), (h) mitotic poisons (colchicine), (i) photodynamic dyes (acridine orange and neutral red), (j) antifolic compounds (methotrexate and aminopterin), (k) organic solvents (benzene and mercaptoethanol), (1) inorganic substances (lead and arsenic), and (m) miscellaneous comopounds (Imuran and piperazine).

However, it should be emphasized that few chemical clastogens have been implicated in chromosomal breakage *in vivo.* To cite but a few examples: Ambient exposure to benzene has been noted to be associated with both chromosome breakage and subsequent development of leukemia (Tough and Court Brown, 1965; Hartwich *et al.,* 1969). On the other hand, LSD, while producing chromosome breaks *in vitro,* has not been shown to be effective *in vivo* (Stenchever and Jarvis, 1970).

Unit 9 Molecular Cytogenetic Techniques

Structure

9.1 FISH, GISH

9.1.1 *In situ* hybridization technique

Under normal temperature and ionic conditions DNA remains in a duplex state by the base pairing through the hydrogen bonds. By heating in buffer solution or by increasing pH the two strands can be separated. But if again the temperature is lowered or pH is reduced then the separated strands will join again and reassociate.

This fact was shown by Julius Marmur and Paul Doty in 1960. This type of reassociation of DNA strands in called molecular hybridization or nuclear hybridization. It may also take place between the complementary strands of DNA or RNA or between DNA and RNA.

Hybridization technique which can be used to localize specific nucleic acid fragments that reside in their original site *(in situ)* within cells, is known as *in situ* hybridization.

In situ hybridization is a version of hybridization analysis, in which an intact chromosome is examined by probing it with a labeled DNA molecule.

For this technique to work, DNA in the chromosome must be made single stranded and denatured by breaking the hydrogen bonds between the base pairs. Only then the DNA or RNA probe can be hybridizid with the chromosomal DNA. For this purpose, without destroying the chromosome morphology a dry preparation is made into a glass microscope-slide and then treated with formaldehyde.

The position where the hybridization occurs provides the information about the map location of this gene, thus physical mapping of chromosome could be done by this method.

To locate the region of hybridization two types of markers viz., radioactive marker and fluorescent marker are used. Besides these other types of markers are also used for this purpose.

To prepare a DNA profile, the nucleotides are synthesized in which one of the phosphorus atoms replaced by 32_P or 33_P. One of the oxygen atom in the phosphate group is replaced with 35_S or one or more of the hydrogen atom is replaced with 3_H. Radioactive nucleotides act as substrates for DNA polymerases and so are incorporated into a DNA molecule by any strand synthesis reaction catalyzed by a DNA polymerase.

The labeled nucleotides or individual phosphate groups can also be attached to one or both ends of a DNA molecule by the reaction catalyzed by T4 polynucleotide kinase or terminal deoxy nucleotidyl transferase.

The rodioactive signal can be detected by scintillation counting, but for most molecular biology, the position of the hybridization is detected by exposure of an X-ray sensitive film (autoradiography) or radiation sensitive phosphorescent screen (phosphorus imaging). 32_P has a high emission energy and the resolution is lower. But low emission isotopes such as 35_S or 3_H give less sensitivity but greater resolution.

9.1.2 Fluorescent *in situ* hybridization (FISH)

Concept

To solve these problems in 1980's non radioactive fluorescent DNA labels were developed. These labals combine high sensitivity with high resolution and are ideal for *in situ* hybridization. The different fluorolabels of different colours have been disigned for the probes and it is possitive to hybridize the chromosome and the individual hybridization signals enable the location of relative position of the probe sequence to be mapped.

Fluorescent dyes (Fluorochromes) live quinacrine (Q) and quinacrine mustard (QM) are used to obtain specific patterns of cross striations or bands appear with alternate fluorescent and non fluorescent bands. The bands obtained with quinacrine are called Q bands while those obtained with quinacrine mustard are called Q M-bands. Bands are obtained when fluorescent deyes attach to the specefic regions of the chromosomes (See Fig. 9.2).

Technique

During the *in situ* hybridization a sample of dividing cells is dried on a microscope slide and treated with formaldelyde so that the chromosomes become denatured but do not lose their characteristic metaphase morphologies. The probe is added to the denatured chromosomes, which will be hybridized with the complementary DNA region of the chromosome. The position at which the probe hybridizes to the chromosomal DNA is visualized by detecting the fluorescent signal emitted by the labeled DNA.

Problem

If a probe be a long fragment of DNA, then a potential problem is that it is likely to contain repetitive DNA sequences and so may hybridize many portions of the chromosome. If these sequences are not blocked then the probe will hybridize non specifically to any copy of these genome and will repeat in the target DNA. To block the repeat sequences, the probe is pro-hybridized with a DNA fraction enriched for repetitive DNA.

Advanteage of FISH over other methods

Health and environmental issues have meant that radioactive markars have become less popular in recent years and they are now largely superseded by non-radioactive alternatives.

Other drawbacks of radioactive markers in *in situ* hybridization and that the radioactive label has high emission energy (eg., P^{32}) and then it scatters its signal and so gives poor resolution. On the other hand , if H^3 is used, the emission energy is less but its sensitively is so low that lengthy exposures are needed.

Disadvantage

The metaphase chromosomes are highly condensed and a fluorescent signal obtained by FISH is marked by measuring its position relative to the end of the short arm of the chromosome (the flepter value). The two markers having at least 1Mb apart to be resolved as separate hybridization signal (Trask *et al*, 1991). Therefore, using FISH, the highly condensed nature of mataphase chromosomes means that only low resolution mapping is possible. Therefore, FISH provides a fough idea of its map position.

More advanced FISH

In 1996 Heiskanen *et al* solved the problem of FISH and a range of higher resolution FISH technique has been developed. Higher resolution is achieved by changing the condensing pattern of the metaphase chromosome. There are two ways of doing this.

(1) Mechanically stretching of metaphase chromosomes : Centrifugation generates shear forces, which can result in the chromosomes becoming stretched upto 20 times of the normal length. Thus the resolution is significantly improved and markers that are only 200-300 Kb apart can be distinguished.

(2) Taking non-metaplase chromosomes : Attempts have been made to use prophase nuclei because in this stage the chromosomes are not still sufficiently condensed for individual ones to be identified and so provides no advantage.

Interphase chromosomes become more useful because, then the

chromosomes again become less condensed. Using anaphase stage the resolution down to 25 kb is possible but their chromosome morphology is lost, so there are no external reference point against which the position of the probe to be mapped. This technique is used to construct map in small region of the chromosome after obtaining preliminary information.

Fibre-FISH

Interphase chromosomes are most unpacked of all cellular DNA. To improve the resolution of FISH better than 25 kb, it is therefore, necessary to abandon intact chromosomes. This approach is called fiber FISH. In this approach the DNA is prepared by gel stretching or molecular combing. This can distingish markers that are less than 10 Kb.

To carry out gel stretching (Fig. 9.1a), molten agarose containing chromosomal DNA molecules is pipetted into a microscope slide, coated "with a restriction enzyme (Schwartz *et al,* 1993). As the gel solidifies, the DNA molecules become stretched. It is not understood why this happens but it is thought that fluid movement on the glass surface during gelation might be responsible. Once the gel is solidified it is washed with $MgCl_2$ solution, which activates the restriction enzyme. A fluorecent dye such as DAPI (4,6 diamine 2-phenylindolo dihydrochloride) is added which stains the DNA so that the fibres can be seen when the slide in examined with a high power fluorescence microscope. The restriction enzyme cuts the DNA molecule. As the molecules gradualy coil up, the gaps representing the cut sites become visible. The relative positions of the cuts are to be recorded.

In molecular combing (Michalet *et al,* 1997) (Fig. 9.1b), the DNA fibres are prepared by dipping a silicon-coated coverslip into a solution of DNA. It takes 5 minutes to attach DNA to the coverslip by their ends. After that coverslip is removed at a constant speed of 0.3 mm/sec. The force required to pull the DNA molecule through the miniscus causes to line up. Once in the air the surface of the coverslip dries, DNA molecules are arranged as a parallel fibre thus producing a comb of parallel molecules.

9.1.3 Biotin-labelling *in situ* hybridization

Recent advances in nucleic acid technology offer alternative to radioactive-labelling probes. One procedure that is becoming increasingly popular is biotin-labelling of nucleic acid. This is nontoxic, whose half life is longer and can be prepared in adavnce in bulk and stored at -20°C for (repeated) use.

Drosophila salivary gland chromosomes can be hybridized with a biotin labelled nucleic acid probe. After washing, detection can be done by adding a biotin-binding protein called ovidin which is covalently bound to alkaline

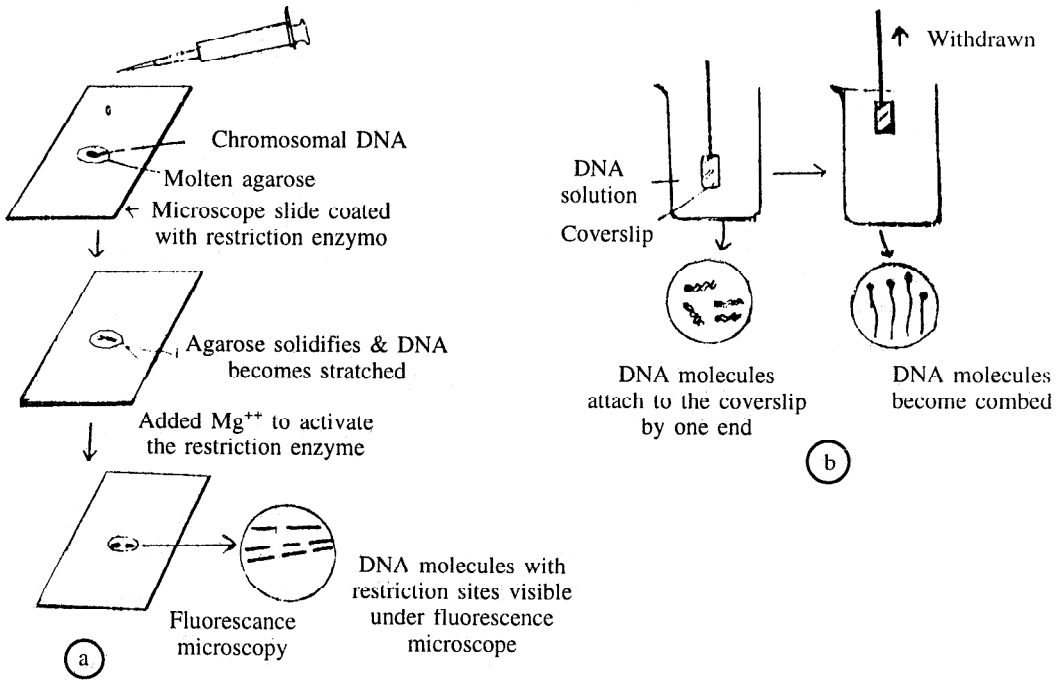


Fig. 9.1 (a) Gel stretching and (b) molecular combing in fibre FISH

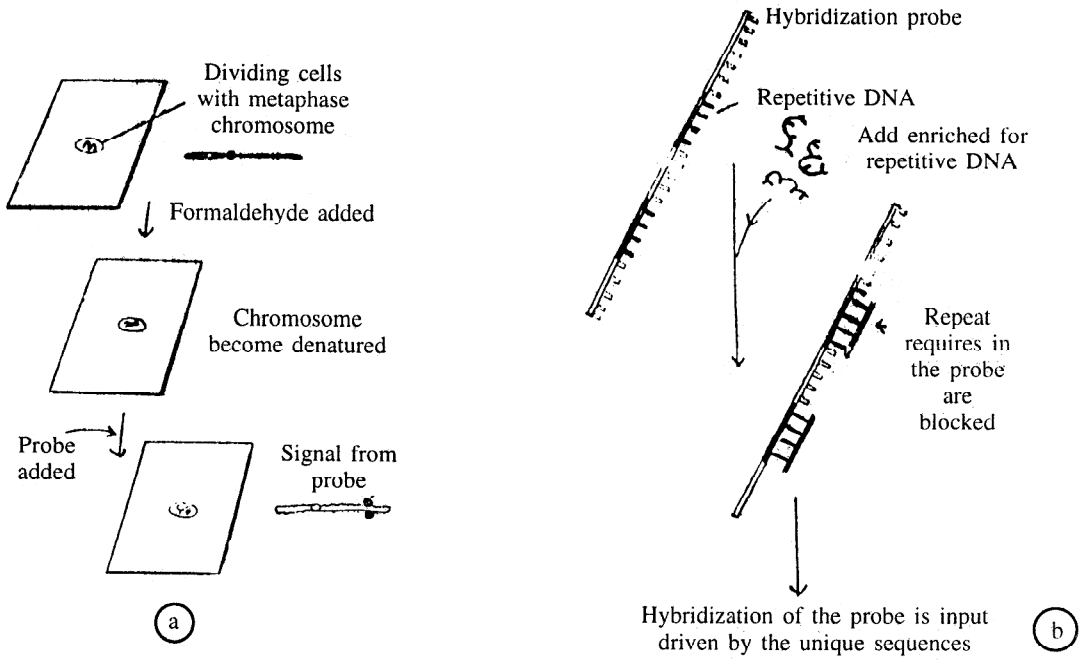


Fig. 9.2 (a) FISH ; (b) Blocking of repetitive DNA sequences in a hybridization probe

phosphatase. After addition of a soluble substrate, the enzyme catalyses a reaction that results in formation of an insoluble blue coloured precepitate at that site of hybridization. The intensity is proportional to the amount of biotin in the hybrid.

In situ hybridization with a biotin labelled probe has been particularly useful in chromosome mapping of DNA clones in *Drosophila* because the logical map of the poly-tene chromosomes of this organism is known at high resolution.

9.2 DNA finger printing

The techniques of DNA diagnosis have found application in a quite different area, the identification of medicine. This is important in areas as diverse as identifying cell cultures, determining family relationships in studies of animal behaviour, immigration problems to identify criminals or murderer, disputed parentage and in forensic medicine.

The most accurate method of identification technique based on recombinant technology is called DNA finger printing or DNA typing or DNA profiling.

Principle

DNA finger printing is based on **sequence polymorphism** that occurs in human genome and the genome of other organisms. The sequence polymorphisms are slight sequence differences, usually single base pair changes, that occur from individuals to individuals once in every few hundred base pairs on average. Each difference from the consensus human genome sequence is generally present in only a fraction of the human population but every individual has some of them. These polymorphic locus is called minisatellite or VNTR (variable numbers tandem repeat) thus forming a haplotype which shows mendelian inheritance among their offsprings. This locus is made up of a variable number of identical sequences joint together in tandem.

One family of minisatellites in the human genome share a common "Core" sequence, The core is G-C rich sequence of 10-15 bp showing on asymmetry of purine/ pyrimidine distribution on the two strands. These repeats are written as (C-A) n (G-T)n, occur in 100000 blocks in every genome and appear to be uniformly distributed throughout the genome (value of n varies from 1 to 40). The successful application of these (C-A)n (G-T)n repeats has led to the use of a variety of other di-, tri- **and** tetranucleotide sequences for mapping.

Technique

(1) DNA of the sample is first isolated whose DNA finger printing has to be made. Usually in forensic case the DNA **is** prepared from dried stains, sperms

in vaginal swabs that had been stored for as long as two years. Sufficient DNA can be isolated from freshly pulled hair roots, polymorphism in mitocondrial DNA and class II HELA gene DQ have been analysed from the shed hairs of several months old containing less than 1 ng of DNA.

(2) Sufficient quantities of intact DNA from forensic samples will always be a problem. PGR may have a great impact in this area. From a small quantities of DNA PCR can produce a large number DNA. These are used as probes. This probe is more redioactive.

(3) The DNA from the individual whose DNA is to be compared with the forensic sample is isolated and are cut into fragments by restriction enzymes.

(4) DNA fragments after digestion of DNA from the genome are first separated according to their size by agarose gel electrophoresis. These ds DNA are denatured by soaking the gel in alkali to make it ss DNA.

(5) The DNA fragments are transferred to nito-cellulose paper by the southern blot technique. The paper is then immersed in a solution containing a radioactively labeled DNA probe. Fragments to which the probe hybridizes are revealed by autoradiography.

Detection of forensic problem :

The power of DNA fingerprinting was demonstrated by Alec Jeffreys in 1985, when a man had been aceured of two rape murders committed three years apart and had made a confussion. Lastly the real murderer was caught and DNA finger printing confirmed the identification.

DNA from a semen sample obtained from a rape and murder victim was analysed along the DNA samples from the victim and two suspects.

Each of the DNA samples was cleaved into fragments and separated by gel electrophoresis.

Radioactive DNA probes were used to identify a small subset of these fragments that contained sequences complimentary to the probe. The sizes of the fragments identified varied from one individual to the next. The different patterns for the three individuals (victim and two suspects) tested. One rape suspects DNA exhibits a banding pattern identical to that of the semen sample taken from the victim.

More than one probe may be used to make a positive identification.

9.3 Separation of chromosomes by flow cytometry

The dividing cells with condensed chromosomes are carefully broken open so that a mixture of intact chromosomes is obtained. The chromosomes are then stained with fluorescent dye. The amount of dye that a chromosome binds

depends on its size. Thus larger chromosomes bind more dye and fluoresence more brightly than smaller ones. The mixture of chromosomes is diluted and passed through a fine aperture, producing a stream of droplets, each one containing a single chromosome. The.droplets are passed through a detector that measures the amount of fluorescence and thus identifies which droplits contain the particular chromosome (being sought). An electric change is applied to these droplets by a charger and then the droplets reach the electric plates, the changed ones are deflected into a separate beaker.

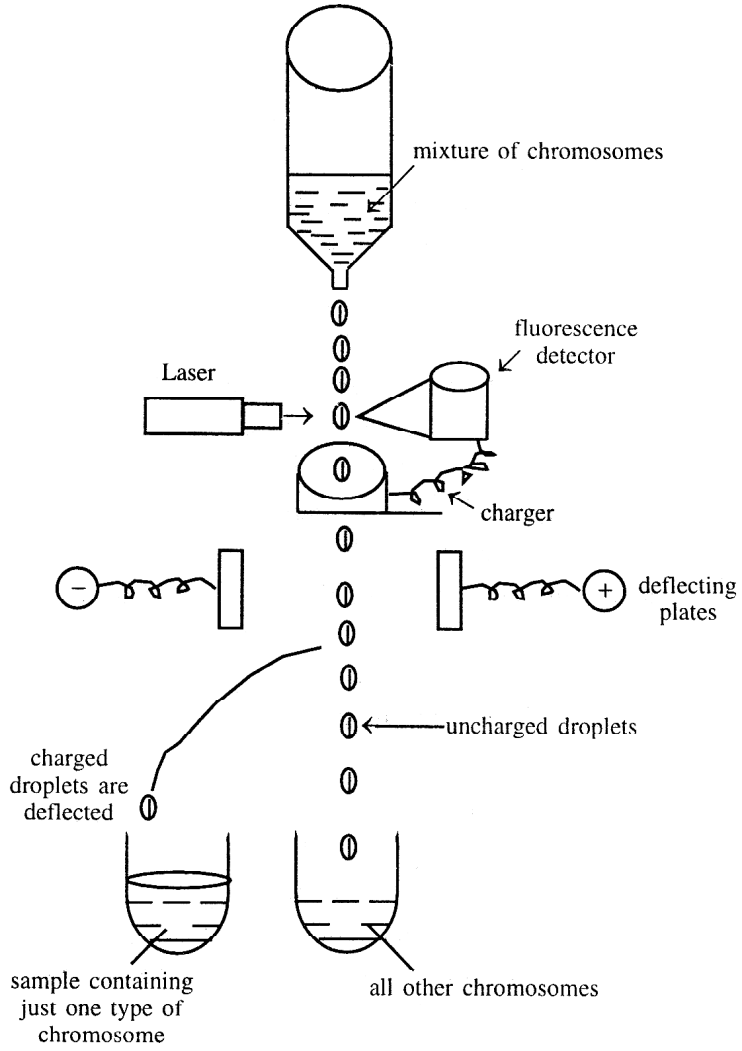


Fig. 9.3 Flow cytometry

If two chromosomes are of equal size as in human chromosome number 21 and 22, then the dye, tioechst 33258 and chlomomycin-A preferably bind to the A-T and G-C rich DNA respectively and by this differential staining activity, these chromosomes can be distinguished properly (Fig. 9.3).

9.4 Chromosome painting

This is a method for visualizing each of the chromosomes in distinct bright colours and thus it simplifies greatly the distinction between chromosomes of similar size and shape and the karyotyping of the chromosome. Usually such a painting is done at the stage in the cell cycle (mitosis) when chromosomes are specially compact and easy to visualize, usually at mitosis. But sometimes, selective chromosome painting may be done in the interphase stage to see its orientation in the nucleus.

Method

(1) The probes which are used for chromosome painting are specific for sites scattered along the length of each chromosome.

(2) The probes are labeled with one of two dyes that fluorence at different wave lengths. For example, DNA molecules derived from chromosome-1 are labeled with one specific dye combination, chromosome-2 with another and so on.

(3) The labeled probe can hybridize only the chromosome from which it was derived, each chromosome is differently labeled. After the probes are hybridized to chromosomes the excess is removed, the sample is placed in a fluorescent microscope in which a detector determines the fraction of each dye present at each fluorescing position in the microscopic field. The information may be conveyed, to a computer and a special program assigns a false colour image to each type of chromosome.

Use

Chromosome painting is very useful in karyotyping the chromosomes. It can be done in interphase stage to locate the specific chromosomes and their arrangement in the nucleus. A combination of chromosome banding with FISH, called multicolor FISH can detect chromosomal translocations which are associated with certain genetic disorder and specific types of cancers. For example, in chronic myeloid leukemia (CML), the lymphocytes contain the Philadelphia chromosome(small), which is produced by the translocation of chromosome no-22 and chromosome no-9. The translocations can be detected by classical banding analysis technique.

Unit 10 Genome Analysis

Structure

10.1 C-value paradox

The haploid DNA content in an individual is described as its C-value.

The anomalies of the gene contents by two different methods one on the basis of knowledge about the rate of mutation per locus and other on the basis of general method used for DNA content, is called C-value paradox.

DNA content in eukaryotic cells are much higher than that in the prokaryotic cells and a wide range of variations are observed among different species even among same species. The content of DNA also depends on the number of chromosomes in the cells (i.e., ploidy of the chromosome). Example : See Table 10.1.

Table 10.1 DNA content of some organisms

Class of organisms	Species	**Haploid DNA (Picogram)**	**Dalton**	Base pairs
Phages	∅ × 174 T4	2.6×10^{-6} $20.7 \times lO^{-5}$	1.7×10^{6} 126×10^{6}	5400 200000
Animal virus	Adenovirus	21.7×10^{-6}	13×10^{6}	21000
Prokaryotes	*E. coli*	4.4×10^{-3}	2.7×10^{9}	$4.2 \times I0^{6}$.
Unicellular-eukaryotes	*S. cerevisial*	$14 \times lO^{-3}$	8.5×10^{9}	$1.4 \times I0^{7}$
Multicellular enkaryotes	*D. melanogaster*	0.18	0.11×10^{12}	0.17×10^{9}
	Homo sapiens	2.8	18×10^{12}	2.8×10^{9}

C-value paradox takes its name from the inability to explain the content of a genome in terms of an anticipated function. There are two aspects of the

paradox. First there are huge variations in C-values between certain species whose apparent complexity does not vary correspondingly. There can be rather substantial variations even between certain closely related species.

The range of C-values is found in different evolutionary phyla. There is some increase in the minimum genome size that is found in each phylum as the complexity increases.

For example, in prokaryotes, the genome size are very small. In eukaryotes, a vast increase occurs in genome size. In yeast, *Saccharomyces cerevisiae* has a genome size of 2.3 × $]0^7$ bp, only 5 times greater than that of *E. colt.*

The modest increase in genome size just over two folds is adequate to support the slime mold, D *discoideum,* able to live in either unicellular or multicellular modes.

Another increase in complexity is necessary to produce the first fully metazoan organisms. For example, *C. elegans* has a DNA content of 8×10^7 bp. Then any close relationship between complexity of the organisms and content of DNA is obscure, although it is necessary to have a genome of more than $I0^8$ bp, to make an insect of more than 4×10^8 bp, to assemble an echinoderm of more than 8×10^8 bp, to produce a bird or amphibians and more than 2×10^8 bp to develop into a mammals.

In some cases the spread of genome size is quite small. For example, birds, reptiles and mammals, all show a little variation within the phylum, with a range of genome size in each case about two fold. But in other cases, there is quite a wide range of values, often more than 10 folds. This reflects some surprizing discrepencies between genome size and complexity of the organism.

An extraordinary C-value is found in amphibian where the smallest is below 10 bp while the largest are almost 10^{11} bp. It is hard to believe that this could reflect a 100 fold variation in number of genes in different amphibians.

There are some cases where rather closely related species show surprising variations in total genome size. For example two amphibian species may have 10 fold increase where morphologies are very similar. Yet if the gene number is roughly similar most of the DNA in the species with the larger genome cannot be concerned with coding for protein. So the question, what could be its function?

The second aspect of C-value paradox is the apparent absolute excess of DNA compared with the amount that could be expected to code for proteins.

Actually, eukaryotic DNA has an excess length and the excess is encountered because genes are much larger than the sequences needed to code for proteins. For example, human haploid cell has DNA amount equal to 1.8 x $]0^{12}$ dalton = 87 cm of DNA which is equal to 2.8×10^9 base pairs, then this genome could contain approximately as many as 3×10^6 genes assuming 1000 bp per gene coding for nearly 300 amino acids.

However, the number of genes estimated in humans on the basis of the rate of mutation per locus, as estimated by Muller (1967), the frequency of deleterious mutations per locus in human is 10^{-5} to 10^{-6} in each generation. If the number of gene is 3×10^6 (as calculated in general method), then it will yield 30 deleterious mutaions in each generations at the rate of 10^{-5} mutations per locus. This will be an unbearable genetic load. The actual frequency of deleterious mutations per generation per individual has been estimated to be 0.5 against an expected frequency of 30. This mutation frequency at the rate of 10^{-5} per locus will be an estimate of 5×10^4 genes.

Thus it is supposed that the actual number of genes in human should be 50000 and not 3 million an estimated from DNA content.

This anomalous situation has been described by some workers as C-value paradox.

10.2 Satellite DNA

Large proportions of DNA in eukaryotes has been shown to be present in the form of multiple copies of identical DNA sequences, thus is called repetitive DNA or Satellite DNA. The remaining DNA in the cell is found in the form of single copy of DNA sequences which is known as unique DNA.

When the denatured DNA (single stranded) is led to reassociate, then it was observed that from the heterogenous populations, the smaller molecular weight DNA associate easily. Britten and his Coworkers (1966, 68) have demonstrated that many vartebrate DNAs reassociates easily when it is broken into small pieces. This observation gave rise to the hypothesis that certain short sequences of bases are repeated hundred times in DNA, this is the **satellite DNA.**

This repetivive DNA generally contributes at least 20% DNA and can reach upto 90% in some cases. It is believed that these repetitive sequences do not carry any genetic informations and therefore do not form genes, but play some other structural or regulatory role.

Repetitive DNA consists of short identical genes which are repeated in tandem, several hundred or thousand times. Such DNA is found in the region of the chromosome adjacent to the centromere. In many case the base compositions of the repeating sequences are unlike that of the rest of the DNA. It is, therefore easy to separate repetitive DNA by ultracentrifugation.

The satellite DNA can be isolated by density gradient centrifugation in neutral caesium chloride as they have distinctive bouyant densities. The fractions can be separated as a band from the main band of DNA, this band is called satellite band.

The satellite band in found on the left of the main band if lighter and on the right side if havier than the DNA of the main band.

In *Drosophila virilis,* there are three highly repetitive DNA each consisting of a repeating sequence of seven nucleotide pair and about 25%. of the DNA is satellite DNA (See Table 10.2).

Table 10.2 Satellite DNA of *Drosophila virilis* (Gall *et al.,* 1974)

Repetitive DNA	Bauyont density	Repeat sequence
Sat DNA-I	1.692	5' ACAAACT 3' 3' ṪĠṪṪṪĠȦ5'
Sat DNA-II	1.688	5' ATAAACT 3' 3' ṪȦṪṪṪĠȦ 5'
Sat DNA-III	1.671	5' ATAAATT 3' 3' ṪȦṪṪṪȦȦ 5'

In human, 30% of DNA is repetitive and in designated as sat I, II and III.

In mouse, 10% of the DNA is highly repetitive and renaturates within a few second, 20% is moderately repetitive and reassociates at an intermediate rate, 70% is-single copy DNA which renaturates very slowly. There are about a million copies of repetiting sequences of about 300 bp.

In prokaryotes, the repeated base sequence in not found.

Two remarkable features of satelite DNA are—

(I) Remakable (relative) uniformity within the same species.

(II) Great variability between closely related species.

The satellite DNA often lies in heterochromatic region of chromosomes and its location can be demonstrated by cytological hybridization by incubating the cells in the radioactive solution and is determined by a autoradiography.

The function of highly repetitive DNA is unknown. This can replicate but cannot transcribe RNA for protein synthesis. This is probably because the short sequences lack promoter sites on which RNA chains can be initiated by RNA polymerase. Repetitive DNA is, therefore, inert and is partly dispensible.

In the African clawed toad, *Xenopus laevis,* the genes for 40s precursor RNA which give rise to 28s and 18s RNA are repeated about 450 times. The genes are tandemly arranged and are separated by a spacer region of about 5000 bp. Genes for 5s **r** RNA are also separated by spacer regions and are arranged

in clusters of 100 to 1000 repetitive units at the ends of the most of the 18 chromosomes.

10.3 Complexity

Complexity is the total length of different sequences of DNA present in a given proparation. The double stranded DNA is denatured and converted into single stranded DNA by heating the DNA solution. This is accompanied with increase in optical density, which is called hyperchromicity.

Again when it is allowed to cool, the single stranded DNA is transformed into a double stranded DNA, again the optical density is decreased, this is called hypochromicity. The 50% resaturation is achieved usually at a specifie temperature which is called melting temperature (Im). The formation of double stranded DNA **is** actually measured over different values of a parameter which is described as C_0. t (concentration x time). It is the product of DNA concentration and time of incubation in a reassociation reaction.

Complexity of the genome can be described under two heads—

A. Kinetie complexity

The reassociation of DNA in the solution depends on the random collisions betwe en the complementary strands, which follow the second order of **kinetics,** since concentration of both the complementary strand will influence the rate of reaction. The rate of reaction, when expressed through differential calculus **is** as follows :—

$\frac{dc}{dt} = -Kc^2$ where, C = Concentration of single stranted DNA at time 't'

K = reassociation rate constant

or $\int \frac{dc}{c^2} = -K \int dt$

or, $-\frac{1}{C} = -Kt + A$ A is constant

or, $\frac{1}{C} = Kt - A$

When t = o, then C = C_0

Now, $\frac{1}{C_0} = KXO - A$

or, $A = -\frac{1}{C_0}$

The equation is

$$\frac{1}{C} = Kt + \frac{1}{C_0}$$

When the reaction is half complete then time is ($t_{1/2}$) and in $t_{1/2}$ time the concentration is C. Then—

$$\frac{1}{C} = Kt_{1/2} + \frac{1}{C_0}$$

or, $\frac{C_0}{C} = k.C_0.t_{1/2} + 1$

We know $C = \frac{1}{2}C_0$

then, $1 + K.C_0.t_{1/2} = 2$

or, $K = \frac{1}{C_0.t_{1/2}}$

or, $C_0.\ t_{1/2} = \frac{1}{k}$

So, during reassociation of DNA occurs at the rate constant K (nt. moles/lit/see) in equal to the receprocals of $C_0.t_{1/2}$

$C_0.\ t_{1/2}$ is the product of DNA concentration and time (t) required to proceed to half completion of the reaction; it is directly proportional to the unique length of reassociating DNA.

The $C_0.t_{1/2}$ of a reaction indicates the total length of different sequence that are present. This is described as the complixity. It is usually given in base pairs, but can be expressed in daltons or any other mass unit.

A higher $C_0.\ t_{1/2}$ means, slower reaction and lower $C_0.\ t_{1/2}$ means faster reaction.

If there is no repetitive DNA (because the repetitive DNA reassociates faster), the $C_0.\ t_{1/2}$ of a reaction will be directly prortional to the DNA content. In view of this $C_0.\ t_{1/2}$ will indicate the length of all the different sequences in a genome, which will be less than the length of the total DNA in a genome when

there is repetition. This is called kinetie complexity.

Kinetie complexity is the complexity of a DNA component measured by the kineties of the DNA association.

This can be calculated by knowing the $C_0 . t_{1/2}$.

For example, *E. coli* has a genome = 0.004 pg DNA = 4.2×10^6 bp with $C_0 . t_{1/2} = 4$.

The Kinetie complexity of the genome

$$= \frac{C_0 . t_{1/2} \text{ of DNA of any genome} \times \text{base pair}}{C_0 . t_{1/2} \text{ of } E.coil \text{ genome}}$$

$$= \frac{4.2 \times 10^6 \times C_0 . t_{1/2} \text{ of the genome}}{4}$$

In eukaryotes, the genome contain more than one pure components. For example. calf DNA has two component, each with characteristics $C_o.t_{1/2}$ value. In wheat, more than two such components are found. Proportion of each component is determined by using the formula $1 - \frac{C}{C_0}$, where C = concentration at $t_{1/2}$ for corresponding component.

Form this proportion, chemical complexity on the component can be determined.

B. Chemical Complexity : Chemical complexity is the amount of a DNA component measured by chemical assay.

For example, if the genome size is 12×10^8 bp and the component represents 25% of the genome then the chemical complixity of this component is 3×10^8 bp.

Chemical complixity = size of the genome × % of the component in this group.

If the kinetic complexity is known from earlier equation, then repetition frequency (f) of repetitive DNA component can be determined using the following formula——

$$f = \frac{\text{chemical complexity}}{\text{kinetie complexity}} = \frac{C_0 . t_{1/2} \text{ of non repetitive DNA}}{C_0 . t_{1/2} \text{ of repetitve DNA}}$$

Following table shows the reassociation of a eukaryotic genome starting at a C_0. t of 10^4 and terminating at C_0.t of 10^4. Reaction falls into three types of components and their results are as follows—

	Fast Component	Intermediate Component	Slow Component
% of genome	25	30	15
$C_0 \cdot t_{1/2}$	0.0013	1.9	630
ECinetie complexity (bp)	340	6×10^5	3×10^8
Repetition frequency	500000	350	1

There is a good relationship between the kinetic complexity and chemical complexity of eukaryotic genome.

Usually *E. coli* is used as a standard. Its components are taken to identical with the length of the genome. Thus, complexity of any DNA can be determined by compairing its $C_0 \cdot t_{1/2}$ with that of standard DNA of know DNA complexity.

$$\frac{C_0 \cdot t_{1/2} \text{ of (DNA of any genome)}}{C_0 \cdot t_{1/2} (E.coli \quad \text{DNA})} = \frac{\text{Complexity of nay genome}}{4.2 \times 10^6 \text{ bp}}$$

According to the table, the slow component represents 45% of the total DNA, so the concentration in the reassociation reaction is 0.45 of the measured concentration (total amount of DNA).

If DNA were isolated as a pure component, free of other fractions, it would renature with $C_0 \cdot t_{1/2}$ of $0.45 \times 630 = 283$.

Suppose that under these conditions *E. coli* DNA reassociates with a $C_0 \cdot t_{1/2}$ of 4.0, Comparing these two clues, we see :

$$\text{The kinetie complexity} = \frac{C_0 \cdot t_{1/2} \text{ of DNA of any genome} \times 4.2 \times 10^6}{4}$$

$$= \frac{0.45 \times 630 \times 4.2 \times 10^6}{4}$$

$$= 3 \times 10^8 \text{ bp}$$

$$\text{Then the whole genome is} = 3 \times 10^8 \times \frac{1}{0.45} = 6.6 \times 10^8 \text{ bp}$$

$$\text{First component complexity} = \frac{0.0013 \times 0.25 \times 4.2 \times 10^6}{4} = 340$$

$$\text{Second component complexity} = \frac{1.9 \times 0.30 \times 4.2 \times 10^6}{4} = 6 \times 10^5$$

Reversing the argument if we took three DNA preparations, each containing a unique sequence of the appropriate length 340 bp, 6×10^5 bp and 3×10^8 bp respectively and mix them in the proportions = 25 : 30 : 45, each would renature as though it was a single component, together the mixture would display the same kineties as those determined for the whole genome.

Non repetitive DNA complexity can estimate the genome size. The complexity of the slow components comprise sequences that are unique in the genome upon denaturation each single stranded sequence is able to renaturate only with the corresponding complementary sequences. It is usually the major component in eukaryotis. It is called non-repetitive DNA. According to the table the compexity of non-repetitive DNA is 3×10^8 bp. If this fraction is unique and represent 45% of the genome, then the whole genome would have a size of $3 \times 10^8 \div 0.45 = 6.6 \times 10^8$ bp. This provides an independent assessment of genome size. The value is approximately 7×10^8, obtained from the result of chemical complexity.

Eukaryotic genomes certainly contain repetitive sequences. Intermediate component occupies 30% of the genome. According to chemical complexity the total amount is $0.30 \times 7 \times 10^8 = 2.1 \times 10^8$ bp.

But kinetic complexity of this component is only 6×10^5 pp.

Thus repetition frequency = $= \frac{2.1 \times 10^8}{6 \times 10^5} = 350$

Thus intermediate components behaves as though consisting of a sequence of 6×10^5 bp that present in 350 copies in every genome. Repetition frequency (f) is the number of copies present per genome.

Highly repetitive DNA takes the name from the very large number of copies of the basic reassociating sequence present. The fast component consists 310 bp long in 500000 copies per genome. Because of the short length of the reassociating unit sometimes this is also referred to as simple sequence DNA.

$$\text{The repetition frequency (f)} = \frac{C_0 . t_{1/2} \text{ of non repetitive DNA}}{C_0 . t_{1/2} \text{ of repetitive DNA}}$$

$$= \frac{630}{0.0013} = 500000 \, (\text{Approx.})$$

Unit 11 Linkage Map, Cytogenetic Mapping

Structure

11.1 Physical and molecular maps

STS, i.e. Sequence Tagged Site is a short DNA sequence generally between 100-500 bp in length that is easily recognisable and occurs in the chromosome only once (i.e. unique).

STS mapping is a physical mapping procedure that locates the positions of sequence tagged sites (STSs) in a genome.

11.1.1 Qualities of STS

(1) Its sequence must be known, so that a PCR (polymerase chain reaction) assay can be set up to test the presence or absence of the STS on different DNA fragments.

(2) It must have a unique location in the chromosome being studied once in the genome. If the STS sequence occurs in more than one position, then the mapping data will be ambiguous. So STSs do not include sequences found in repetitive DNA.

11.1.2 Sources of STS

(1) **Expressed sequence tags (ESTs) :** These are short sequences obtained by the analysis of cDNA clones. cDNA is prepared by converting mRNA into dsDNA. Thus ESTs represent the genes that are expressed in the cell.

(2) **SS4Ps :** These are also used in genetic mapping as well as physical mapping of genes.

(3) **Random genomic sequences (RGS) :** These are obtained by sequencing random pieces of cloned genomic DNA.

11.1.3 Principles of STS mapping

(1) To map a set of STSs, a collection'of overlapping DNA fragments from single chromosome or from the entire genome is needed.

(2) The data from which the map will be derived are obtained by determining fragments which contain STSs. This can be done by hybridization analysis but PCR is generally used which is quieter and automated process.

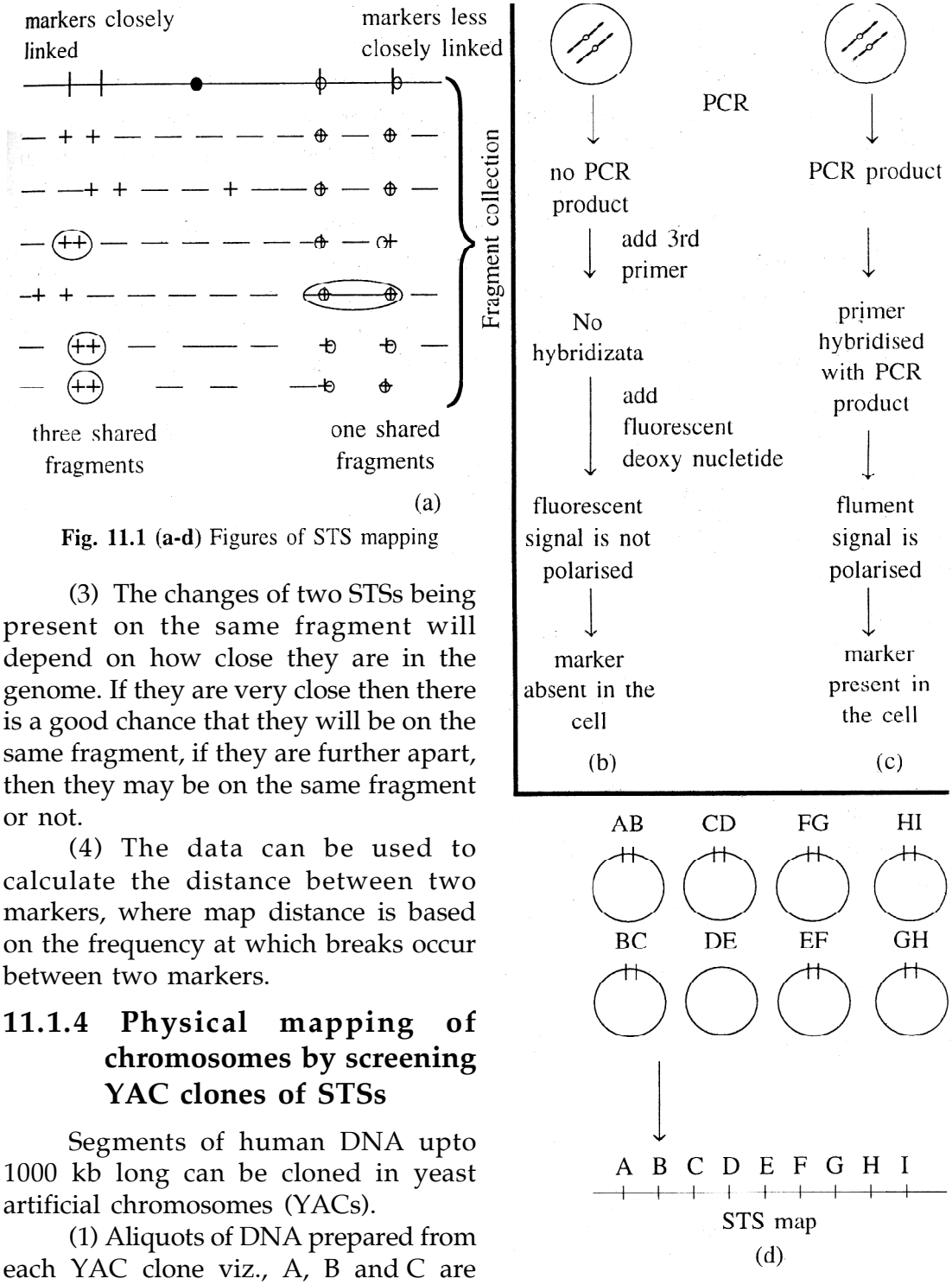


Fig. 11.1 (a-d) Figures of STS mapping

(3) The changes of two STSs being present on the same fragment will depend on how close they are in the genome. If they are very close then there is a good chance that they will be on the same fragment, if they are further apart, then they may be on the same fragment or not.

(4) The data can be used to calculate the distance between two markers, where map distance is based on the frequency at which breaks occur between two markers.

11.1.4 Physical mapping of chromosomes by screening YAC clones of STSs

Segments of human DNA upto 1000 kb long can be cloned in yeast artificial chromosomes (YACs).

(1) Aliquots of DNA prepared from each YAC clone viz., A, B and C are

subjected to PCR amplification using primer pairs (1-6) corresponding to the ends of various STSs. Only those clones containing STSs with ends complementary to particular primers will be amplified.

(2) Electrophoretic analysis then shows that YAC clones contain STSs.

(3) The illustration is very simple showing 6 primer pairs. Clone-A contains STSs no. 1,3 and 5; clone-B contains 2 and 1 and clone-C contains 3, 4, 5 and 6 primer pairs.

(4) The three YAC clones can be ordered showing their relative positions. In the mapping of human chromosome-21, about 1,20,000 clones from two separate YAC libraries were screened. In addition, 14,000 YACs isolated from a library prepared specifically from chromosome-21 were screened individually. By the use of 198 STSs, researchers identified 810 positive clones and ordered them into a contiguous map.

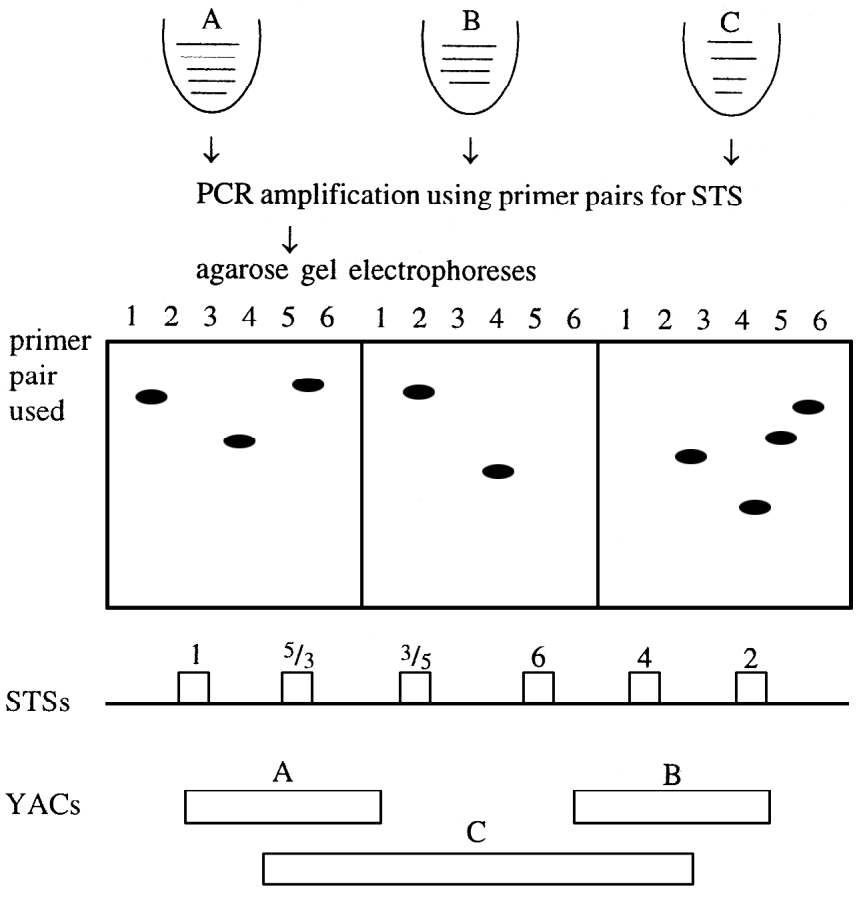


Fig. 11.2 Ordering of contiguous overlapping YAC

11.1.5 Fragments of DNA for STS mapping

At first, collection of DNA fragments sparing the chromosome or genome is required. The collection is called mapping reagent. There are two method as follows—

A. **Radiation hybrid method** :

(1) Human cells are exposed to X-ray doses of 3000-8000 rads, which causes the chromosomes to break randomly into fragments. Higher doses produce smaller fragments.

(2) These fragments can be propagated if the irradiated cell is fused with non-irradiated hamster (or other rodent) cells. Fusion is achieved chemically with polyethylene glycol or by exposure to Sendai virus.

(3) The hamster cell line that is unable to make either thymidine kinase (TK) or hypoxanthine phosphoribosyl transferase (HPRT) is used for the purpose. Those cells incapable of taking up human chromosome fragments are unable to survive in HAT (hypoxanthine, aminopterin and thymidine) medium. Those cells taking up chromosome fragments can synthesize TK and HPRT and are able to grow in HAT. Thus the hybrid cells are collected. These hybrid cells are then used as a mapping reagent in STS mapping.

(4) To map, as many markers as possible are us.ed. A pair of primers is designed for every DNA marker that was to be tested. Each primer pair is specific for one marker and will not give a PCR product with any other part of the genome.

(5) The success or failure of PCR is determined by agarose gel electrophoresis. The presence of a band of the expected size in the gel indicates that the PCR has worked.

(6) Another procedure has been designed where a third specific primer for each marker is added to the reaction mixture along with a fluorescently labelled dideoxy-nucleotide. If PCR has been successful, then the third primer anneals'to the product and is extended by the fluorescent dideoxy nucleotide and emits signals from this.

B. Use of clone library as the mapping reagent for STS analysis

A clone library can also be used as a mapping reagent in STS analysis. The clone library is prepared by using the genome or chromosome which is broken into fragments and are integrated into a high capacity vector.

The single specific chromosome can be isolated by flow cytometry technique and a clone library of a chromosome is made possible.

The data obtained from STS analysis is used for preparation of the physical map.

11.2 Restriction mapping of genes

Restriction mapping of genes is a physical mapping which locates the relative positions on a DNA molecule of the recognition sequences for restriction endonucleases.

Genetic mapping using RFLPs as DNA markers can locate the positions of polymorphic restriction sites within a genome, but very few of the restriction sites in a genome are polymorphic, so many sites are not mapped by this technique. Restriction mapping is very useful to solve a problem but the limitation of the technique is that it is applicable only to relatively small DNA molecules.

Two methods are employed for restriction mapping *viz.,* partial digestion method and double digestion method.

(1) In **partial** digestion method, the circular or linear DNA is treated with a particular restriction enzyme. But the enzyme is prevented from the complete digestion of DNA molecule. So the DNA is incubated for a short time or using a suboptimal incubation temperature. This leads to the partial digestion and will produce many cut products and uncut products. If the DNA molecule is linear, the terminal end of the DNA is labeled with P^{32} by the polynucleotide kinase reaction prior to cleavage and only radioactively labelled fragments are considered in agarose get electrophoresis, the other fragments are ignored/discarded of the molecular weight of each enhanced band is invariably the sum of the molecular weights of two fragments which arc considered to be adjacent. Thus the relative positions of the fragments can be ordered.

(2) Second method of the restriction mapping is the double digestion method, although partial digestion method is usually followed. In double digestion method, three samples of a particular DNA species are taken. Each two of these is treated with two separate restriction enzymes and the third sample is treated with both the restriction enzymes. Thus the three sets of fragments are compared following the agarose gel electrophoresis. The terminal end of the DNA molecule is also labeled with radioactive P^{32} prior to the restriction enzyme cleavage.

11.3 DNA foot printing

When transcription factor binds to a DNA sequence, it protects that sequence from digestion by nucleases. Researchers take advantage of this property by isolating chromatin from cells and treating it with DN-digesting enzymes, such as DNA ase-I, that destroy sections of the DNA that are not protected by bound transcription factors. Once the chromatin has been digested, the bound protein

is removed and the DNA sequences that had been protected are identified. This method is called DNA foot printing. This is also used to locate the binding sites of proteins on RNA.

Method

(1) A pure DNA fragment that is labeled at one end with ^{32}P is isolated.

(2) This molecule is then cleaved with a nuclease or a chemical that makes random single-stranded cuts in the DNA.

(3) The DNA molecule is then denatured to separate into two strands.

(4) The resultant fragments from the labeled strand are separated on a gel and detected by autoradiography.

The pattern of bands from DNA cut in the presence of a DNA-binding protein is then compared with that from DNA cut in its absence. The protein covers the nucleotides at its binding site and protects from DNA ase. The labeled fragments that shows no cleavage will show an area which is missing in the electrophoretic gel, is leaving a gap is called "foot print".

11.4 Gene mapping by human pedigree analysis (microsatellite mapping)

Recombination (CO) mapping is very difficult in human because

(1) It is impossible to preselect the genotypes of parents and set up crosses.

(2) The data for the calculation of recombination frequencies have to be obtained by examining the genotypes of the members of successive generations of existing families.

(3) The data obtained are very limited and their interpretation is often - difficult because in human test cross rarely occurs and the number of family members and offsprings are not many.

Therefore, gene mapping in human may be done by pedigree analysis.

Let us, suppose a family of two parents and six children were studied with a genetic disease. The diseased state is due to one allele and the healthy state is due to second allele. Diseased allele is dominant over healthy allele.

The pedigree showed that mother is affected because four of her children are affected by the disease (3♂ + 1♀).

The grand mother is affected. The grand father is dead but? We can assume that he was also affected. We can include them in the pedigree analysis.

The aim is to map the position of the gene for the genetic disease. For **that** purpose one is studying its linkage to a microsatellite marker M, four alleles of which *viz.*, M_1, M_2, M_3 and M_4 are present in the living family members. One has now to calculate the number of children who are recombinants.

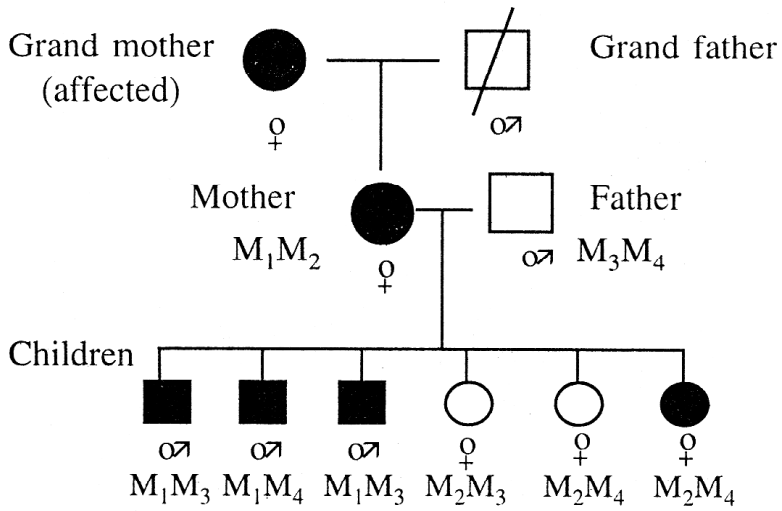


Fig. 11.3 Pedigree of a family with inheritance of genetie disease (solid means affected person)

The pedigree can be interpreted by two different hypotheses because child-1, 2 and 3 have the disease allele and microsatellite allele M], the child-4 and 5 have the healthy allele and microsatellite allele M_2.

In hypotheses-1, the mother would have the genotype = $\frac{\text{Disease allele} - M_1}{\text{Healthy allele - } M_2}$

The child no-1, 2, 3, 4 and 5 all would have parental genotype. Only the child no-6 would be a recombinant. The recombination percentage is 16.66 i.e., the disease gene and **the** microsatellite allele are relatively closely linked.

In hypothesis- II, the genotype of the mother would be = $\frac{\text{Healthy - } M_1}{\text{Disease - } M_2}$

Here the child **no-1,** 2, 3, 4, 5 are recombinants while child no-6 is with parental genotype. The recombination percentage is 88.33, which means that the disease gene and microsatellite gene are far apart on the chromosome.

		Possible genotypes of the mother	
		Hypothesis-I	Hypothesis-II
		Disease-M_1	**Healthy-M_1**
Child-1	**Disease-M_1**	**Healthy-M_2**	**Disease-M_2**
Child-2	**Disease-M_1**	Parental	Recombinant
Child-3	**Disease-M_1**	Parental	Recombinant
Child-4	**Healthy-M_2**	Parental	Recombinant
Child-5	**Healthy-M_2**	Parental	Recombinant
Child-6	**Disease-M_2**	Recombinant	Parental
	Recombination frequency	16.66%	83.33%

Fig. 11.4 Probable interpretation of the pedigree

Imperfect pedigrees are analysed statistically by using a measure called "lod score" (Morton, 1955). This stands for logarithms of the odds that genes are linked. This is used to determine whether the two markers lie on the same chromosome or not. If lod analysis establishes the linkage then the data will give confidence about their recombination frequencies.

If the number of the family members are larger the result would be more satisfactory. Atleast three generations are to be tested. Atleast four grand parents and atleast eight second generation children could be sampled.

Unit 12 Genetics of Cell Cycle

Structure

12.1 Genetic regulation of cell division in yeast and eukaryotes

12.1.1 Introduction

A cell reproduces by performing an orderly sequence of events in which it duplicates its contents and then divides into two. This cycle of duplication and division is known as the cell cycle. The basic organization of the cell cycle and its control system are essentially the same in all eukaryotic cells. Three **eukaryotic** systems in which cell-cycle is commonly studied are yeasts, frog embryo and cultured mammalian cells.

Yeasts are tiny, single-celled fungi. Two species are generally used in studies of cell cycle. The fission yeast viz. *Schizosaccharomyces Pombe,* is a rod shaped cell that grows by elongation at its ends. Division occurs by the formation of a septum or cell plate in the centre of the rod. It has a typical eukaryotic cell cycle with G,, S, G_2 and M phases. In contrast with that happening in higher eukaryotic cells, the nuclear envelope of the yeast cell does not break down during M-phase. The microtubules of the mitotic spindle are formed inside the nucleus and are attached to spindle pole bodies (SPB) at its periphery. The cell devides by forming a partition (cell plate) and splitting into two. The mitotic chromosomes are readily visible.

The budding yeast *Saccharomyces cerevisiae,* also called baker's yeast is a oval cell and divides by forming buds which first appears during G_1 and grows until it separates from the mother cell after mitosis. It has normal G_1 and S-phase but does not have a normal G_2 phase. The microtubule based spindle begins to form inside the nucleus early in the cycle during S-phase. Nuclear envelope remains intact during mitosis and the spindle forms within the nucleus.

12.1.2 Genetic regulation of cell cycle in *S.pombe* (Fig. 12.1)

(1) *ede 2* is identified as a crucial regulator by its involvement at both stages of cell cycle block, i.e. between G_2 and M-phase and in G_1 at start.

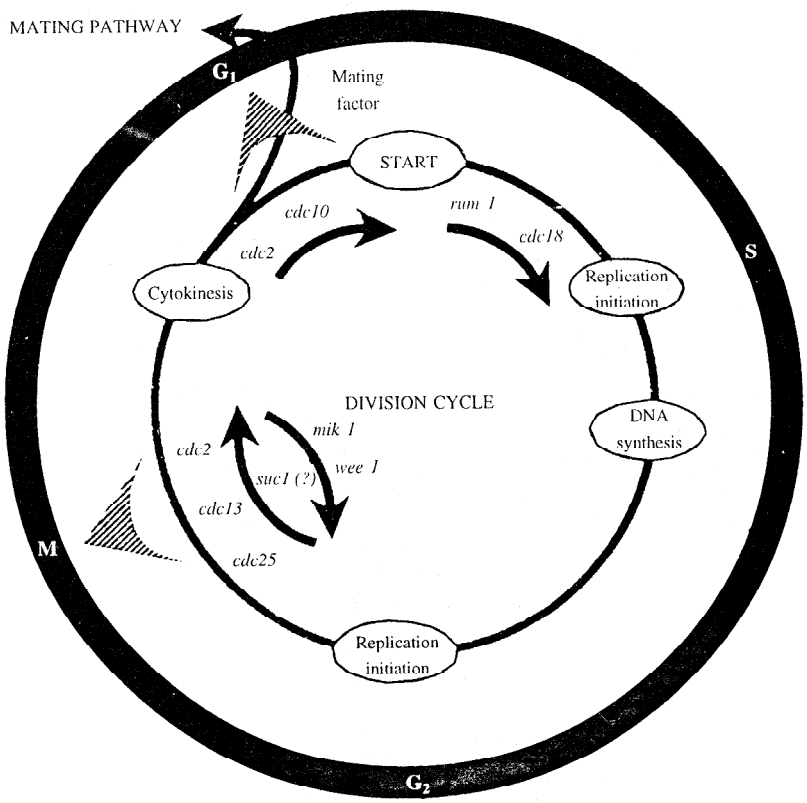


Fig. 12.1 The cell cycle in *S. pombe* requires *cdc* genes to pass specific stages, but may be retarded by genes that respond to cell size *(wee 1).* Cells may be diverted into the mating pathway early in G1

(2) During G_2 and M-phase cdc 2 has a partner which is the product of *Cdc 13* generating an M-phase kinase (resembling the p^{34}-B cyclin dimer of animal cells). The activity of the cdc 2 catalytic subunit in these dimers are controlled by phosphorylation (in the same way as p^{34} in animal cell).

But the difference that in yeast there is no Thr-14 so there are only two relevant sites Tyr-15 where phosphorylation is inhibitory and Thr-161 where phosphorylation is required.

(3) Under normal conditions, the cell division cycle is related to the size of the cell. In the poor growth condition the cells increase in size more slowly. The genes involved in control cell size are identified, *wee* 1 usually inhibits cells from initiating mitosis until their size is adequate. It has been suggested that *wee* 1 is part of a check point that prevents the activation of cdc 2 until an adequate mass has been attained. *wee* 1 codes for a kinase and can phosphorylate serine or

threonine and tyrosine. It inactivates cdc 2 by phyrophorylating Tyr-15. Another gene viz., *mik 1* has similar effects.

(4) The product of *cdc 25* is required for the dephosphorylation of cdc 2 in the cdc 2/cdc 13 dimer. It is probably responsible for the key dephosphorylating event in activating the M-phase kinase. The level of cdc 25 increases at mitosis and its accumulation over a threshold level could be important, cdc 25 executes the checkpoint that ensures s-phase to be completed before M-phase can be activated.

(5) The products of *wee 1* and *cdc 25* plays antagonistic roles. The kinase activity of *wee I* acts on Tyr-15 to inhibit cdc 2 function. The phosphatase activity of *cdc 25* acts on the same site to activate cdc 2.

(6) During mitosis the cdc 2 of cdc 2/cdc 13 dimer is in the active state that locks the phosphate at Tyr 15 and has the phosphate at Thr 161. At the end of mitosis kinase activity is lost and cdc 13 is degraded cdc 2 does not change at this point.

(7) During G, the active form of cdc 2 has a different partner, the B-like cyclin, cig 2 encoded by cig 2 gene. The dimer in converted from inactive state to the active state by dephosphorylation of Tyr-15 residue of cdc 2.

(8) Progression of Gj into *S* is controlled by activation of cdc 2-G| cyclin. In *S. pombe* it is cdc 2/ cig 2.

(9) Transcription of *cdc 18* is activated as a consequence of passing START and cdc 18 is required to enter S-phase. Over expression of cdc 18 allows multiple cycle of DNA replication without mitosis.

(10) For cdc 18 to be active, cdc 2/cdc 13 must be inactive. Again when M-phase kinase is active, it causes cdc 18 to be inactive possibly by phosphorylating it and prevents initiation of another S-phase.

(11) Activity of cdc 2/cdc 13 M phase kinase is influenced by a factor rurn-1, which controls entry into S-phase. When rum-1 is depleted, premature entrance into mitosis occurs and over expression causes cells to fail to enter mitosis. This suggests that M_{m-1} is an inhibitor of the M-phase kinase. It is expressed between G_1 and G_2 and keep M-phase kinase in an inactive state.

12.1.3 Genetic regulation of cell cycle in *S.cerevisiae* (Fig. 12.2)

Cell cycle in *S.cerevisiae* consists of three cycles that separate after START and join before cytokinesis. The cells may be diverted into the mating pathway early in G_1.

(1) Chromosome cycle: In this cycle, duplication and separation of chromosomes, completion of S-phase and nuclear division occurs. Mutation of *cdc 8* stops this cycle in S-phase. Mutation in the chromosome cycle do not stops the cytoplasmic cycle.

(2) **Cytoplasmic cycle** : It consists of bud emergence and nuclear migration into the buds. This cycle can be halted before bud emergence by cdc 24 mutation but the mutation does not prevent chromosome replication.

(3) **Centrosome cycle** : This cycle consists of duplication and separation of spindle polar body (SPB) and organizes microtubules to allow chromosome segregation within the nucleus. Blocking of the cycle by cdc 31 does not prevent S-phase or bud emergence.

Completion of entire cell cycle requires all three constituent cycles because nuclokinesis needs both chromosome and centrosome cycles but cytokinesis requires all the 3 cycles.

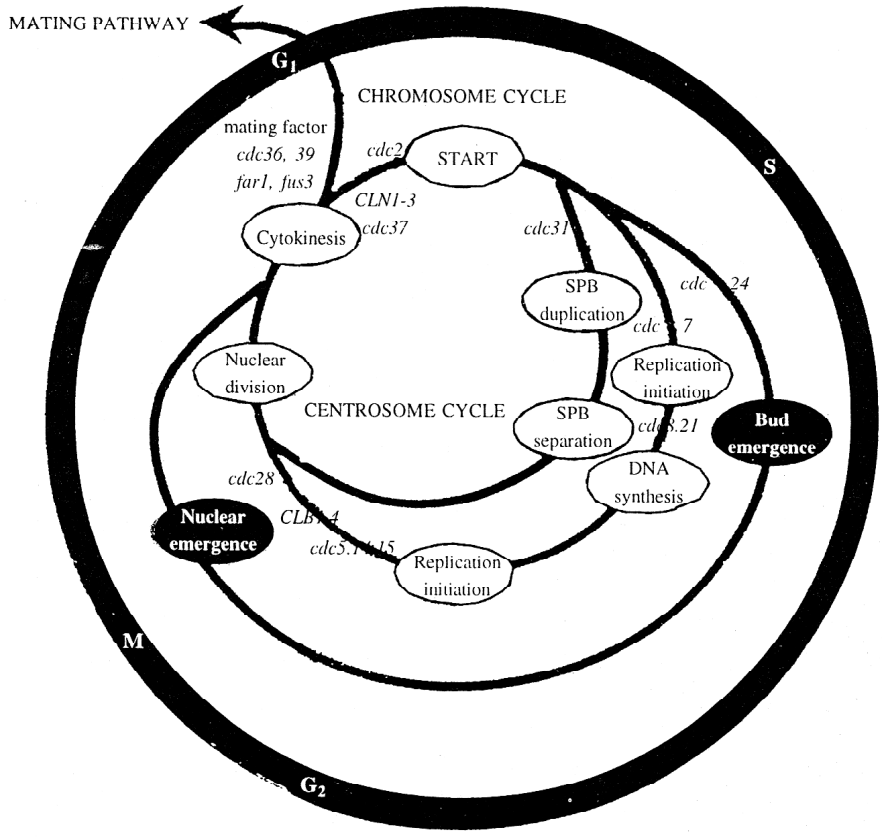
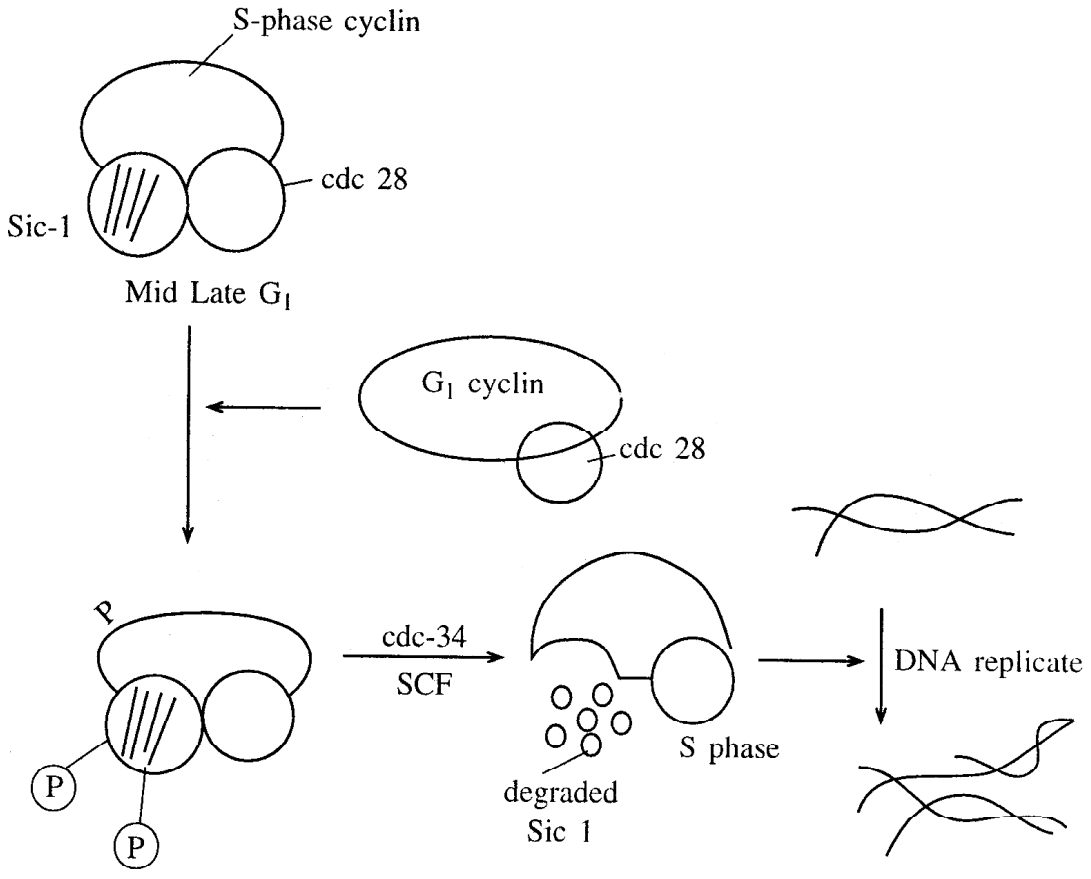


Fig. **12.2** The cell cycle in *S. cerevistae* consists of three cycles that separate after START and join before cytokinesis. Cells may be diverted into the mating pathway early in G1

(1) *S. cerevisiae* expresses a single cyclin dependent protein kinase (cdk) encoded by *cdc 28 gene* which interacts with several cyclins during different phases of the cell cycle.

(2) Just after cytokinesis, the decision on whether to initiate a division cycle is made before the START. The cells can diverted into mating type pathway by mating factors and *cdc 36* and *cdc 39* which appear to block the cell cycle before START and really function in the mating type pathway. Mutants block cell cycle by diverting cells into mating even in absence of the mating gene.

(3) After cytokinesis the mother ceil and the bud both the cell remain in the G_1 phase of the cell cycle. The ability to pass START is determined by environmental conditions. Prior to bud formation, spindle body duplication and DNA replication, the yeast cell exhausts its nutrients. When *S.cerevisiae* cells in G_1 have grown sufficiently in the growth medium, they begin a programme of



Fig, 12.3 Mechanism of Genetic regulation

gene expression that leads to entry into mitosis. Once G_1 cells reach the critical size, they become committed to completing cell cycle even if they are shifted to low nutrient medium.

(4) The crucial gene in pasing START is *cdc-28* which is homologns to *cdc-2* in *S.pombe.* Three cyclins are active in G,. CIn 1, Cln 2 and Cln 3 are encoded by *CLN-},* CLN-2, *CLN-3* respectively. Mutations in any one or two of these genes fail to block the cell cycle, thus the *CLN* genes are functionally redundant.

(5) The complexes formed between cdc 28 and the three G_1cyclins (Cln 1, Cln 2 and Cln 3) have protein kinase activity and constitute the hypothesized S phase promoting factor (SPF_S). In wild type yeast cells, Cln 3 is expressed at a nearly constant level throughout the cell cycle. Cln 1 and Cln 2 are expressed during the second half of G_1 and they increase rapidly and when their accumulation exceeded a critical threshold level, triggers the passage of *start* (START) into S-phase. After that its concentration declines gradually and are eliminated by the time of mitosis.

(6) cdc 28-Cln 3 phosphorylates and activates SBF and MBF. These induce transcription of *CLN 1* and *CLN 2* genes as well as several other genes required for DNA replication, including genes encoding DNA polymerase, RPA (ssDNA binding proteins), DNA ligase and certain enzymes acquired for deaxyribonucleoside triphosphate synthesis.

(7) cdc 28-Cln 1 and ede 28-Cln 2 phosphorylate APC in late G_1 and inactivate it.

(8) Two B-type cyclin genes *CLB 5* and *CLB 6* are also regulated by MBF and transcribed beginning in late Gj. The corresponding proteins Clb 5 and Clb 6 accumulate because of the inactivation of APC.

(9) At late G_1 ede 28-Clb 5 and ede 28-Clb 6 heterodimers accumulate and are inactivated by Sic-1 (an S-phase inhibitor), but it has no effect on ede 28-Cln complexes. Sic 1 is degraded falling polyubiquitination by E_2 associated with E_3. Once Sic is degraded, ede 28-Clb 5 and ede 28-Clb 6 kinases induce DNA replication.

(10) Initiation of DNA replication needs both assembly of pre replication complex and an active cdc 28-Cln complex. A second heterodimeric protein kinase ede 7-Dbf 4, which is expressed in G_1 is also required to trigger initiation. Once replication has initiated, Mem proteins and ede 45 move away from the origin along with DNA polymerases. Mem proteins are homologous to helicase, associated with replication fork movements.

(11) Later in S-phase, transcription of the genes *CLB 3* and *CLB 4* begins, encoding two additional B-type cyclins, Clb 3 and Clb 4, which also form

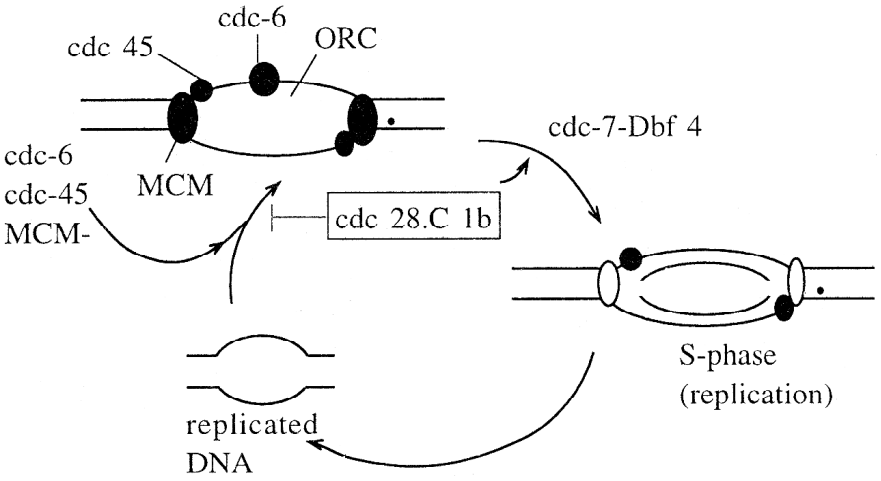


Fig. 12.4 Role of cdc in replication

heterodimeric protein kinases with cdc 28. These two cdc 28-Clb 3 and cdc 28-Clb 4 complexes also initiate the formation of mitotic spindle at the beginning of mitosis.

(12) As cells complete chromosome replication and enter G_2, two more B-cyclins are expressed. These are Clb 1 and Clb 2 encoded by *CLB 1* and *CLB 2* genes. These function as mitotic cyclins, associating with cdc 28 to form complexes that arc required for chromosome segregation and nuclear division.

12.2 Molecular basis of cellular checkpoints

12.2.1 Introduction

In most cells there are several points in the cell cycle, called *checkpoints* in which the cycle can be arrested if previous events have not been completed. Four checkpoint control can arrest the passage through cell cycle. These are (a) Gj-arrest due to DNA damage, (b) S-arrest due to unreplicated DNA, (c) C_2-arrest due to DNA damage, (d) M-arrest due to improper spindle formation.

In the checkpoints, the control system can be regulated by extra cellular signals from other cells. These signals either promote or inhibit cell proliferation. Checkpoints generally operate through negative intracellular signals.

12.2.2 The DNA replication checkpoint

Most of the cells by DNA replication checkpoint mechanism, avoided entry into cell division until the last nucleotide in the genome has been copied. The

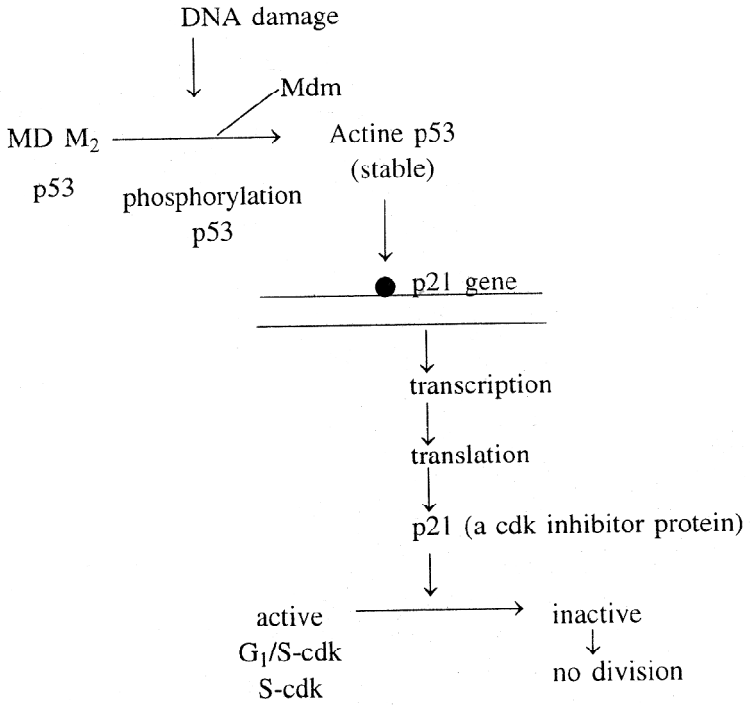


Fig. 12.5(a) Mechanism of G_1 arrest checkpoint

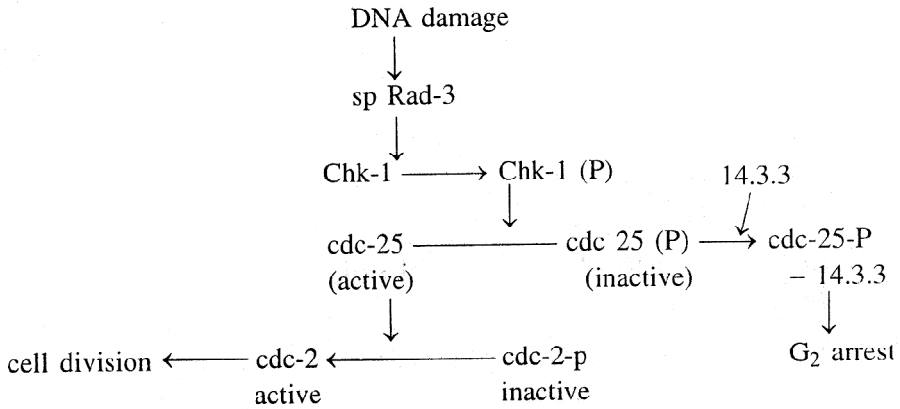


Fig. 12.S(b) Mechanism of G_2 arrest checkpoint

molecular mechanism has not been discovered but any signal from unreplicated DNA or unfinished replication forks send a negative signal to the cell cycle control system that blocks the activation of M-Cdk. Thus normal cells treated with chemical inhibitors of DNA synthesis viz., hydroxyurea, do not progress

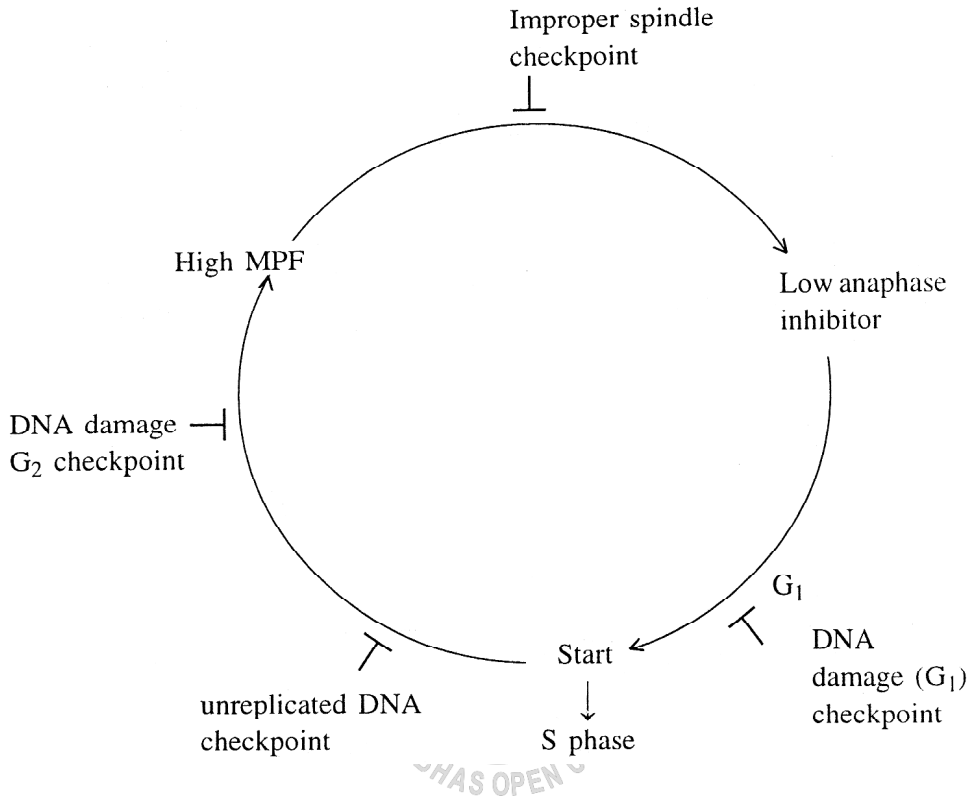


Fig. 12.5(c) DNA replication checkpoint

into mitosis. The block activates a checkpoint mechanism that arrests the cells in S-phase, delaying mitosis. But if caffeine is added along with hydroxyurea checkpoint mechanism fails and the cells proceed into mitosis according to their normal schedule with incompletely replicated DNA. As a result, the cells die.

12.2.3 The spindle attachment checkpoint

The effect of colchicin which inhibits spindle assembly, shows the presence of this checkpoint. In most cell types, a spindle attachment checkpoint mechanism operates to ensure that all chromosomes are properly attached to the spindle before sister chromatid separation occurs. During metaphase kinetochore regions of the chromosomes are attached to microtubules and any kinetochore that is not properly attached to the spindle, sends out a negative wait signal to the cell cycle control system that blocks ede 20-APC activation which is needed for sister chromatid separation. The nature of signal is not clear. But it has been seen that several proteins, including Mad 2 are recruited to unattached kinetochores which

are required for spindle attachment checkpoint to function. Even a single unattached kinetochore results inhibition of cdc 20-APC activation by binding with Mad 2.

In mice **MAD** 2 and BUB 1 and in humans **MAD** 2 genes have been recently identified. MAD 2 protein remains concentrated at the kinetochore until completion of microtubules attachment. This protein is continuously migrating into the cytoplasm and broadcasting signal throughout the cytoplasm. In mammals MAD 2 is associated with p^{55} CDC but in budding yeast MAD 2 is with cdc 20 and in fission yeast Mad 2-slp 1 (sip = sleepy) form a large complex with APC. Hct 1, another protein is associated with APC. APC is kept in check by MAD 2 and when APC is active the cell initiate anaphase by catalizing degradation of the two proteins with the help of Hct and cdc 20.

cdc 20/slp 1 promotes degradation of CUT2 but the activity of cdc 20/slp 1 is inhibited by MAD 2. Hct 1 promotes degradation of CLB-1, APC catalyzes this degradation by ubiquitination. This ubiquitination is inhibited by spindle attachment checkpoints.

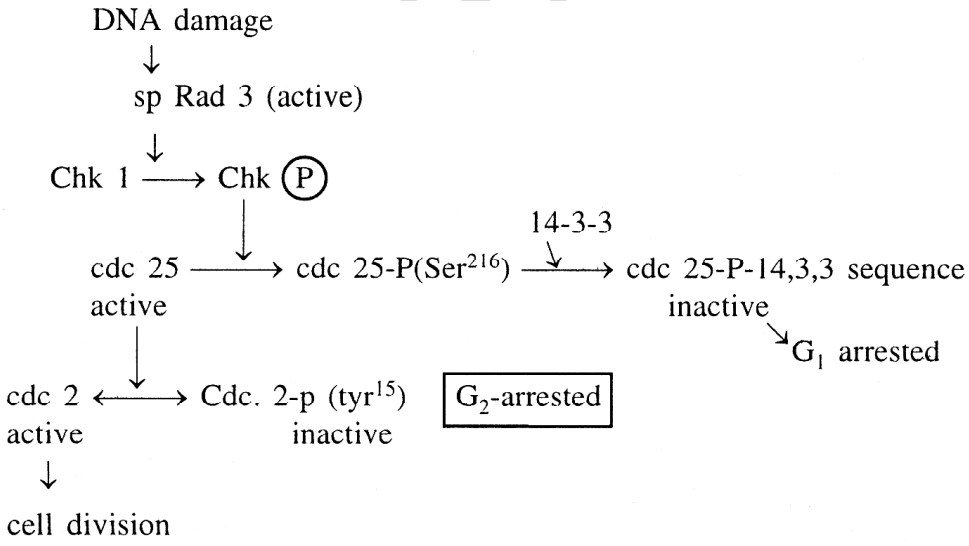


Fig. 12.6 DNA damage checkpoint

In mammals BUB 1 and BUB 3 proteins along with CNEP-E protein induce changes that lastly shut down the transmission of wait signal. MAD 2 dissociates from p^{55} cdc and loses its control on APC. APC is activated and catalyzes breakdown of cyclins and facilitates anaphase separation of chromatid.

12.2.4 DNA damage check points

The cell cycle control system can readily detect DNA *damage* and *arrest* the cell cycle at DNA damage checkpoints. These two checkpoints are one is in late G\ which prevents entry into S-phase and the other is in late G_2 which prevents entry into, mitosis.

G_2 checkpoint depends on a (*similar*) mechanism that delays entry into mitosis in response to incomplete DNA replication. The damaged DNA sends a signal to a series of protein kinases that phosphorylate and inactivate the phosphatase-cdc 25. This blocks the dephosphorylation and activation of M-cdk, thereby blocking entry into mitosis when the damaged DNA is repaired, the inhibitory signal is turned off and the cell division continues.

In the yeast *S.pombe,* the checkpoints sense the DNA damage and transduce inhibitory signal. Four genes including *RAD 9,* RAD 17, *RAD 24* and MEC 3 will sense this damage. The model proposed that DNA damage activates a protein sp Rad 3 (sp prefix means *S. pombe).* It brings about phosphorylation of Chk 1 and Chk 1-Ⓟ functions as a kinase, which brings about phosphorylation at Ser^{216} of Cdc 25. It then promotes binding of cdc 25 to a protein 14-3-3, coded by *rad 24* and *rad 25,* leading to sequestration of cdc 25. *Then cdc 25 is not available for activation of cdc 2 P (tyr^{15}) and the cell division is arrested at G_2-M transition.*

G_1 checkpoint blocks progression into S-phase by inhibition of G_1S-Cdk and S-Cdk complex.

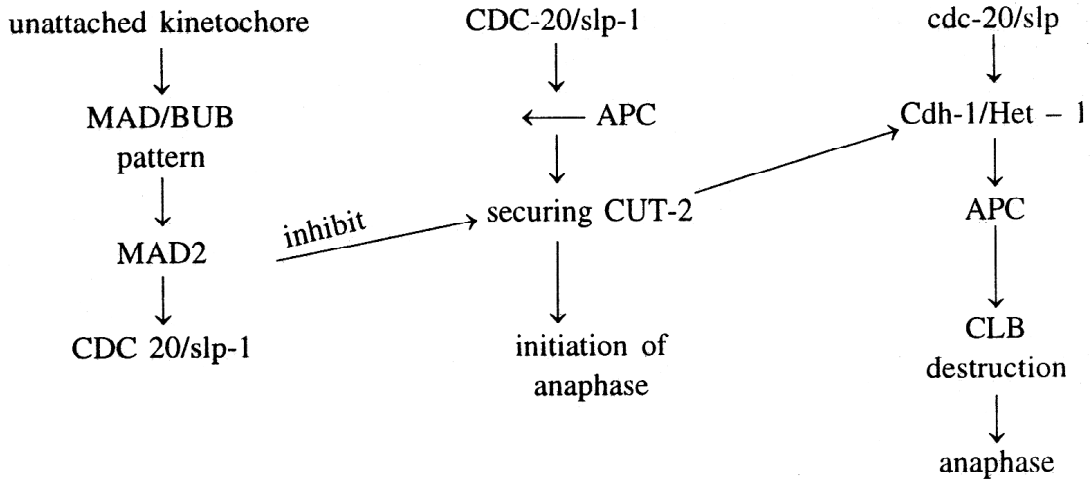


Fig. 12.7 Spindle attachment checkpoint

(1) In mammalian cells a gene regulatory protein p^{53} is being activated by DNA damage and it stimulates transcription of many genes.

(2) A CKI protein, called p^{21} is encoded by such activated genes which binds to G_1/S-Cdk and S-Cdk and inhibits their activity and thus blocks entry into mitosis.

(3) Actually in an undamaged cell, p^{53} is very unstable and is present at a low concentration. It interacts with another protein Mdm2, that causes destruction of p^{53} by ubiquitination mechanism. Damaged DNA activates protein kinase, phosphorylates p^{53} and reduce the binding of Mdm2. As a result p^{53} concentration rises and stimulates gene transcription of p^{21} which inactivates G_1/S-Cdk and S-Cdk activity.

(4) A rare genetic disease, Ataxia telangiectasia is caused by a defect in one of the protein kinases that phosphorylates and activates p^{53} in response to radiation and due to the loss of the DNA damage checkpoints, they suffer from increased rate of cancer.

GROUP
B (II)
Molecular Biology

Unit 1 History and Scope of Molecular Biology

Molecular Biology is one of the most rapidly growing domains of life sciences in the 21st century. Though the subject is relatively new in comparison to the specialized areas like—cytology, bacteriology, morphology and embryology; it has a history which is interesting and worth mentioning. Antony van Leeuwenhoek's invention of the microscope in around 1650 opened up the micro-world of biology. Advancement of knowledge in the fields of biochemistry, microbiology, virology and genetics in the early part of the 20th century gave rise to a new domain of activity which also attracted the attention of chemists and physicists. The remarkable development in physics and technological advancement opened up new frontiers in the biological arena one of which came to be known as **Molecular Biology** - the term was coined by **Warren Weaver** of the Rockefeller Foundation. Molecular biology attempts to explain the phenomena of life and its evolution, starting from the macromolecular properties that generate them. The two main macromolecules that contribute in the life process and remains the focus of the molecular biologist are the nucleic acids, (DNA & RNA)- the constituent of genes, and proteins, which are the active ingredients of life processes.

Charles Darwin's publication of **On *the Origin of Species*** (1859) had a profound impact on biological thinking. The rediscovery of the work of the Austrian monk Gregor Mendel probably might have been the first impetus to the development of molecular biology. August Weismann's description of reduction division in 1'887 followed by the description of meiosis, spermatogenesis by Walter Sutton (1903) lead to the proposal 'Chromosomal Theory of Inheritance". Acceptance of chromosomal theory of inheritance by 1935 became the stepping stone toward the journey for the search of chemical and physical nature of hereditary factors.

Several important developments took place during the early part of the 20th century that suggested the existence of genes within the chromosomes. Thomas Hunt Morgan (1909) showed that phenotypic change in *Drosophila* is linked to the events of crossing over in the chromosomes while Alfred Sturtevant (1910) moved a step ahead and mapped gene on chromosomes. Although existence of a new substance called nuclein was identified in the sperm cells by Friedrich Meischer in 1869, which later came to be known as deoxyribonucleic acid (DNA), it was Robert Fuelgen (1924) who first showed the coexistence of DNA along with proteins in

chromosomes by cytochemical staining. Appreciation of the nucleic acids quickly led to findings that there are two types of nucleic acids- RNA and DNA that differed in their sugar moieties. In 1929 Phoebus Levene showed that there are four types of DNA molecules - each of which he referred to as nucleotide. Each nucleotide had a deoxy-ribose sugar unit, a phosphate group and a nitrogenous base. The four nitrogenous bases were identified as Adenine, Guanine, Thiamine and Cytosine. He also suggested that the nucleotides are linked together through their phosphate sugar back bone but he made a mistake by considering the nucleotides to be present in short sequences and that the bases repeated in the same fixed order. However, the DNA molecule exists as a polymer was confirmed by Torbjorn Caspersson and Einar Hammersten (1934).

During this time Fred Griffith (1928) used pneumoccocus to describe gene transformation and George Beadle & Edward Tatum came up with "one gene - one enzyme" theory to demonstrated the existence of a precise relationship between genes and proteins. Following these discoveries, numerous research groups confirmed the importance of the gene in the life and development of organisms. It became apparently clear that genes are present in chromosomes and chromosomes are made of DNA and proteins. Initially the scientific community could not consider DNA as a hereditary material because of its utter simplicity. Rather, complex nature of the proteins was a preferred candidate to store hereditary information. Nevertheless, the chemical nature of genes and their mechanisms of action remained a mystery. Molecular biologists committed themselves to the determination of the structure of gene and the description of the complex relations between, genes and proteins. In the 50's, two important events took place almost simultaneously. Oswald Avery, Maclyn MacLeod and Colin McCarty in 1944 for the first time came out with the suggestion that genes are made up of DNA. Although their work was severely criticized by the scientific community then, Alfred Hershey and Martha Chase (1952) confirmed through their ingenious experiment that the genetic material of the bacteriophage is DNA.

Then came the discovery that revolutionized the world of science. Watson and Crick (1953) published their paper entitled "Molecular structure of Nucleic Acids" *Nature* 171, 737-738 (1953) where they gave a detail account of the double helix structure of the DNA molecule. The discovery of the double helical structure is history by itself. In brief there were three groups working to elucidate the structure of the DNA molecule. The first group worked at King's College, London and was led by Maurice Wilkins and was later joined by Rosalind Franklin. The second group working on DNA was Francis Crick and James D. Watson was at Cambridge

and the third group was at Caltech where the noble laureate Linus Pauling was leading the show. Inspiration and data from the works of Erwin ChargafPs work, published in 1947 and the X-ray diffraction patterns of DNA fibers produced by Maurice Wilkins and Rosalind Franklin at King's College facilitated Watson and Crick to design the double helix model using metal rods and balls in which they incorporated the known chemical structures of the nucleotides, as well as the known position of the linkages joining one nucleotide to the next along the polymer. For their pioneering work, Watson, Crick and Wilkins were awarded Nobel Prize in physiology in the year 1962. In their paper on the structure of DNA double helix, Watson and Crick wrote at the end of the paper, 'It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material'—a path breaking thought that allows the preservation of hereditary information from generation to generation.

New thought processes began to evolve centering round the DNA molecule. Subsequently, Crick in 1957 proposed the 'Central Dogma' that explains the flow information of genes to proteins and the relationship between DNA, RNA and proteins. Controversies arose regarding the replication mechanism. However, ingenious experiment by Meselson-Stahl put an end to all controversies by showing the DNA replication is semi conservative in nature. Crick and coworkers showed that the genetic code was based on non-overlapping triplets of bases, called codons, and Har Gobind Khorana (1961) and others deciphered the genetic code not long afterward. These findings represent the birth of molecular biology which is still advancing at rapid pace everyday adding new information and creating history.

Following the deciphering of the genetic code, Arthur Kornberg described the action of DNA polymerase; Stanley Cohen (1968) discovered plasmids and antibiotic resistance gene. These findings gave birth to recombinant DNA technology which was immediately followed by the development of DNA sequencing technique by Walter Gilbert, allan Maxam, Fred Sanger in the year 1975. Simultaneously, Ceasar Milstein, Geo Kholer and Niles Jeme developed the art of monoclonal antibody production. Split gene concept was put forward by Richard Roberts and Philip Sharp in 1977 that marked the difference between the eukaryotic and prokaryotic genome. Another giant leap was made in the field of molecular biology when Kary Mullis (1985) successfully synthesized DNA polymers *in vitro* - a technique better known as Polymerase Chain reaction (PCR). Scientist then dared to sequence the 3 billion nucleotides present in human and launched the flamboyant project "Human Genome Project" in the year 1989 with a

target to complete the entire human sequence by 2010. Thanks to the technological development and discovery of efficient DNA polymerases that enabled to complete the project in the year 2003. The history does not end here. New branches like proteomics and genomics have come up to understand the functioning of cellular genes and to utilize the information for betterment of the man kind.

Unit 2 DNA Replication

Structure

2.1 Introduction

The fundamental biological process of reproduction requires the faithful transmission of genetic information from parent to offspring. Genetic information is stored in the form of an array of nucleotide sequences. The life process has evolved mechanisms to replicate the array of nucleotide sequences with great accuracy, that too at an astounding speed. The single, circular chromosome of *E. coli* contains about 4.7 million base pairs. Duplicating at a rate of more than 1000 nucleotides per minute, replication of the entire chromosome would require almost 3 days. Yet, these bacteria are capable of dividing every 20 minutes making minimum errors during the replication process. A huge amount of genetic information and an enormous number of cell divisions are required to produce a multicellular adult organism; even a low rate of error during copying would be catastrophic. Interestingly, mechanism to correct base pairing mistakes that occur during the process of replication has evolved in both prokaryotes and eukaryotes.

In 1953, Watson and Crick wrote at the end of their paper on the structure of DNA double helix, 'It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material'. They recognized and explained that an inherent copying mechanism exists in the double helix DNA molecule. During replication, the two strands of the DNA helix unwind and unzip; each strand serves as a template for the new DNA molecule to be synthesized on it - a process termed as **semi conservative** replication.

2.2 Semiconservative replication

Semi conservative replication hypothesize that during replication of DNA double helix the strands unwinds and each strand serves as a template on which

new daughter strands are synthesized following the base paring rule. That is, an **'A'** at one position on the mother strand signals the addition of a **'T'** on the corresponding position on the newly forming strand. Similarly the presence of **'G'** on the mother strand will signal the addition of **'C'** on the daughter strand. This type of base pairing, which determine the nucieotide sequence of the new strand is known as **complementary base pairing.** Once the bases are aligned, DNA polymerase enzyme link the new incoming nucieotide with the previous aligned nucieotide of the daughter strand and eventually hydrogen bond is established between the base pairs. The process continues until the entire length of the mother DNA strand is copied. The newly formed DNA helix comprises of a new strand and an old original strand (conserved strand). Such pattern of DNA double helix duplication is called **semiconservative** replication. However, alternative mechanisms of replication were proposed which are known as **conservative replication** and **dispersive replication** (Fig. 2.1).

According to conservative replication, the original two strands serve as templates for the formation of new DNA strands. But, one of the double helix would consist entirely of original DNA strands, while the other helix would consists of two newly synthesized strands.

Dispersive replication suggests that the two DNA helix formed after replication comprises of interspersed blocks of new and old strands. However, evidences of such type of replication are still lacking.

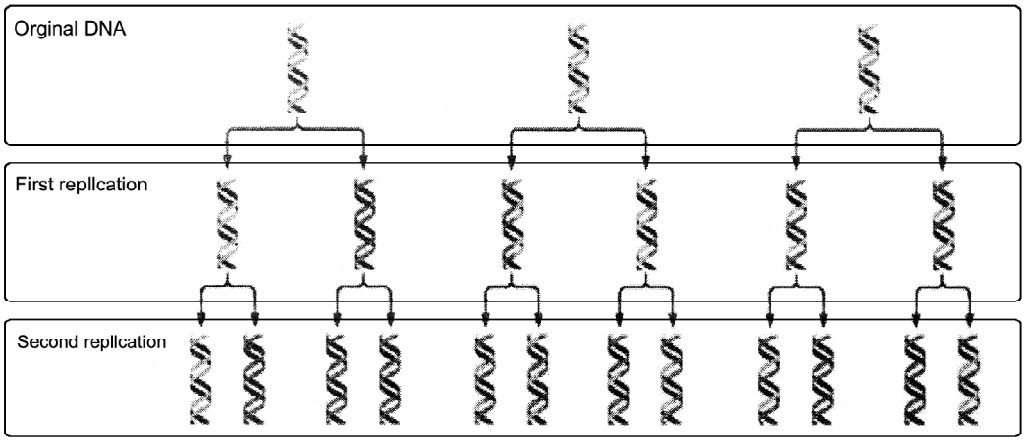


Fig. 2.1 : Proposed mechanisms of DNA replication: Semiconservative, Conservative and Dispersive

2.2.1 Experimental evidence of semiconservative replication

In 1958, M. Meselson and F. Stahl performed an imaginative experiment that confirmed the semiconservative nature of DNA replication. They grew *E. coti* in a medium containing a heavy isotope of nitrogen, ^{I5}N for several generations. After

growing for several generations in ^{15}N medium, practically all nitrogen atoms in the DNA of the bacterial cells were labeled with ^{15}N. They then transferred some of the cells from ^{15}N medium to new medium in which the nitrogen was ^{14}N. The bacteria were allowed to divide for one cycle only. Similarly, in another tube, the cells transferred to ^{14}N medium and were allowed to divide for two generation only. Any DNA synthesized after the transfer would contain a mixture of light & heavy isotopes. The newly cultured cells were then isolated, their DNA extracted and subjected to **equilibrium density gradient centrifugation** in Cesium chloride gradient to determine the density of the respective DNAs. (Cesium chloride [CsCl] centri-fuged at 50000 rpm, 250,000g for 2 days that produce the gradient).

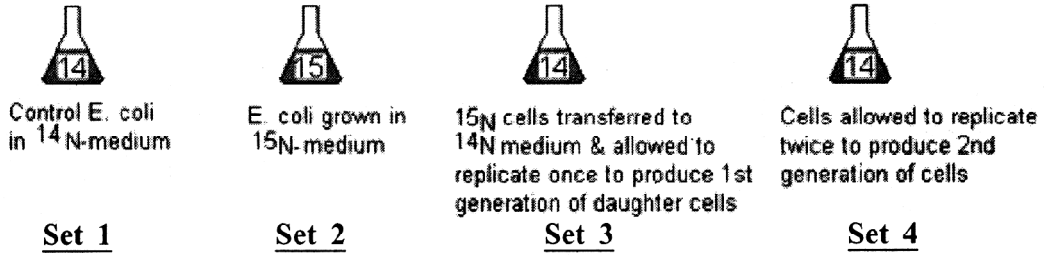


Fig. 2.2: *E. coli* grown in medium having either ^{15}N or ^{14}N as the only source for nitrogen for DNA replication

When DNA from each of these sets of cells was prepared and centrifuged in a cesium chloride gradient, the DNA formed bands depending on their respective density (**Fig. 2.3).**

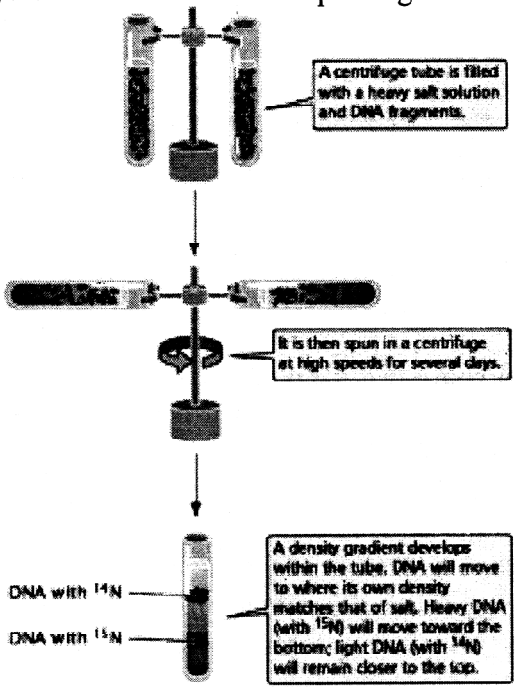


Fig. 2.3 Meselson and Stahl used equilibrium density gradient centrifugation

$^{I5}N^{I5}N$ -DNA being heavier, they formed band at bottom end of the tube **(Set 2).** $^{15}N^{14}N$-DNA produced after one round of cell division in ^{I4}N containing medium was lighter than the ^{I5}N type DNA and formed a band above the $^{15}N^{I5}N$ bands **(Set 3)**. In accordance with the semi conservative replication, the cells that were allowed to divide twice in ^{I4}N- medium should have a mixture of $^{I5}N^{14}N$ & $^{I4}N^{14}N$ DNA and should form two bands in CsCl gradient. As per prediction, the cells from the **Set 4** produced two bands. The lower band corresponded with the $^{15}N^{I4}N$ band observed in **Set 3** and the upper band corresponded to the $^{I4}N^{14}N$ band of Set 1 **(Fig. 2.4).**

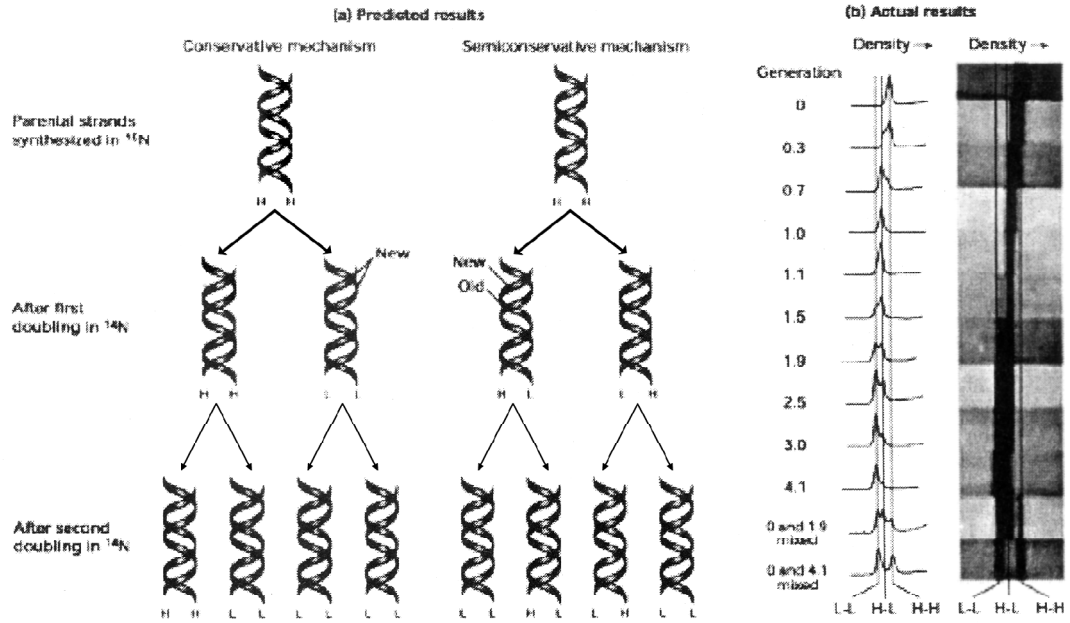


Fig. 2.4: Meselson and Stahl's experiment showing that DNA replication is semiconservati ve.

This classic experiment confirmed the prediction of the semi conservative mode of replication envisaged by Watson and Crick and disapproved all notions of conservative and dispersive models for DNA replication. Later, autoradiographic study by J. Cairns (1963) on bacterial DNA replication confirmed the observation of Meselson and Stahl (1958). The study also elucidated the circular nature of bacterial DNA and showed that DNA replication occurs simultaneously on both the strands at one or two moving 'Y' shaped forked junctions in the circular DNA. There are, however, several different ways that semiconservati ve replication can take place, differing principally in the nature of the template DNA—whether it is linear or circular— and in the number of replication forks. Individual units of

replication are called **replicons,** each of which contains a **replication origin.** Replication starts at the origin and continues until the entire replicon has been replicated. Bacterial chromosomes have a single replication origin, whereas eukaryotic chromosomes contain many.

2.2.2 Theta replication:

A common type of replication that takes place in circular DNA, such as that found in *E. coli* and other bacteria, is called **theta replication (Fig. 2.5),** because it generates a structure that resembles the Greek letter theta (e). In theta replication, double-stranded DNA begins to unwind at the replication origin, producing single-stranded nucleotide strands that then serve as templates on which new DNA can be synthesized. The unwinding of the double helix generates a loop, termed a **replication bubble.** Unwinding may be at one or both ends of the bubble, making it progressively larger. DNA replication on both of the template strands is simultaneous with unwinding. The point of unwinding, where the two single nucleotide strands separate from the double-stranded DNA helix, is called a **replication fork.** If there are two replication forks, one at each end of the replication bubble, the forks proceed outward in both directions in a process called **bidirectional replication,** simultaneously unwinding and replicating the DNA until they eventually meet. If a single replication fork is present, it proceeds around the entire circle to produce two complete circular DNA molecules, each consisting of one old and one new nucleotide strand.

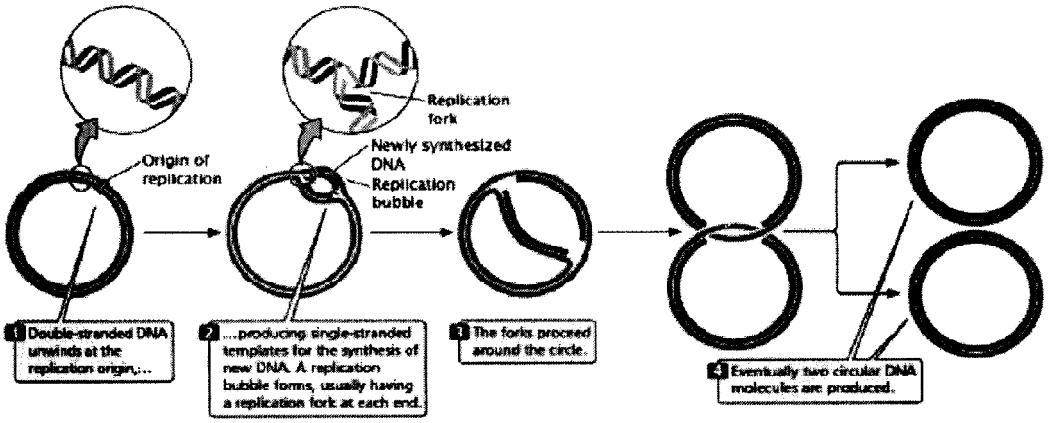


Fig. 2.5a. Theta replication in *E.coli* and other organisms possessing circular DNA

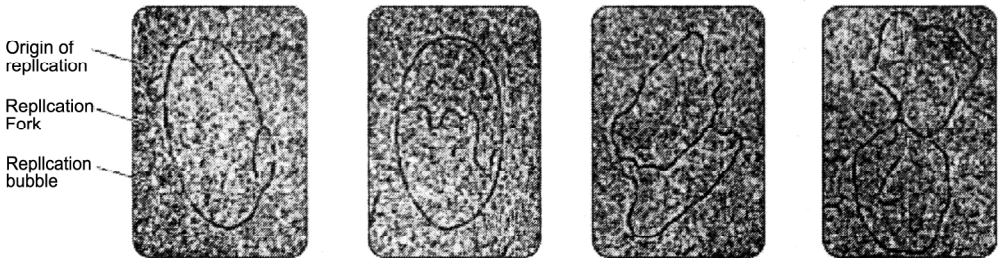


Fig. 2.5b: Experimental evidence produced by J. Cairns (1963) to show the Theta mode of replication in *E.coli*

Rolling-circle replication takes place in some viruses and in the F factors of *E. coli.* This form of replication is initiated by a break in one of the nucleotide strands that creates a 3'-OH group and a 5'-phosphate group. New nucleotides are added to the 3' end of the broken strand, with the inner (unbroken) strand used as a template. As new nucleotides are added to the 3' end, the 5' end of the broken strand is displaced from the template, rolling out like thread being pulled off a spool. The 3' end grows around the circle, giving rise to the name rolling-circle model. The replication fork may continue around the circle a number of times, producing several linked copies of the same sequence. With each revolution around the circle, the growing 3' end displaces the nucleotide strand synthesized in the preceding revolution. Eventually, the linear DNA molecule is cleaved from the circle, resulting in a doublestranded circular DNA molecule and a single-stranded linear DNA molecule. The linear molecule circularizes either before or after serving as a template for the synthesis of a complementary strand (Fig. 2.6).

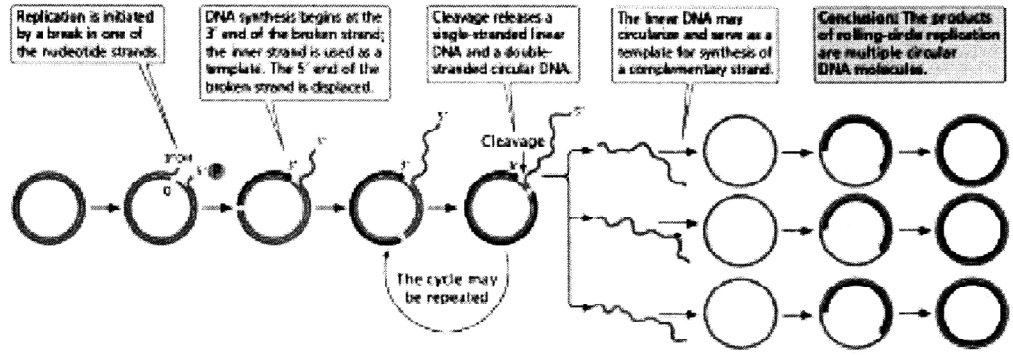
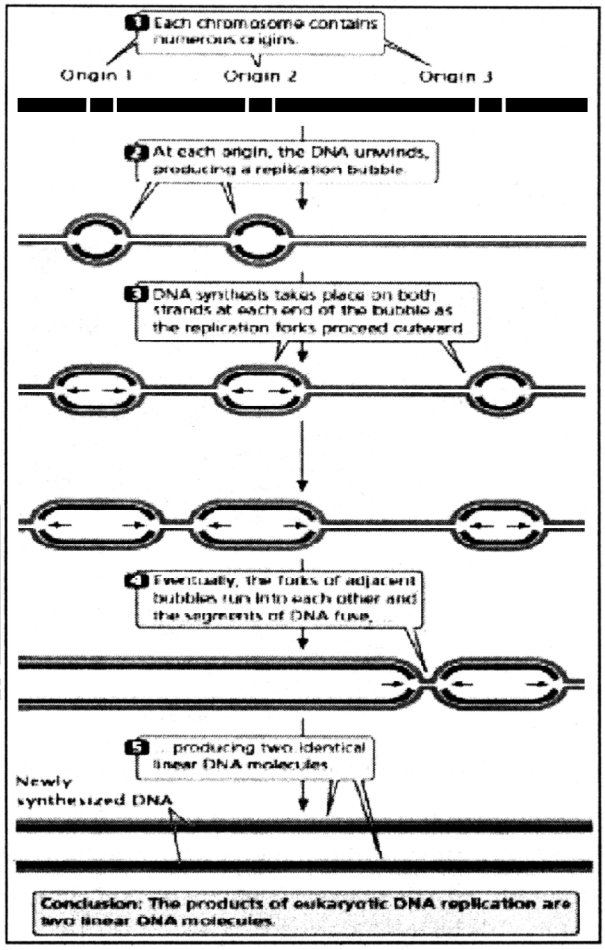


Fig. 2.6: The linear molecule serving as a template for the synthesis of a complementary strand.

2.2.3 Linear eukaryotic replication :

The large linear chromosomes in eukaryotic cells contain too much DNA and need to replicate speedily within a reasonable time frame and therefore cannot afford to have a single origin of replication fork as found in bacteria. Multiple

replication origin point exists in eukaryotic that proceed at a rate ranging from 500 to 5000 nucleotides per minute at each replication fork, considerably slower than bacterial replication but still replicate in a matter of minutes or hours, not days. This rate is possible because replication takes place simultaneously from thousands of origins. Typical eukaryotic replicons are-from 20,000 to 300,000 base pairs in length. At each replication origin, the DNA unwinds and produces a replication bubble. Replication takes place on both strands at each end of the bubble, with the two replication forks spreading outward. Eventually, replication forks of adjacent replicons run into each other, and the replicons fuse to form long stretches of newly synthesized DNA **(Fig. 2.7).** Replication and fusion of all the replicons leads to two identical DNA molecules.



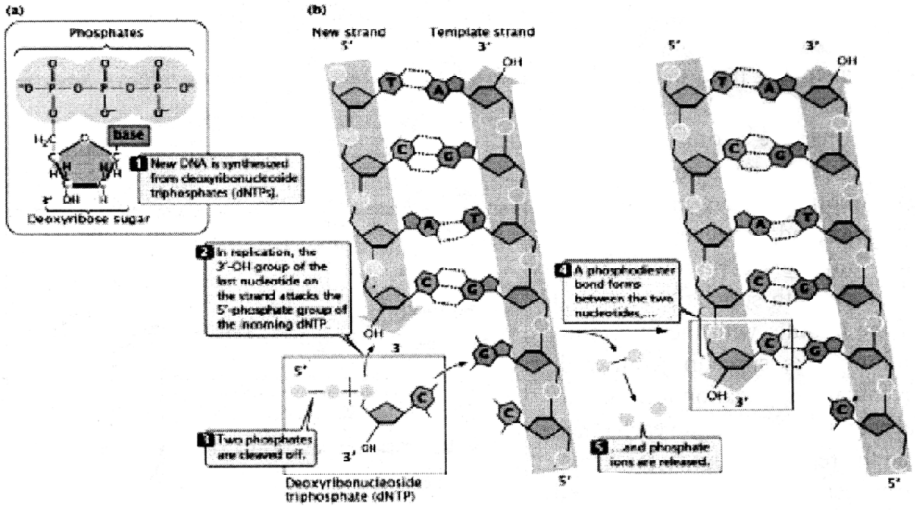
Fie. 2.7: DNA replication on linear chromosomes

Like all other metabolic processes, DNA replication is under the control of several proteins and enzymes, engaged in an intricate and coordinated interplay. Our understanding of DNA replication is primarily derived from physical, chemical and biochemical studies of enzymes and nucleic acids from *Es-cherichia coli,* their phages and their mutants. Prokaryotic and eukaryotic mechanism of DNA replication differs in many ways though the basic mechanisms are same.

2.3 DNA replication model

In the simplest model of DNA replication, the mother DNA strand is unzipped to produce a 'Y' shaped replication fork. At the replication fork, enzymes and protein factors facilitate the addition nascent nucleotides to the newly forming

DNA strand by way of complimentary base pairing and the event should occur simultaneously on both the strands. During DNA synthesis, nucleotides are added to the 3'-OH group of the growing nucleotide strand (Fig. 2.8). The 3'-OH group of the last nucleotide on the strand attacks the 5'-phosphate group of the incoming dNTP. Two phosphates are cleaved from the incoming dNTP, and a phosphodiester bond is created between the two nucleotides.



Fie: 2.8: Inclusion of a nascent nucleotide into a growing DNA molecule

But there lies a problem. All known DNA polymerases can add nucleotides in the 5'⟶3' direction of the growing strand only. As the DNA double helix is anti-parallel in nature, simultaneous synthesis of both the strands is difficult to conceive as one strand will be synthesized in 5'⟶3' direction and the other strand has to be synthesized in 3 '⟶5' if both the strands has to be synthesized simultaneously - which is not possible. Interestingly, mother-nature has evolved mechanism that allows both the strands of the DNA double helix to be synthesized simultaneously.

2.3.1 Continuous and discontinuous DNA replication

Auto radiographic studies confirmed that DNA synthesis occurs simultaneously on both the strands. The works of Okazaki and his colleagues enabled to explain the basic mechanism underlying the simultaneous synthesis of both the strands in DNA double helix. They studied the incorporation of radioactive thymidine at different phases of DNA synthesis and found that the radioactive materials were present only in short DNA fragments (100- 1000 nucleotides) extracted just a few moments after the radioactive pulse was inhibited. As the time

elapsed, the radioactive materials could be detected in high molecular weight DNA strands. Normally, this should not have happened as the feeding of radioactive material was stopped. They predicted that during DNA synthesis, continuous synthesis occurs on the 3'→5' template but on the 5'→3' template short DNA segments called Okazaki **fragments** are synthesized, which are subsequently linked together by the action of DNA polymerases. Thus, **continuous replication** occurs in one of the templates in the direction of the movement of the replication fork while **discontinuous replication** occurs in the other strand. The strand that allows continuous synthesis in known as **leading strand** and the strand on which discontinuous synthesis occurs is called the **lagging strand** (Fig. 2.9).

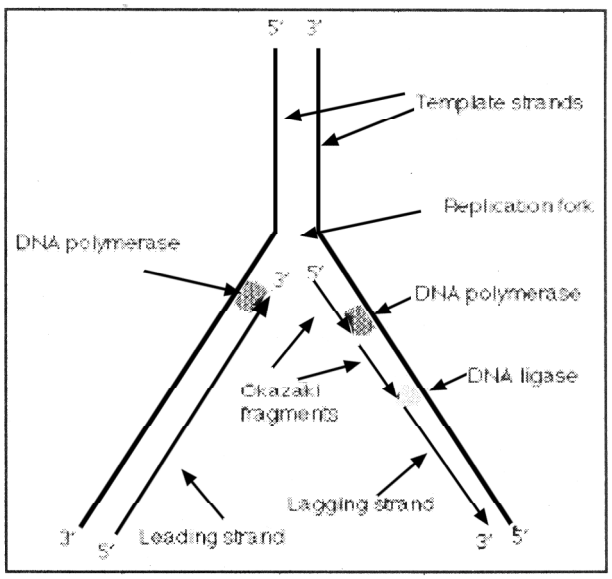


Fig: 2.9: DNA synthesis is continuous on one template strand of DNA and discontinuous on the other.

2.3.3 The fidelity of DNA replication

Overall, replication results in an error rate of less than one mistake per billion nucleotides. How is this incredible accuracy achieved? Answer lies in the activity of the DNA polymerase. These enzymes are very particular in pairing nucleotides with their complements on the template strand. Most of the errors that do arise in nucle-otide selection are corrected in a second process called **proofreading.** When a DNA polymerase inserts an incorrect nucleotide into the growing strand, the 3'-OH group of the mispaired nucleotide is not correctly positioned for accepting the next nucleotide. The incorrect positioning stalls the polymerization reaction, and the 3'→5' exonuclease activity of DNA polymerase removes the incorrectly paired nucleotide. DNA polymerase then inserts the correct nucleotide. Together, proofreading and nucleotide selection result in an error rate of only one in 10 million nucleotides. A third process, called **mismatch repair** corrects errors after replication is complete. Any incorrectly paired nucleotides produce a deformity in the secondary structure of the DNA; the deformity is recognized by specialized enzymes that excise an incorrectly paired nucleotide and replace it with the correct

nucleotide. Methylation on the old DNA strand allows the mismatch repair enzymes to distinguish between old and new strand.

2.3.3 Speed of replication :

The single molecule of DNA that is the *E. colt* genome contains 4.7 x 10^6 nucle-otide pairs. DNA replication begins at a single, fixed location in this molecule, the **replication origin,** proceeds at about 1000 nucleotides per second, and thus is done in no more than 40 minutes. And thanks to the precision of the process (which includes a "proof-reading" function), the job is done with only about one incorrect nucleotide for every 10^9 nucleotides inserted. In other words, more often than not, the *E. colt* genome (4.7 x 10^6) is copied without error.

2.4 Replication in eukaryotes

Our understanding of the DNA replication in eukaryotic cells is limited but development of new experimental techniques is rapidly changing the imbalance of knowledge. The replication machinery is similar to bacterial system but basic differences include: (1) replication on linear chromosomes associated with multiple proteins, (2) multiple replication origins in their chromosomes; (3) more types of DNA polymerases, with different functions; and (4) nucleosome assembly immediately following DNA replication. As the yeast *(Saccharomyces cerevisiae)* replication system is very similar to mammalian cells, isolation of various mutant yeasts, unable to produce specific gene products required for various aspects of replication has added to our understanding of eukaryotic replication. Further, the monkey virus **SV40** has single origin of replication where the viral encoded **large T antigen** binds along with several other proteins that are synthesized by the host DNA. Scientists utilize this knowledge to simulate DNA replication *in vitro* and understand the function of various cellular replication proteins.

2.4.1 Initiation of replication

In eukaryotes, cells have much more DNA and their polymerases synthesize DNA at a much lower rate. To compensate these difficulties, the eukaryotic cells replicate their genome in small portions, termed **replicons** (Fig. 2.10). From the radioactive studies it has been estimated that each replicon is approximately 15 to 100 urn in length (50 to 300 kb) and the replication fork proceeds in both direction. The heterochromatin regions tend to replicate late in the S phase as such the Barr body is the last to replicate while the active X chromosome is replicated at an early stage in females.

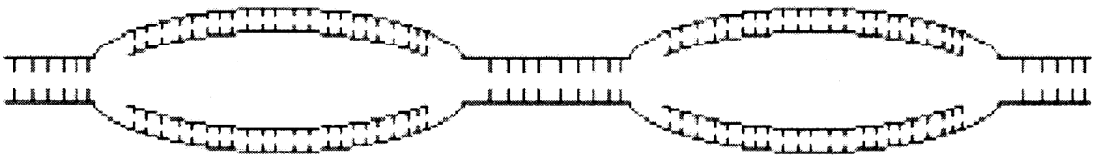


Fig: 2.10: Multiple site of origin of DNA replication and each is known is replicon

Initiation of replication in eukaryotes is much more complicated than in prokaryotes. In yeast cells, the site of origin of replication if removed and inserted in another DNA molecule, the hybrid DNA molecule acquires the ability to replicate *in vitro* or *in vivo.* As the sequences at the site of origin of replication promote replication of the DNA in which they are contained, they are referred to as **autonomous replicating sequences (ARSs).** There are about 400 ARSs scattered throughout the genome of yeast cells and each ARSs has a conserved 11 bp sequence that allow the binding of essential multiprotcin complex called the **origin recognition complex (ORC).** In lormal cells, ORC proteins remain bound to the ARSs all through the cell cycle. The binding of other proteins to the ORC-ARS complex allows initiation of replication. Mutated ARSs fail to bind ORC proteins and thus cannot initiate DNA replication at that site.

In mammals, virtually any type of purified naked DNA is suitable for initiation of replication **with** cellular extracts suggesting that, unlike yeast, mammalian DNA might not possess specific sites at which replication is initiated. However, in vivo studies on intact chromosomes indicate that replication does begin at specific sites along the DNA and initiation is not a random event. It appears that, mammalian DNA molecules have numerous sites where replication can be initiated, but because of the presence of nucleosome and higher order of organization of mammalian chromosome, most of the initiation sites remain suppressed while promoting initiation at specific sites that serve as replication origins. One such replication origin site has been located in the β-globin gene cluster.

As the Eukaryotic cells utilize thousands of origins, the cell needs to ensure that each segment of the DNA is replicated once during cell division. The precise replication of DNA is accomplished by the separation of the initiation of replication into two distinct steps. In the first step, the origins are licensed (see licensing factors), meaning that they are approved for replication. During replication, only the licensed sites can bind to replication initiation factors. The preliminary initiation factors at first displace the **replication licensing factors** and then induce the formation of replication bubble. The sites of origin of replication do not bind to any further replication licensing factors during the progression of the S phase until it enters

the mitotic phase thereby ensuring the replication of genome only once per cell cycle.

Two-replication fork are formed at each site of origin of replication and bidirectional DNA synthesis occur in a manner, which is similar in all organisms - whether it is virus, prokaryotes or eukaryotes. The replication forks are not randomly distributed in the nuclear matrix. There are 50-250 sites in the nucleus - called **replication foci,** where synthesis takes place. Each foci contain approximately active 40 replication forks. The clustering of replication forks may provide mechanisms for coordinating the replication of adjacent replicons over individual chromosomes. The nuclear matrix also seems to play an important during replication. The substances necessary for replication remain bound the nuclear matrix and are made available during the process of replication.

DNA replication in eukaryotic cells is limited to S phase of the cell cycle. Approximately 10^3 to 10^5 replication events occur in a coordinated manner, though not identically at all origins. This leads to great variation in the duration of S phase. Moreover, the associated histones and non histone proteins get synthesized either during.Gl or S phase.

Replication **'tool kit'** consists of helicase, single stranded DNA binding proteins, topoisomerases, primase, DNA polymerase, and DNA ligase. The DNA in eukaryotes is also synthesized in semi-discontinuous manner, although the Okazaki fragments of the lagging strand are much smaller (250 nucleotides in length).

2.4.2 Some proteins required for replication

The following table shows a glimple of different proteins required for replication.

DNA Polymerase Sub Units	Polymeras Activity	Exonuclease Activity	Cellular Function
α (alpha)	Yes	No	Initiation of nuclear DNA synthesis and DNA repair
β (beta)	Yes	No	DNA repair and recombination of nuclear DNA
γ (gamma)	Yes	Yes	Replication of mitochondrial DNA
δ (delta)	Yes	Yes	Leading & lagging-strand synthesis of nuclear DNA, DNA repair, and translesion DNA synthesis
ε (epsilon)	Yes	Yes	Unknown; probably repair and replication of nuclear DNA
ζ (zeta)	Yes	No	Translesion DNA synthesis
η (eta)	Yes	No	Translesion DNA synthesis

DNA Polymerase Sub Units	Polymerase Activity	Exonuclease Activity	Cellular Function
θ (theta)	Yes	No	DNA repair
ι (iota)	Yes	No	Translesion DNA synthesis
κ (kappa)	Yes	No	Translesion DNA synthesis
λ (lambda)	Yes	No	DNA repair
μ (mu)	Yes	No	DNA repair
σ (sigma)	Yes	No	Nuclear DNA replication (possibly), DNA repair, and sister-chromatid cohesion

Other DNA polynferases (ζ, η, θ, κ, λ, μ) allow replication to bypass damaged DNA (called translesion replication) or play a role in DNA repair. Many of the DNA polymerases have multiple roles in replication and DNA repair.

2.4.3 Eukaryotic DNA polymerase

Eukaryotic cells contain several DNA polymerases of which the most important are α,β,μ,γ,δ and ε (Table 1). **DNA polymerase** α, which contains primase activity, initiates nuclear DNA synthesis by synthesizing an RNA primer, followed by a short string of DNA nucleotides. After DNA polymerase α has laid down from 30 to 40 nucleotides, **DNA polymerase** δ completes replication on the leading and lagging strands. **DNA polymerase** β does not participate in replication but is associated with the repair and recombination of nuclear DNA. The γ polymerase is encoded by nuclear gene but located within the mitochondria and is responsible for the replication of the mitochondrial DNA;. γ polymerase like enzyme replicates chloroplast DNA in plants. Similar in structure and function to DNA polymerase δ , **DNA polymerase** e appears to take part in nuclear replication of both the leading and the lagging strands, but its precise role is not yet clear.

Inhibitors of polymerase activity : Aphidicolin (α, δ, & ε); N-ethylmalemide (α, β, δ, d); butylphenyl-dGTP (α,); dideoxynucleoside 5'triphosphate (γ).

2.4.4 The enzymes of DNA replication

1. **Topoisomerase** is responsible for initiation of the unwinding of the DNA. The tension holding the helix in its coiled and supercoiled structure can be broken by nicking a single strand of DNA. Try this with string. Twist two strings together, holding both the top and the bottom. If you cut only one of the two strings, the tension of the twisting is released and the strings untwist.

2. **Helicase** accomplishes unwinding of the original double strand, once supercoifing has been eliminated by the topoisomerase. The two strands very much want to bind together because of their hydrogen bonding affinity for each other, so the helicase activity requires energy (in the form of ATP) lo break the strands apart.
3. **DNA polymerase** proceeds along a single-stranded molecule of DNA, recruiting free dNTP's (deoxy-nucleotide-triphosphates) to hydrogen bond with their appropriate complementary dNTP on the single strand (A with T and G with C), and to form a covalent phosphodiester bond with the previous nucleotide of the same strand. The energy stored in the triphos-phate is used to covalently bind each new nucleotide to the growing second strand. There are different forms of DNA polymerase, but it is DNA polymerase III that is responsible for the processive synthesis of new DNA strands. DNA polymerase cannot start synthesizing de novo on a bare single strand. It needs a **primer** with a 3'OH group onto which it can attach a dNTP. DNA polymerase is actually an aggregate of several different protein subunits, so it is often called a holoenzyme. The holoenzyme also has proofreading activity, so that it can make sure that it inserted the right base, and nuclease (excision of nucleotides) activities so that it can cut away any mistakes it might have made.
4. **Primase** is actually part of an aggregate of proteins called the **primeosome.** This enzyme attaches a small RNA primer to the single-stranded DNA to act as a substitute 3'-OH for DNA polymerase to begin synthesizing from. This RNA primer is eventually removed by **RNase H** and the gap is filled in by **DNA polymerase I.**
5. **Ligase** can catalyze the formation of a phosphodiester bond given an unattached but adjacent 3'-OH and 5'phosphate. This can fill in the unattached gap left when the RNA primer is removed and filled in. The DNA polymerase can organize the bond on the 5' end of the primer, but ligase is needed to make the bond on the 3' end.
6. **Single-stranded binding proteins** (SSB) are important to maintain the stability of the replication fork. Single-stranded DNA is very labile, or unstable, so these proteins bind to it while it remains single stranded and keep it from being degraded.

2.4.5 The replication fork

Why can DNA polymerase only act from 5' to 3' The reason is the relative stability of each end of DNA. A triphosphate is required to provide energy for the bond between a newly attached nucleotide and the growing DNA strand. However,

this triphosphate is very unstable and can easily break into a monophosphate and an inorganic pyrophosphate, which floats away into cell. At the 5' end of the DNA, this triphosphate can easily break, so if a strand has been sitting in the cell for a while, it would not be able to attach new nucleotides to the 5' end once the phosphate had broken off. On the other hand, the 3' end only has a hydroxyl group, so as long as new nucleotide triphosphate are always brought by DNA polymerase, synthesis of a new strand can continue no matter how long the 3' end has remained free.

This presents a problem, since one strand of the double helix is 5' to 3', and the other one is 3' to 5'. How can DNA polymerase synthesize new copies of the 5' to 3^1 strand, if it can only travel in one direction? This strand is called the **lagging strand,** and DNA polymerase makes a second copy of this strand in spurts, called **Okazaki fragments,** as shown in the diagram. The other strand can proceed with synthesis directly, from 5' to 3', as the helix unwinds. This is the **leading strand.**

2.4.6 Nucleosome assembly

ukaryotic DNA is complexed to histone proteins in nucleosome structures that contribute to the stability and packing of the DNA molecule (see Fig. 2.11). The disassembly and reassembly of nucleosomes on newly synthesized DNA probably takes place during replication, but the precise mechanism for these processes has

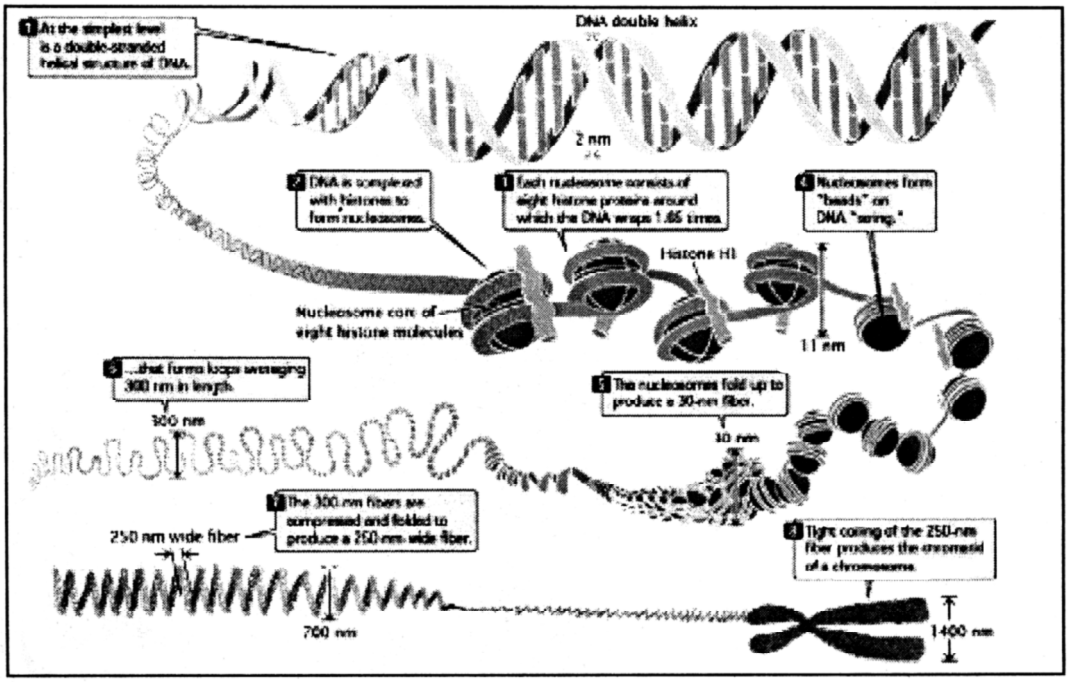


Fig. 2.11: Several levels of organization of eukaryotic chromosome.

not yet been determined. The unwinding of double stranded DNA and the assembly of the replication enzymes on the single-stranded templates probably require the disassembly of the nucleosome structure. Electron micrographs of eukaryotic DNA show recently replicated DNA already covered with nucleosomes indicating that nucleosome structure is reassembled quickly. Before replication, a single DNA molecule is associated with histone proteins. After replication and nucleosome assembly, two DNA molecules get associated with histone proteins. After replication and nucleosome assembly, two DNA molecules get associated with histone proteins. Whether the original histones remain together, attached to one of the new DNA molecules, or do they disassemble and mix with new histones on both DNA molecules is still not known. Experiments with radioactive labeled histones suggest that newly assembled octamers consist of a random mixture of old and new histones.

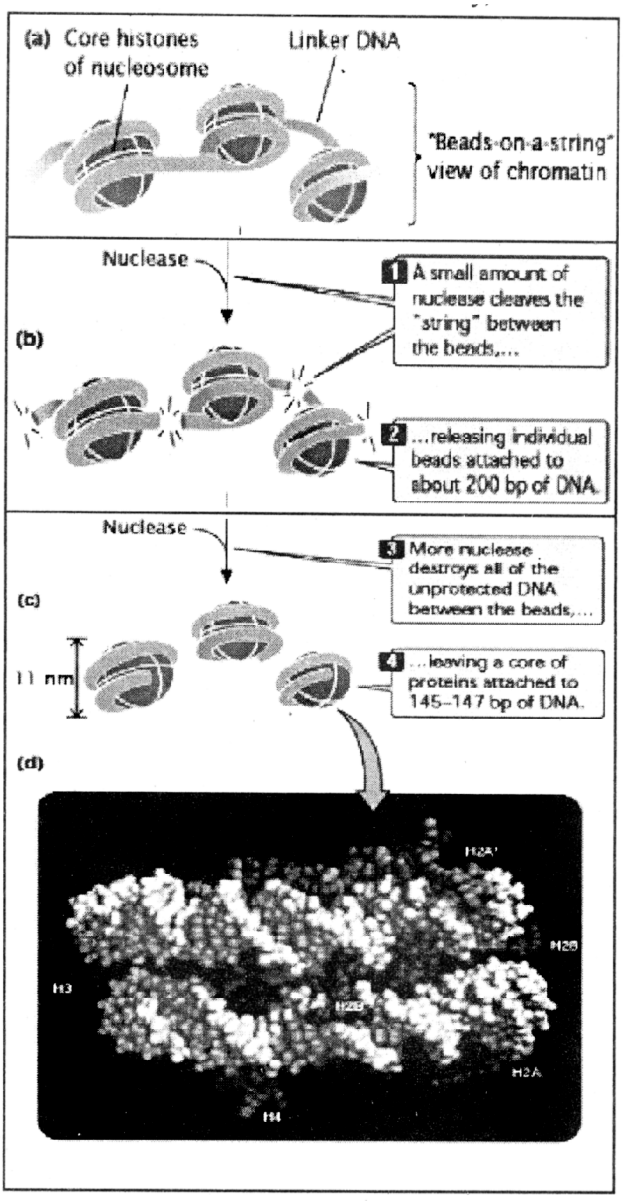


Fig. 2.12: The nucleosome model

2.4.7 The nucleosome

Chromatin has a highly complex structure with several levels of organization. The simplest level (Fig. 2.12) is the double helical structure as proposed by Watson and Crick (1953). At a more complex level, the DNA molecule is associated with proteins and is highly folded to produce a chromosome. Chromatins viewed under electron microscope, frequently looks like beads on a string. Partial digestion of chromatin with nuclease produces beads. Each individual bead has attached 200

bp of DNA. Further digestion with more nuclease chews up the entire DNA between the beads and leaves a core of proteins attached to a fragment of DNA (Fig. 2.12). These experiments demonstrated that chromatin is not a random association of proteins and DNA but has a fundamental repeating structure having the the simplest level of chromatin structure, the **nucleosome**

The nucleosome is a core particle consisting of DNA wrapped about two times around an octamer of eight histone proteins (two copies each of H2A, H2B, H3, and H4), much like thread wound around a spool (Fig. 2.12d). The DNA in direct contact with the histone octamer is between 145 and 147 bp in length, coils around the histones in a left-handed direction, and is supercoiled. It does not or kinks, in its helical structure as it winds around the histones.

The fifth type of histone, **H1,** is not a part of the core particle but plays an important role in the nucleosome structure. The precise location of H1 with respect to the core particle is still uncertain. The traditional view is that H1 sits outside the octamer and binds to the DNA where the DNA joins and leaves the octamer (Fig. 2.11). However, the results of recent experiments suggest that the H1 histone sits inside the coils of the nucleosome. Regardless of its position, H1 helps to lock the DNA into place, acting as a clamp around the nucleosome octamer. Together, the core particle and its associated H1 histone are called the **chromatosome,** the next level of chromatin organization. The H1 protein is attached to between 20 and 22 bp of DNA, and the nucleosome encompasses an additional 145 to 147 bp of DNA; so about 167 bp of DNA are held within the chromatosome. Chromatosomes are located at regular intervals along the DNA molecule and are separated from one another by wrap around the octamer smoothly; there are four bends, **linker DNA,** which varies in size among cell types—most cells have from about 30 bp to 40 bp of linker DNA. Nonhistone chromosomal proteins may be associated with this linker DNA, and a few also appear to bind directly to the core particle.

2.4.8 DNA synthesis at the ends of chromosomes

A fundamental difference between eukaryotic and bacterial replication arises because eukaryotic chromosomes are linear and thus have ends. As the 3'-OH group is needed by DNA polymerases to elongate, at the initiation of replication by RNA primers provide the 3 '-OH group synthesized by primase. RNA primers must be removed and replaced by DNA

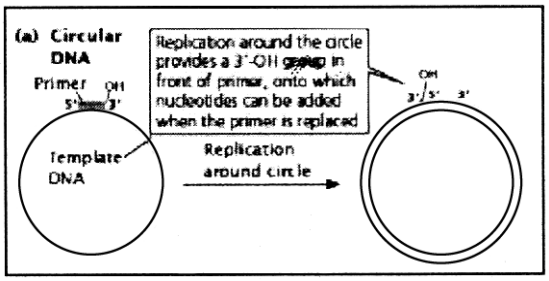


Fig. 13a: Replication at the ends of circular DNA where the 3'-OH group is available

Fig. 13a: Replication at the ends of circular DNA where the 3 '-OH group is available

mucleotides subsequently which is done by DNA polymerase I. In a circular DNA molecule, elongation around the circle eventually provides a 3'-OH group immediately in front of the primer (Fig. 2.13a). After the primer has been removed, the replacement DNA nucleotides can be added to this 3'-OH group.

In linear chromosomes with multiple origins, the elongation of DNA in adjacent replicons also provides a 3'-OH group preceding each primer (Fig. 2.13b.). At the very end of a linear chromosome, however, there is no adjacent stretch of replicated DNA to provide this crucial 3'-OH group. Once the primer at the end of the chromosome has been removed, it cannot be replaced with DNA nucleotides, which produces a gap at the end of the chromosome (Fig. 2.13c), suggesting that the-chromosome should become progressively shorter with each round of replication, leading to the eventual elimination of the entire telomere and destabilization of the chromosome, and cell death. But chromosomes don't become shorter each generation and destabilize. Interestingly, the ends of chromosomes that preserve the integrity of the chromosome structure and avoid being getting shorter with each cycle of cell division.

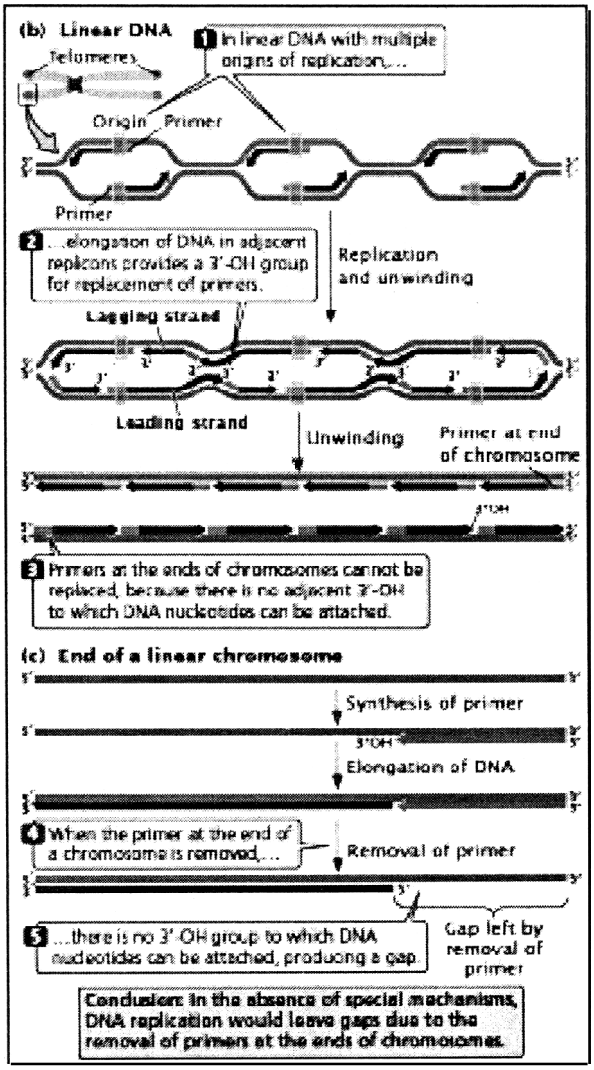


Fig. 2.13b : Replication at the ends of linear DNA

The telomeres possess several unique features, one of which is the presence of many copies of a short repeated sequence. In the protozoan *Tetrahymena,* this telomeric repeat is CCCCAA., with the G-rich strand typi-cally protruding beyond the C-rich strand (Fig. 2.14a): The single-stranded protruding end of the telomere can be extended by **telomerase,** an enzyme with both a protein and an RNA component (also known as a ribonucleoprotein). The RNA part of the enzyme contains from 15 to 22 nucleotides

that are complementary to the sequence on the G-rich strand. This sequence pairs with the overhanging 3' end of the DNA (Fig.2.14b) and provides a template for the synthesis of additional DNA copies of the repeats. DNA nucleotides are added to the the end of strand one at a time (Fig.2.14c) and, after several nucleotides have been added, the RNA template moves down the DNA and more nucleotides are added to the 3' end. Usually, from 14 to 16 nucleotides are added to the 3' end of the G-rich strand.

5' end of	3-CCCCAA	toward
Chromosome	3-GGGGTTGGGGTT	centromere

In this way, the telomerase can extend the 3' end of the chromosome without the use of a complementary DNA template. How the complementary Orich strand is synthesized is not yet clear. It may be synthesized by conventional replication, with priniase synthesizing an RNA primer on the 5' end of the extended (G rich) template. The removal of this primer once again leaves a gap at the 5' end of the chromosome, but this gap does not matter, because the end of the chromosome is extended at each replication by telomerase; no genetic information is lost, and the chromosome does not become shorter overall. The extended single-strand end may fold back on itself, forming a terminal loop by nonconventional pairing of bases, displacing a part of the original telomeric duplex. The loop structure is formed and stabilized by specific telomere-binding proteins (Fig. 2.15).

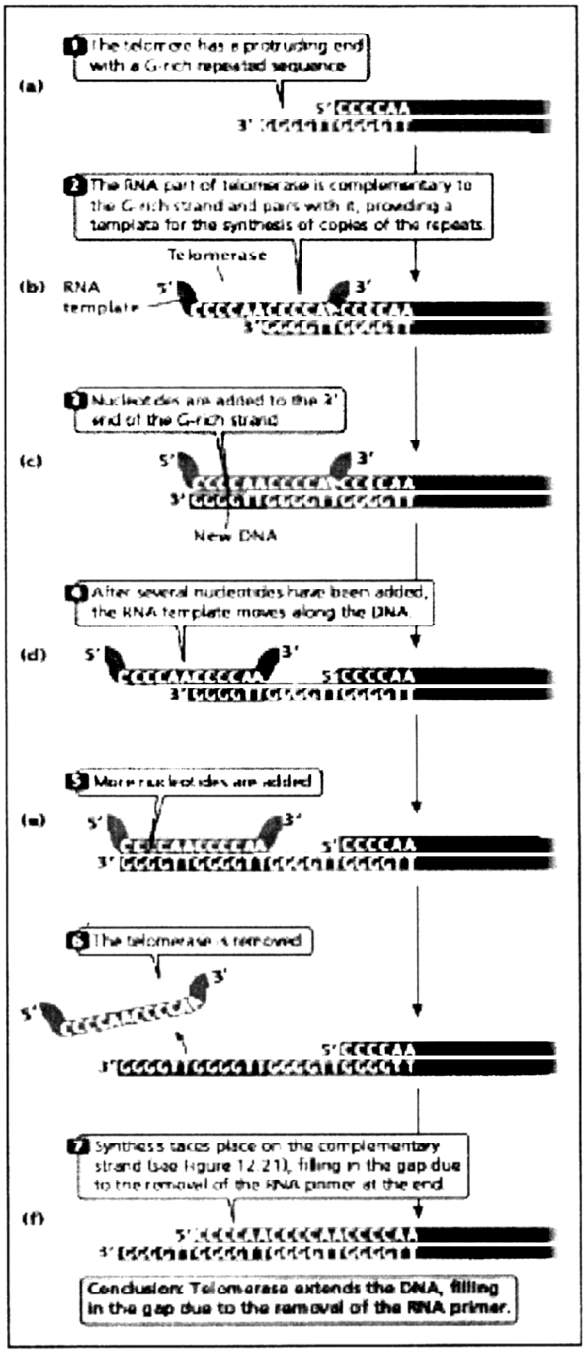


Fig. 2.14: DNA synthesis at the end point of linear chromosomes by telomerase enzyme

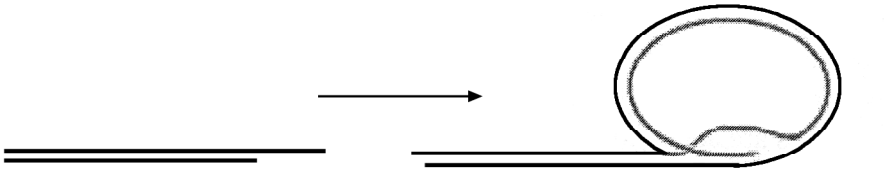


Fig. 2.15: Loop formation by the single stranded DNA strand at telomeric region by unconventional base pairing

2.4.9 Telomerase enzyme

Telomerase is a multi-subunit enzyme that is comprised of a RNA component - hTR, and a protein component - hTERT (Nakamura and Cech, 1998). hTR contains an 11 bp sequence that provides the template for the synthesis of telomeric repeats which are added to the chromosome, whereas hTERT, the reverse transcriptase component, catalyzes the synthesis reaction. Thus, addition of TTAGGG repeats to the 3' ends of chromosomes compensates for losses due to the end-replication problem. In humans, telomerase activity is absent in most normal cells but present in majority of tumors (Kim *et al,* 1994). However, activity has been detected in high levels in germ cells, early embryos (Xu and Yang, 2001), activated T and B cells and germinal centres of lymphoid organs.

Telomerase is present in single-celled organisms, germ cells, early embryonic cells, and certain proliferative somatic cells (such as bone-marrow cells and cells lining the intestine), all of which must undergo continuous cell division. Most somatic cells have little or no telomerase activity, and chromosomes in these cells progressively shorten with each cell division. These cells are capable of only a limited number of divisions; once the. telomeres shorten beyond a critical point, a chromosome becomes unstable, has a tendency to undergo rearrangements, and is degraded. These events lead to cell death. The shortening of telomeres may contribute to the process of aging. Genetically engineered mice that lack a functional telomerase gene do not express telomerase in somatic or germ cells and therefore experience progressive shortening of their telomeres in successive generations. After several generations, these mice show some signs of premature aging, such as graying, hair loss, and delayed wound healing. Through genetic engineering, it is also possible to create somatic cells that express telomerase. In these cells, telomeres do not shorten, cell aging is inhibited, and the cells will divide indefinitely. Telomerase also appears to play a role in cancer. Cancer tumor cells have the capacity to divide indefinitely, and many tumor cells express the telomerase enzyme. Telomerase activation alone does not lead to cancerous growth in most cells, but it does appear to be required along with other mutations for cancer to develop.

The length of the telomeric sequence varies from chromosome to chromosome and from cell to cell, suggesting that each telomere is a dynamic structure that actively grows and shrinks. The telomeres of *Drosophila* chromosomes are different in structure. They consist of multiple copies of the two different retrotransposons *Het-A* and *Tart,* arranged in tandem repeats. Apparently, in *Drosophila,* loss of telomere sequences during replication is balanced by transposition of additional copies of the *Het-A* and *Tart* elements. Farther away from the end of the chromosome, from several thousand to hundreds of thousands of base pairs form telomere-associated sequences. They, too, contain repeated sequences, but the repeats are longer, more varied, and more complex than those found in telomeric sequences.

2.4.10 Licensing: positive control of replication

The average human chromosome contains 150 x 10^6 nucleotide pairs which are copied at about 50 base pairs per second. The process would take a month (rather than the hour it actually does) but for the fact that there are many places on the eukaryotic chromosome where replication can begin. Replication begins at some replication origins earlier in S phase than at others, but the process is completed for all by the end of S phase. As replication nears completion, "bubbles" of newly replicated DNA meet and fuse, finally forming two new molecules.

In order to be replicated, each origin of replication must be bound by:

- An **O**rigin **R**ecognition **C**omplex of proteins (**ORC**). (These remain on the DNA throughout the process).
- Accessory proteins called **licensing factors**. (These accumulate in the nucleus during G_1 of the cell cycle. They include) :

  CDC-6 and CDT-1, which bind to the **ORC** and are essential for coating the DNA with

  MCM proteins. Only DNA coated with MCM proteins (there are 6 of them) can be replicated.
- Once replication begins in S phase,

  CDC-6 and CDT-1 leave the ORCs (the latter by ubiquination and destruction in proteasomes).

  The MCM proteins leave in front of the advancing replication fork.

2.4.11 Geminin: negative control of replication

G_2 nuclei also contain at least one protein — called **geminin** — that prevents assembly of MCM proteins on freshly-synthesized DNA (probably by sequestering Cdtl).

As the cell **completes mitosis,** geminin is degraded so the DNA of the two daughter cells will be able to respond to licensing factors and be able to replicate their DNA at the next S phase.

Some cells deliberately cut the cell cycle short allowing repeated S phases without completing mitosis and/or cytokinesis. This is called **endoreplication.** How these cells regulate the factors that normally prevent DNA replication if mitosis has not occurred is still being studied. Endoreplication is described on a separate page.

2.4.12 Post-replicative modification of DNA, methylation

One of the major post-replicative reactions that modifies the DNA is **methylation.** The sites of natural methylation (i.e. not chemically induced) of eukaryotic DNA is always on cytosine residues that are present in CpG dinucleotides. However, it should be noted that not all CpG dinucleotides are methylated at the C residue. The cytidine is methylated at the 5 position of the pyrimidine ring generating 5-methylcytidine.

Methylation of DNA in prokaryotic cells also occurs. The function of this methylation is to prevent degradation of host DNA in the presence of enzymatic activities synthesized by bacteria called restriction endonucleases. These enzymes recognize specific nucleotide sequences of DNA. The role of this system in prokaryotic cells (called the restriction-modification system) is to degrade invading viral DNAs. Since the viral DNAs are not modified by methylation they are degraded by the host restriction enzymes. The methylated host genome is resistant to the action of these enzymes.

The precise role of methylation in eukaryotic DNA is unclear. It was originally thought that methylated DNA would be less transcriptionally active. Indeed, experiments have been carried out to demonstrate that this is true for certain genes. For example, under-methylation of the MyoD gene (a master control gene regulating the differentiation of muscle cells through the control of the expression of muscle-specific genes) results in the conversion of fibroblasts to myoblasts. The experiments were carried out by allowing replicating fibroblasts to incorporate 5-azacytidine into their newly synthesized DNA. This analog of cytidine prevents methylation. The net result is that the maternal pattern of methylation is lost and numerous genes become under methylated. However, lack of methylation nor the presence of methylation is a clear indicator of whether a gene will be transcriptionally active or silent.

The pattern of methylation is copied post-replicatively by the **maintenance methy-lase system**. This activity recognizes the pattern of methylated C residues

in the maternal DNA strand following replication and methylates the C residue present in the corresponding CpG dinucleotide of the daughter strand.

The phenomenon of **genomic imprinting** refers to the fact that the expression of some genes depends on whether or not they are inherited maternally or paternally. **Insulin-like growth factor-2 (IGF2)** is a gene whose expression is required for normal fetal development and growth. Expression of IGF2 occurs exclusively from the paternal copy of the gene. Imprinted genes are "marked" by their state of methylation. In the case of IGF2 an element in the maternal locus, called an **insulator element,** is methylated blocking its function. The function of the un-methylated insulator is to bind a protein that when bound blocks activation of IGF2 expression. When methylated the protein cannot bind the insulator thus allowing a distant enhancer element to drive expression of the IGF2 gene. In the maternal genome, the insulator is not methylated, thus protein binds to it blocking the action of the distant enhancer element.

2.5 Mechanism of replication

2.5.1 Replication in prokaryotes

Replication in prokaryotes and viral DNA usually starts at specific sites on chromosome, referred to as **replication origin.** The origin in E. coli is specifically known as ***OriC***. Approximately 20-30 different proteins, some in multiple copies are required to initiate the DNA replication process. Some of the proteins characterized and their genes are given in Table 1.

Initiation and Unwinding of DNA: function of Helicases & Topoisomerase

In *E.coli,* ***OriC*** comprises of **245** base pairs that contains **four 9 bp** sites with similar sequences at which product of ***dnaA*** (homologous tetramer) binds and initiates the assembly of all other proteins and enzymes necessary for replication. In addition, the origin contains 11 methylation sites

Gene Product or/functions	Gene
Initiator protein; binds at OriC	*dnaA*
1HF protein-DNA binding protein; binds at OriC	*HimA*
F1S protein-DNA binding protein; binds at OriC	***fis***
Heiicase and activator of primase	*dnaB*
Proteins that complexes with dnaB protein and delivers to DNA	*dnaC*
Primase-synthesizes RNA primer	*dnaG*
Single stranded binding proteins (SSB-proteins)	*ssb*
DNA ligase	*Lig*
Gyrase (topoisomerase type. II)	*gyrA,, yrB*
TBP proteins (terminus (ter) binding proteins) stops replication	*Tus*
Topoisomerase I	*topA*
Topoisomerase IV	*parE*

Table 1: Different proteins that assimilate at the initiation complexes

recognized by **DNA methylase** and **three AT rich** direct tandem repeats consisting of 13 base pair each.

Initiation of replication begins with the binding of **dnaA** molecules at 4 sites consisting of 9 nucleotides provided that the 9-mers are fully methylated.

- Region of the DNA bound to dnaA coalesce and are joined by additional dnaA molecules to form a nucleosome like complex, which promotes nearby melting of the double helix at **AT rich** site (Fig. 2.16).

- Physical separation of the two strands requires untwisting of the DNA molecule. Unwinding of the DNA is facilitated by helicases, which is the product of the ***dnaB.*** dnaA, with the aid of dnaC binds to the helicases, which then makes contact with the DNA at the replication fork and separates the two strands to form a bubble by melting the hydrogen bonds. Energy required for unwinding is derived from the hydrolysis of ATP.

- During unwinding tension build up ahead of the replication fork because of the formation of super coils. DNA gyrase (topoisomerase

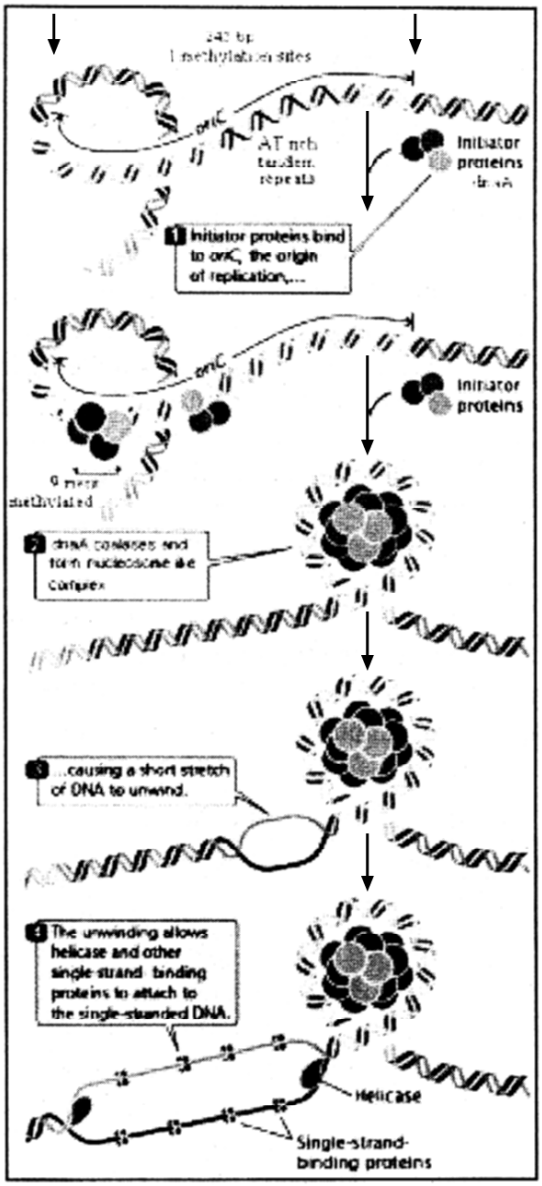


Fig. 2.16: Formation of initiation complex at the site of origin of DNA synthesis

I & **III** nicks single strand while **II** & IV nicks two strands to relieve tensions), reduce trosional strain (torque) that builds up ahead of the replication fork asa result of unwinding The topoiso-merase apparently nick one strand of the double stranded DNA ahead of the replication fork. The nicked DNA molecule then rotates, relieving the tension and avoid the formation of super coils. There are some indication that topoisomerase also induce negative super coils ahead of the replication fork thereby reliving the tensions and also assist the helicases in the process of unwinding **(Fig. 2.17).**

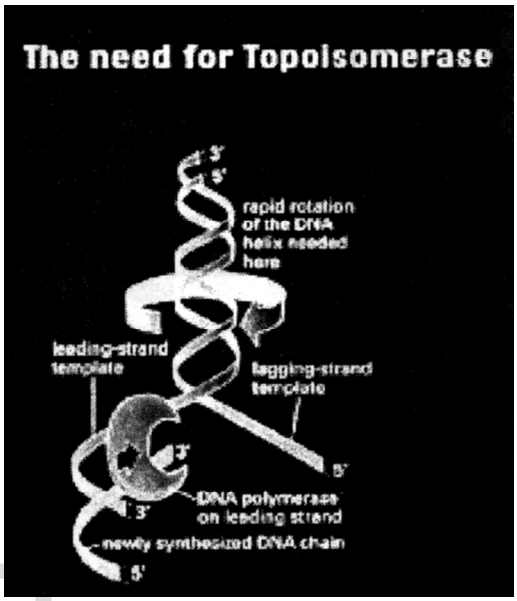


Fig. 2.17: Unwinding of the DNA helix ahead of the replication fork by topoisomerase releases tension and avoids DNA breaks.

2.5.2 Formation of Replisome:

The assembly of all protein factors and enzymes at the site of origin of replication is called a **replisome** (Fig. 2.18).

- Immediately after the bubble is formed **Single-strand binding proteins (SSBs)** binds to the single DNA strands and stabilizes them to avoid any unwanted reactions and- also to prevent the single strand DNA to reunite and form a duplex again, until replication at that region is complete. Formation of the bubble creates a '**Y**' shaped structure called a replication fork. A replication fork moves in one direction.
- Assembly of dnaB-dnaC complex is followed by the addition of four other poly-peptides - n, n', n" and i. This complex constitutes the prepriming complex.

The stage is now ready for the binding of the **Primase** the product of ***dnaG,*** Primase synthesize short RNA sequences complementary to DNA strands at the initiation site. This is because DNA polymerase **III** cannot initiate synthesis of a chain of DNA from free nucleotides and require a RNA primer to provide a free 3'-"OH end that can be extended by addition of nucleotides. Addition of Primase converts the priming complex to a **primosome.** The primase is much smaller that the usual RNA polymerase and is only 6000dalton. Primase is activated by dnaB, which then starts synthesizing short RNA primers on both the strands. The primers start with two purine nucleotides, most frequently pppAG The primers are usually

10 to 15 bases long. Assembly of the **replisome** is completed by the addition of **DNA polymerase III** (Fig. 2.18).

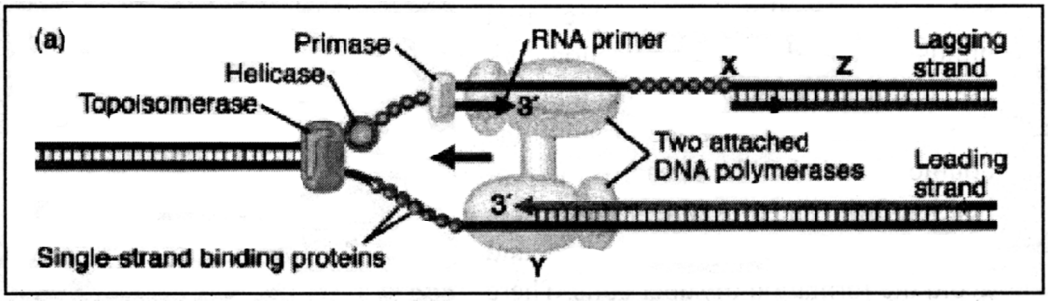


Fig. 2.18 : Replisome complex at the site of origin of DNA synthesis

2.5.3 Elongation

Once the single strands of DNA are stabilized, they serve as templates upon which new strands are synthesized by the enzyme referred to as DNA polymerase. DNA polymerases elongate the polynucleotide strand by catalyzing polymerization reaction. DNA polymerases add nucleotides to the 3-OH group on a nucleotide already paired with the template strand and consequently DNA synthesis can only proceed in the 5'–3' direction. Unlike the RNA polymerases, DNA polymerases require short primer sequences to initiate DNA synthesis on a single strand DNA template. Finally, the addition of new nucleotide is not random. DNA polymerase selects each deoxyribonucleotide that can form a complementary base pair with the nucleotide on the template strand DNA.

2.5.4 Termination of DNA synthesis

Termination occurs at *ter* or t locus, lying across from ***Ori C*** of the circular chromosome, between minutes 28 to 36. This region incorporates 6 **ter** sequences with sequence GTGTGTTGT that bind Tus protein **(terminator protein).** Three ter sequences arrest replication from the right and rest three sequences arrest replication coming from the left. One interesting aspect of E. coli DNA replication is that the cells are viable even if the whole terminator region is deleted and can terminate replication.

- Tus protein is a contra helicases, and functions by interfering with the ATP dependent function of dnaB helicases (rather than simply impending the propagation of this helicases along the double helix)
- Each Ter-Tus site has directional properties and it arrests only those replisomes that reach the Tus site from one specific direction.
- Replisome arriving from opposite direction apparently force dissociation of the TUS protein and thus can proceed unimpeded past the Ter-Tus site.

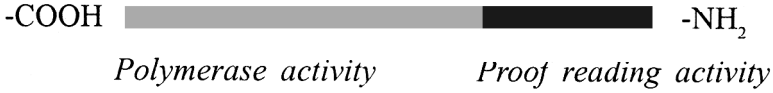
2.6 DNA Polymerases

The best-studied polymerases are those of *E. coli,* which has at least five different DNA polymerases. DNA polymerase I and DNA polymerase III play the major role during DNA replication, the pther three have specialized functions in DNA repair mechanism. All the DNA polymerases share the same fundamental property of adding new nucleotides only to the 3'-OH group on a previously existing paired nucleotide on a DNA template. Arthur Kornberg (1956) first identified and isolated DNA polymerase from the lysate of is. *coli,* which is now known as **DNA polymerase** I. Later, identification of DNA polymerase I deficient *E. coli* clones led to the isolation of two new polymerases - **DNA polymerase II** and **DNA polymerase III.** Characteristic properties of different polymerases isolated from *E. coli* are given in **Table 2.**

DNA Polymerase	5′→3′ Polymerization	3′→5′ Exonuclease	5′→3′ Exonuclease	Function
1	Yes	Yes	Yes	Removes and replaces primer
II	Yes	Yes	No	DNA repair: restarts replication after damaged DNA halts synthesis
III	Yes	Yes	No	Elongates DNA
IV	Yes	No	No	DNA repair
V	Yes	No	No	DNA repair: traitslesion DNA synthesis

Table 2: Properties of DNA polymerases found in *E. coli*

DNA polymerase I : DNA polymerase I participate in lagging strand synthesis by eliminating primer RNAs and also have a role in repair mechanism. This enzyme is coded by locus polA and is a single polypeptide chain. When treated with proteolytic enzymes, it is cleaved into two fragments- the larger fragment is know as **Klenow** fragment (used for *in vitro* synthesis). Two third of the protein chain, beginning from the C terminal end has polymerase activity while the rest one third on N-terminal end contains proofreading exonuclease activity, (no. of units per cell = 400)



DNA polymerase II: The biological function of DNA polymerase II is unclear, although this enzyme is induced when chromosomal DNA is damaged.

DNA polymerase III: DNA polymerase III is a huge multiprotein holoenzyme complex that plays the major role during DNA replication in *E. coli.* No. of units present per cell is approximately 10 to 12. It consists of 10 different polypeptide

chains. The catalytic core is composed of three subunits. The α subunit possess 5′→3′ DNA synthetic activity, ε subunit has the 3′ →5′ exonuclease activity. The θ subunit possibly participate in assembly of the enzyme. The remaining seven auxiliary subunits enhance the processivity of the core by clamping it onto the template. The addition of the τ causes the core to dimerize. The functions of γ and δ subunits are less well defined. The formation of holoenzyme is completed by the addition of the β ? subunit (Table 3). Structural analysis has revealed that DNA polymerase III is an asymmetric dimmer with twin polymerase activity sites, capable of synthesizing DNA strands simultaneously on both leading and lagging strand. The 5′ →3′ DNA synthetic activity and the 3′ →5′ exonuclease activity together allow DNA polymerase III to efficiently and accurately synthesize new DNA molecules.

DNA polymerase	***No. of Units***	***Gene***	***mol. wt.***	***Polymerization 5′→3′***	***Exonuclease activity***	***molecules. per cell***
I	One	*polA*	103.0 kd	Yes	3′ 5′ & 5′	3′ 400
II	One	*polB*	90.0 kd	Yes	3′ 5′	?
III	Ten	——		Yes	3′ 5′	10-12
α = dna*E*,130.0kd; ε = *dnaQ*, 27.5 kd; θ = *holE*, 10.0 kd; *T* (tau) = *dnaX*, 71.0 kd; γ = *dnaX*,47.0 kd						
β = *dnaN*,40.6 kd δ = *holA* 35.0 kd; δ = *holA*, 33.0kd; *x(chi)* = *holB*,15.0 kd: ψ(psi) =*hold;* 12.0 kd						

Table 3: Different subunits of the DNA polymerase III, their molecular weights and cellular function

All the *E. coli* DNA polymerases have 3′ →5′ exonuclease activity. This means that the DNA polymerases check the accuracy of the most recently assembled base pair. If a wrong base pair is included, exonuclease activity removes the erroneous

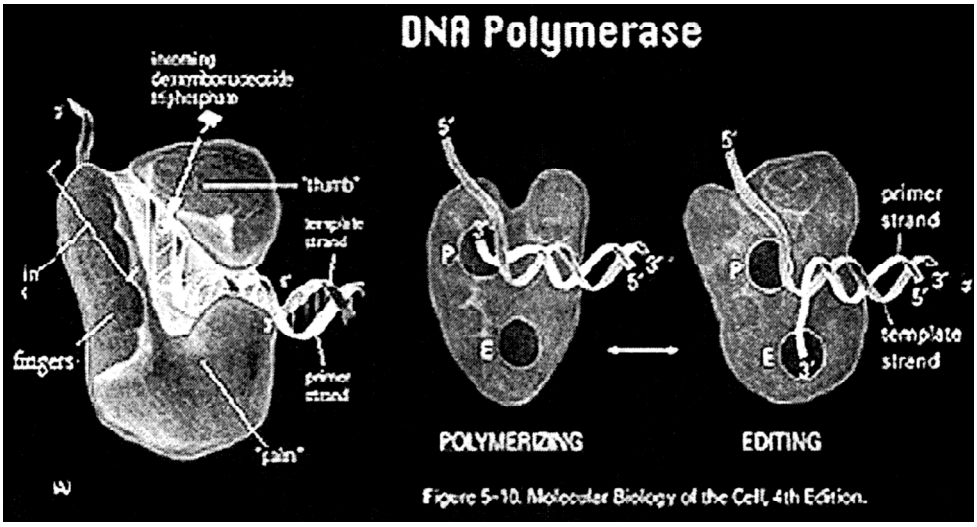


Fig. 2.19

nucleotide by excision and catalyzes the formation of the correct base pair. Thus in DNA replication 3′→5′ exonuclease activity is a proofreading mechanism that helps to keep the frequency of DNA replication errors at very low level (10^9/cycle). In addition, DNA polymerase I has 5′→3′ exonuclease activity and can remove nucfeotides from the DNA 5' end of a DNA strand or of an RNA primer strand. This activity is important for DNA repair.

2.6.1 DNA polymerase activity

- DNA synthesis begins immediately after the addition of DNA polymerase III by complementary base pairing at the 3'-OH of the primer.
- Because of the anti parallel nature of the DNA helix and the unidirectional movement of the DNA polymerase along the replication fork poses a problem, since DNA polymerase can only make new DNA strands in the 5' to 3' direction. Interestingly, both the strands of the DNA double helix are synthesized simultaneously.
- The 3′ →5′ template in the direction of fork movement is synthesized in a continuous manner and is called the **leading strand** (Fig. 2.20).
- The 5′ →3′ template in the direction of fork movement is synthesized in a discontinuous manner and is called the **lagging strand** (Fig. 2.20).
- DNA polymerase III complexes are endowed with the property to synthesize continuously on the leading strand and synthesize discontinuously on lagging strand.

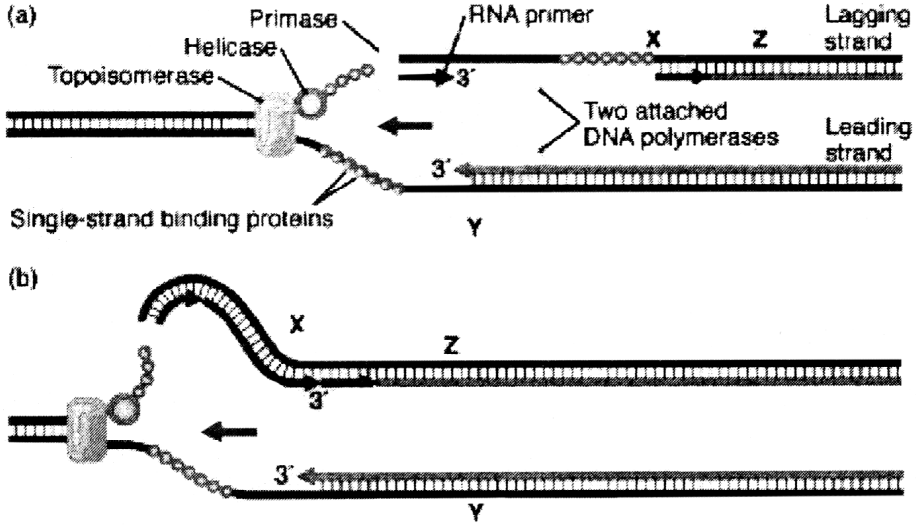


Fig. 2.20: (a) Different enzymes and factors at the site of replication fork,
(b) The replication on leading and lagging strand

- Lagging strand is synthesized discontinuously, in short stretches- known as **Okazaki fragments.** Formation of Okazaki fragments is also initiated by **Primase** at sites selected by pre-priming proteins. Each Okazaki fragments starts with a primer - a sequence of RNA, approximately 10 bases long that provides 3'-OH end for extension by DNA polymerase III.

- When a nascent Okazaki fragment reaches the 5 end of the previously synthesized Okazaki fragment, the lagging strand template is released and un-looped.
  The RNA primers of the Okazaki fragments in *E. coli* are removed by the combined activity of **RNase H** and **DNA polymerase I** that fills the gap. DNA polymerase I follows DNA polymerase III and, using its 5′ →3′ exo-nuclease activity removes the RNA primer. It then uses its 5′ →3′ polymerase activity to replace the RNA nucleotides with DNA nucleotides one at a time.

- Two Okazaki fragments are joined by DNA **ligase** producing intact DNA daughter strand.

Unit 3 Prokaryotic Transcription

Structure

3.1 Introduction

Genomic DNA contains a set of information that governs all cellular activities. These instructions are implemented by synthesis of RNA and proteins. Genetic character of a cell is determined by not only what genes are inherited, but also on how and when the genes are expressed.

Three years after Watson and Crick (1953) published the DNA model, Crick proposed the Central Dogma, describing the two-step process of protein synthesis. According the theme, flow of information is always unidirectional i.e. DNA to RNA and RNA to Protein. However after the discovery of reverse transcriptase in retro viruses by Temin and Baltimore in 1970, the dogma now stands as follows (Fig. 3.1)

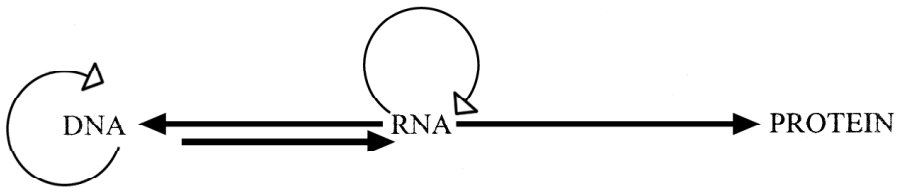


Fig. 3.1: Flow of genetic information from DNA to Protein

information never flows back from protein to RNA, in other words it can be said that acquired characters are never inherited.

The first level of gene expression involves the transfer of information stored in DNA to single stranded RNA molecule by way of a process called **transcription.** In the second step, the information scripted in the RNA molecule is translated into a linear sequence of amino acids and the process is called **translation.**

3.2 Similarities and difference between replication and transcription

Our understanding of the transcriptional process comes from the study of *E. coli.* There are several similarities between the transcription mechanism and DNA replication. Both the synthesis process utilize the similar nucleotide building blocks and : use the same chemical method of attack by a terminal -OH group of the growing chain on the triphosphate group of an incoming nucleotide. Both replication and transcription are fueled by the hydrolysis of the pyrophosphate group that is released upon attack. There are however, a number of important differences between these two distinct processes.

a) One major difference rests on the fact that while DNA replication copies an entire helix, DNA transcription only transcribes specific regions of one strand of the helix. During DNA transcription, only short stretches (about 60 base pairs) of the template DNA helix are unwound. As the RNA polymerase transcribes more of the DNA strand, this short stretch moves along with the transcription machinery. This process is different from that in DNA replication in which the parent helix remains separated until replication is done.

b) There are slight differences in the substrates that are used in DNA replication versus transcription. Transcription machinery utilize ribonucleotide instead of deoxyribonucleotide triphosphates. Additionally, in RNA the thym-ine base is replaced with the base uracil. Both of these differences can be seen in DNA transcription.

c) Another major difference is that DNA replication is a highly regulated proc- I ess that only occurs at specific times during a cell's life. DNA transcription is also regulated, but it is triggered by different signals from those used to control DNA replication.

d) One final difference lies in the capabilities of RNA polymerase versus DNA polymerase. RNA primers are needed to begin replication because DNA polymerase is unable to do it alone. DNA transcription does not require any primer synthesis as the RNA polymerase is capable of initiating RNA syn- I thesis. The structure of the RNA polymerase is necessary for understanding all of the processes that underlie initiation, elongation, and termination and I also explain some of its added capabilities.

3.3 General idea of transcription

Some 50 different protein **transcription factors** bind to **promoter** sites, usually on the 5' side of the gene to be transcribed

Transcription in both prokaryotes and eukaryotes is catalyzed by **RNA polymerase,** that synthesizes a complementary RNA molecule using one strand of the duplex DNA as template

The DNA strand that acts as template is called **template strand** and the DNA strand that is identical in sequence to the RNA strand is called **sense strand** or coding strand or nontemplate strand

5' ...ATGGCCTGGACTTCA... 3' Sense strand of DNA

3' ...TACCGGACCTGAAGT... 5' Antisense strand of DNA

↓ **Transcription of antisense strand**

5' ...HUGGCCUGGRCUUCH... 3' mRNA

↓ **Translation of mRNA**

Met— Ala— Trp— Thr — Ser — Peptide

Transcription of RNA proceeds in the 5′→3′ direction

Several types of RNA molecules are transcribed by RNA polymerases:

messenger RNA (mRNA). This will later be **translated** into apolypeptide.

ribosomal RNA (rRNA). This will be used in the building of ribosomes: machinery for synthesizing proteins by translating mRNA.

transfer RNA (tRNA). RNA molecules that carry amino acids to the growing polypeptide.

small nuclear RNA (snRNA). DNA transcription of the genes for mRNA, rRNA, and tRNA produces large precursor molecules **("primary transcripts")** that must be processed within the nucleus to produce the functional molecules for export to the cytosol. Some of these processing steps are mediated by snRNAs.

small nucleolar RNA (snoRNA). These RNAs within the nucleolus have several functions.

micro RNA (miRNA). These are tiny (-22 nts) RNA molecules that appear to regulate the expression of messenger RNA (mRNA) molecules.

XIST RNA. This inactivates one of the two X chromosomes in female vertebrates.

RNA primers formed during DNA synthesis

telomerase RNA,

ribozymes that act as enzymes

In prokaryotes, there is only a single type of RNA polymerase responsible for synthesizing all types of RNAs.

E. coli RNA polymerase is a holoenzyme comprised of subunits ââ′á(dimer) and ó^{70}.

The ó^{70} subunit is the subunit which binds to the promoter region, but is unable to initiate RNA synthesis.

After the ó^{70} subunit subunit binds, the other subunits bind forming a function RNA polymerase.

After approximately 10 base pairs have been transcribed the ó^{70} subunit leaves and the **core polymerase** continues on.

In eukaryotes, there are three major classes of RNA polymerase:

1. RNA polymerase I transcribes large rRNAs
2. RNA polymerase II transcribes mRNAs
3. RNA polymerase III tRNAs, small rRNAs and other small RNAs

In prokaryotes, the transcription and translation process are **coupled**

In eukaryotes, transcription and translation events are **compartmentalized.** RNA molecules are transcribed in the nucleus. All types of RNAs is then exported into the cytoplasm for translation

Messenger RNAs are processed in eukaryotes prior to translation. Both ends of the RNA are modified in the nucleus. The transcribed intron sequences are removed by splicing to produce a final mRNA ready for translation.

rRNAs and tRNAs are processed in both prokaryotes and eukaryotes. Most rRNAs are synthesized as a single large precursor RNA that is then cleaved into it final products. In tRNA, many individual nucleotides are chemically modified to produce the final tRNA

Transcription proceeds through the steps of **initiation, elongation** and **termination** in all cell types.

3.4 Transcription in prokaryotes

3.4.1 RNA polymerase

The structure of the RNA polymerase is necessary for understanding all of the processes that underlie initiation, elongation, and termination and also explain some of its added capabilities. In prokaryotes a single RNA polymerase transcribes all genes. There are two main segments of the RNA polymerase molecule: the core enzyme, and the sigma subunit. These two pieces are together referred to as the "holoenzyme". The *E. coli* RNA **core enzyme** is tetrameric, containing two â ? one **â'** ?and one â' type subunits. The core enzyme is sufficient for transcriptional elongation, but correct initiation requires the sigma subunit called ó^{70} factor. The ó^{70} subunit binds to the promoter region, but is unable to initiate RNA synthesis. After the ó^{70} subunit binds, the other subunits bind forming a functional RNA polymerase. Specific function of each polypeptide is given in the table 1.

Sub units	***MW***	***Location***	***Possible function***
α *chains*	*40 kd*	*core enzyme*	*Promoter binding*
β *chain*	*155 kd*	*core enzyme*	*Nucleotide binding substrate*
β′ *chain*	*160 kd*	*core enzyme*	*Template binding*
σ *factor*	*32-90 kd*	*Sigma factor*	*Initiation of transcription*

Table l: Different components of RNA polymerase that comprises the holoenzyme

The start site represents the location on the DNA that marks where the first nucleotide of an RNA chain will commence which is also designated as the "plus one position". All positions designated as upstream of the start site are labeled with negative numbers according to their position relative to the start site. Sequences located just upstream of the start site is called the **promoter** region. This region contains the information that signals the RNA polymerase to start transcription. In prokaryotic cells, free RNA polymerase molecules are constantly colliding with DNA helices. The collision leads to a weak association between the DNA and RNA polymerase, which is soon broken. However, when the RNA poJymerase binds to a specific sequence on the DNA, it binds tightly, forming a DNA/RNA polymerase complex. The **ó**? factor has two functions, it recognizes the promoter and it converts the closed promoter complex into an open promoter complex. After approximately 10 base pairs have been transcribed the ó^{70} subunit leaves and the **core polymerase** continues on. The core enzyme can bind to DNA in the absence of ó factor but with low efficiency and no specificity. The primary function of the ó factor is thus to increase the binding efficiency of RNA polymerase at the promoter and decrease (non-specific, binding. A single ó factor (ó70 in *E. coli)* initiates the transcription of

most genes, but other ó factors are present that functions only with specialized genes. For example (ó^{32}, ó^{415}, ó^{54}, ó^{28}) function under adverse conditions like high temperature, starvation, nitrogen deficiency and for chemotaxis. Some bacteriophages (e.g. T4) encode their own ó factor, which subvert the host enzyme into transcribing the phage genes, T3 and T7 phages encode their own RNA polymerases, which are single polypeptides with great affinity for the phage promoters.

3.4.2 Initiation

Initiation of transcription begins with the binding of the RNA polymerase to DNA strand. RNA polymerase binds to the DNA non-specifically and with low affinity. In the presence of the 6 factor, the holoenzyme associate with the DNA at specific sites called promoters.

Promoters: The DNA sequence to which RNA polymerase binds to initiate tran- I scription of gene is called the promoter. The DNA sequence at the promoter region is more or less conserved. Mutation or deletions at promoter region severely affects the transcriptional efficiency of that gene. The promoter is cis-acting and is always located upstream from the point of initiation of transcription. In *E. coli,* there are two promoter motifs situated at -10 and -35 sequence upstream from the start point of RNA synthesis. The consensus sequence at -10 position is **TATA AT** (Pribnow box) and at **-35** is **TTGACA** (Fig. 3.2).

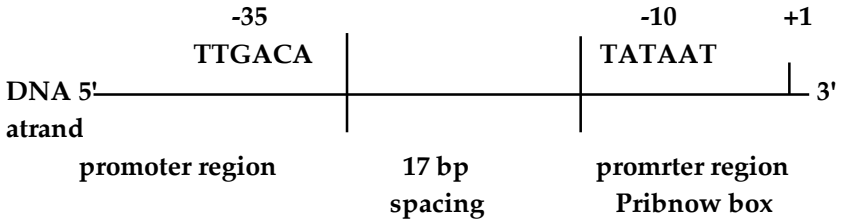


Fig. 3.2: Organization of the promoter region in *E. coli*

The -35 sequence (also called recognition sequence) is essential for binding of the RNA polymerase. At the -10 promoter region, the DNA strands' unwinds when associated with the RNA polymerase, preparing for initiation of transcription. The initial binding between the polymerase and promoter is referred to as a **closed promoter complex** because DNA is wounded. The unwinding of approximately 15 bases of DNA by the RNA polymerase around initiation site is called **open promoter complex.** Ideally, the gap between the two promoter regions is 17 base pairs long. Deviations from this spacing have significant effects on the strength of the promoter region. The closer a promoter region is to matching this canonical promoter sequence, the greater its strength. There is a third promoter element that

is sometimes seen in very strong promoters which is called the **UP element.** It usually is composed of alternating stretches of 5 adenine and thymine bases. It is located upstream of the -35 region.

Transcription usually begins with GTP or ATP and unlike subsequent nucleotides; the first nucleotide retains its triphosphate moiety. A cycle of abortive occurs generating 2-9 base short RNA sequence before actual transcription begins. Once initiation succeeds, 6 factor dissociates from RNA holoenzyme after 9-10 RNA nucleotide polymerizes and then core enzyme completes further elongation of the RNA.

3.4.3 Elongation

Elongation proceeds in the 5′ to 3′ end direction. As RNA polymerase travels down stream, it unwinds the double stranded DNA molecule ahead of it and rewinds the DNA molecule behind it, maintaining an unwounded region of about 17 bp in the region of transcription. Transcription proceeds at the rate of about 30 to 50 nucleotides per second. Energy required for polymerization of nucleotides is obtained from the cleavage of the triphosphate nucleotides (Fig. 3.3)

The RNA-DNA hybrid is very transient. The nascent RNA molecule rapidly separates from DNA and at any given time 2-12 nucleotides associate with each other. Certain proteins influence the rate of the synthesis. For example **NusA** - a protein, binds to the core enzyme that slows down elongation so that ribosome molecules are able to bind.to the nascent RNA molecule right behind the point of synthesis.

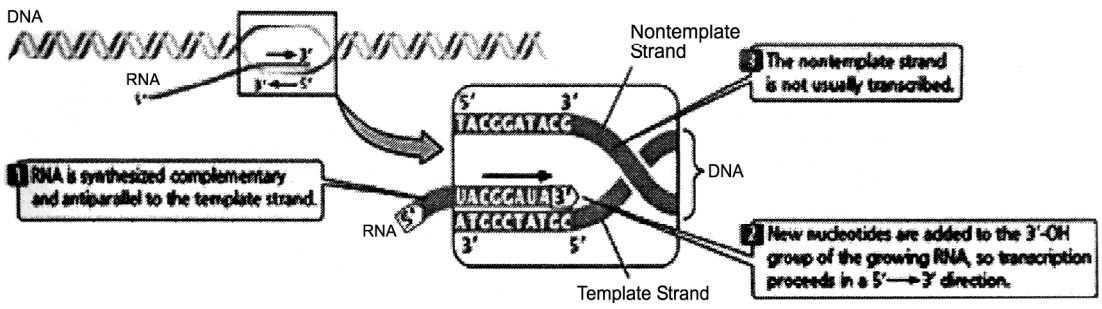


Fig. 3.3: The transcription bubble on the DNA template

3.4.4 Termination

TTermination of RNA synthesis occurs when RNA polymerase reach the end of the gene. Two mechanisms operate in bacteria and viruses: 1. Intrinsic termination or called ρ independent termination 2. ρ dependent termination.

Intrinsic termination (ρ independent termination): In this type, nucleotide sequence near the end of the transcribed RNA specifies the termination of RNA transcription. The sequences at the end of RNA are self-complementary that form a hairpin structure followed by a conserved sequence with the consensus sequence UUUUUUA. This region is called the **intrinsic terminator** (Fig. 3.4). For example:

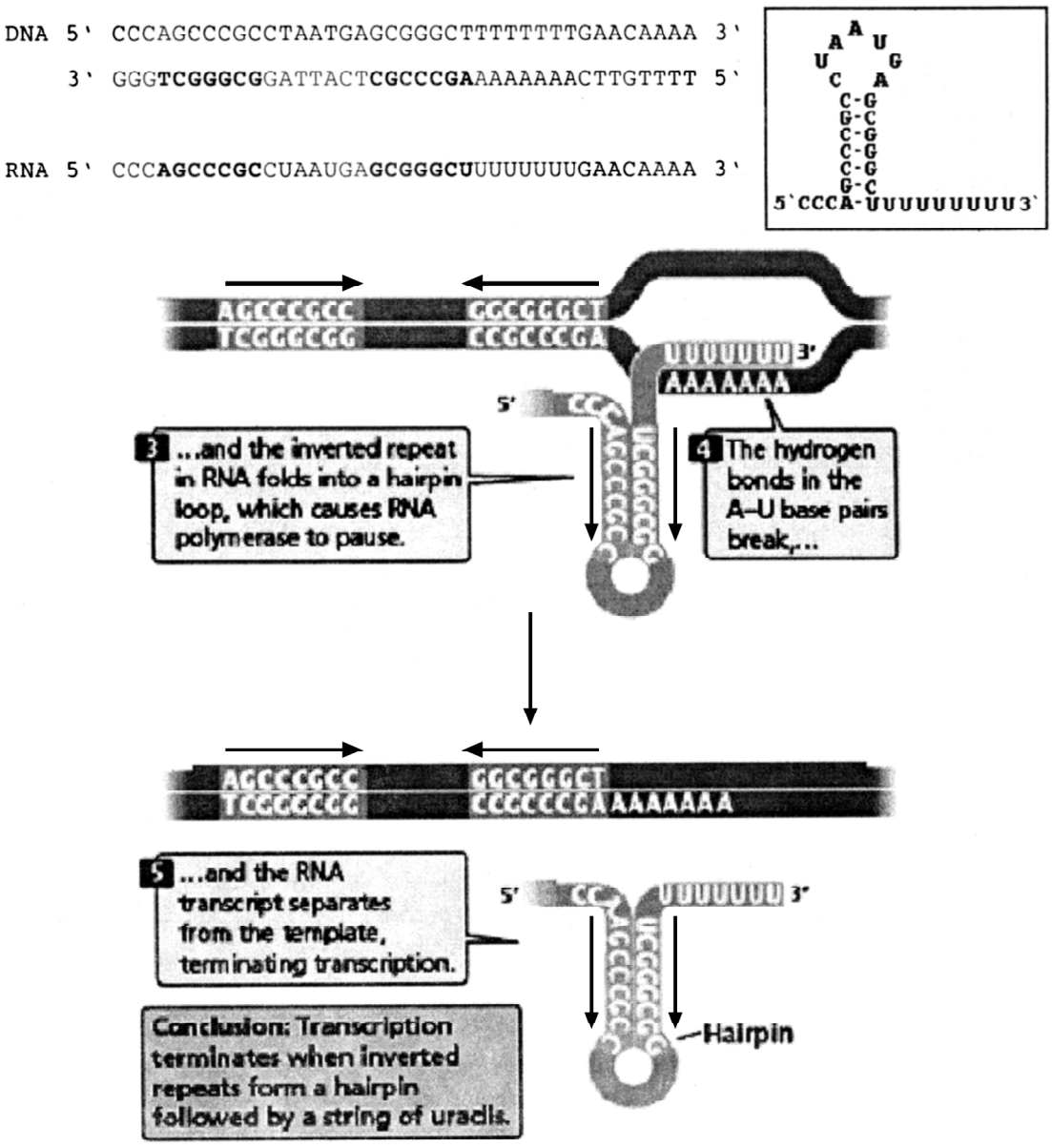


Fig. 3.4. Formation of hairpin loop on the transcribing RNA and ρ independent termination

RNA molecules having self-complementary bases at the termination point form a secondary structure called **hairpin loop** immediately after it is synthesized. The hairpin loop is followed by poly U sequence. Formation of hairpin loop and presence of poly U sequence probably halts the RNA polymerase and cause termination of RNA synthesis..The exact mechanism of action is not clear, probably the formation of secondary structure and the poly U sequence immediately before the termination signal on DNA destabilizes the RNA polymerase and releases the polymerase of the template strand. The weak bonding between the poly A DNA sequence and poly U strand on RNA may also contribute to the termination of RNA synthesis. However, hairpin loop alone not sufficient to terminate polymerase activity.

ρ dependent termination: In ρ dependent termination, the DNA sequences produce a pause in transcription towards the end of transcription and RNA does not produce any secondary structures and also lack poly U sequences. In this type, the p protein plays the principal role in termination. The ρ protein has two domains. One domain bind to RNA and the other domain bind to ATP molecule. Hydrolysis of the ATP molecule activates the ρ factor that then bind to specific site on RNA molecules in the termination region. When RNA polymerase encounters the terminator, it pauses, allowing rho to catch up. The rho protein has helicase activity, which it uses to unwind the RNA-DNA hybrid in the transcription bubble, bringing an end to transcription. Hereto, the exact mechanism of termination needs to be worked out, but most likely it destabilizes the association of the RNA-DNA-polymer association.

3.5 Transcription in eukaryotes

In eukaryotes, the genetic material remains enclosed within the nuclear membrane and is physically separated from the other organelles of the cell. The transcriptional process occurs within the nucleus. Although the transcription proceeds by the same fundamental mechanism as in prokaryotes, the regulatory mechanism is far more complex in eukaryotic cells. There is a significant difference between the transcription of eukaryotic and prokaryotic mRNAs in the initiation process. Eukaryotic promoter involves a large number of factors that bind to a variety of cw-acting elements. Eukaryotic cells possesses three different types of RNA polymerase, each specialized to transcribe different types of RNA. The promoter region on the DNA strand is defined by the nature of the RNA polymerase and transcription factors that will bind to specific sequences and support transcription at the normal efficiency and with the proper control. In fact, RNA polymerase does not make interaction with the upstream region of the promoter. The increased complexity of eukaryotic transcription presumably facilitates the

sophisticated regulation of gene expression needed to direct the activities of the many different cell- types of multicellular organisms.

3.5.1 Eukaryotic RNA polymerases and their promoters

In eukaryotic cells, there are three different types of RNA poiymerases, each located in different locations in the nucleus and is responsible for synthesizing different classes of genes as shown in the table below:

ENZYME	FUNCTION	SENSITIVITY
RNA polymerase **I** (nuckoli)	Transcription of the 45S rRNA precursor (later cleaved into 5.8S, 18S, 28S rRNA (class I genes) RNA processing	Insensitive to a-amanitin, sensitive to actinomycin D
RNA polymerase **II** (nucleoplasm)	Transcription of all protein encoding genes and most genes for small nuclear RNAs (class II genes)	Inhibited by α-amanitin
RNA polymerase **III** (nucleoplasm)	Transcription of tRNA genes, 5S rRNA genes and genes encoding U6 sn RNA and the various sn RNAs (class III genes)	Moderately sensitive to α-amanitin depending on species

Eukaryotic polymerases differ in template specificity, location in the nucleus and susceptibility to different inhibitors. Each RNA polymerase is a complex enzyme having approximately 12 to subunits and weighing about 500 kd. Five subunits of are common to all RNA polymerases. The largest subunits of each polymerase are homologous to each other and to the α, β and β′ subunits of *E. coli* RNA polymerase. There is no counterpart to the bacterial ct factor, and the eukaryotic RNA polymerases are consequentially unable to recognize or bind to their promoters. Weil et al (1979) discovered that RNA polymerases require the assistance of additional proteins not only to bind to promoter region but also to initiate transcription.

RNA polymerase I and its promoter: Apart from the basic subunits required for DNA transcription, **RNA polymerase** I requires specialized 4 Core promoter binding proteins - (It is called SL1, TIF-IB, Ribl in different species) and upstream binding factors called **UBF.** The core binding factor proteins ensures the positioning of the RNA polymerase I at the start point and can initiate transcription at a low basal frequency (Fig, 3.5). SL1 consists of four proteins. One of them, called **TBP** (TATA-binding protein), is a factor that is required also for initiation by RNA polymerases II and III. The UBF factors interact with the core proteins and greatly enhance the transcription frequency. RNA polymerase I most likely exists as a holoenzyme that contains most or all of the factors including the **TBP** required for initiation and is probably recruited directly to the promoter. The **RNA polymerase**

I promoter comprises of two separate regions. The first element surrounds the start point extending from -45 to +20, and is sufficient to initiate transcription. This element is unusually GC rich and includes a short AT rich conserved sequence called the **Inr.** However, presence of an upstream promoter element **(UPE)**, extending from -180 to -107 greatly enhances the efficiency of the primary promoter. The bipartite organization of RNA polymerase I promoter is seen in all organisms although the actual sequence may vary

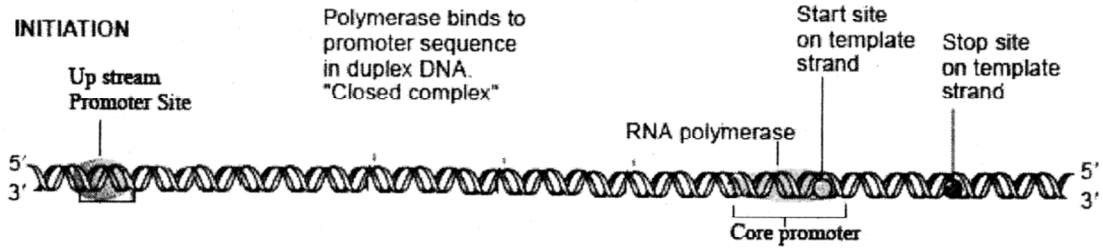


Fig 3.5: RNA polymerase I bipartite promoter

RNA polymerase II and its promoter: Several transcription factors and proteins are required by RNA polymerase II at all promoters for initiation of transcription. The subunits of RNA polymerase II and the general transcription factors are conserved among eukaryotes. Surrounding the startpoint of the core promoter region, the RNA polymerase II and the transcription factors assemble and bind with the DNA. However, the specificity and efficiency of binding to the promoter region depends on certain factors called the activators. The activators bind to target sequence ~ 100 bp upstream of startpoint or further away and influence the formation of initiation complex. Mutational studies led to the identification of three short consensus sequence centered around -30, -75 and -90 bp. Mutation of the TATA box, located at -30 does not prevent initiation but plays a crucial role in positioning the basal factors at precise location and positioning the RNA polymerase to start transcription from the right place.

The enhancing sequence elements located upstream of the start point influences the rate of transcription, probably, by interacting with the basal transcription factors. The CAAT box located at -75 or further away has a strong effect in determining the efficiency of the promoter. The consensus sequence at -90 is GGGCGG, which is very common and often exists in multiple copies.

RNA polymerase III and its promoter: RNA polymerase III uses both downstream and upstream promoters sequences. The promoters for 5S and tRNA genes lie downstream (identified in Xenopus laevis) of the startpoint between +55 and +80 bp. The promoters for snRNA (small nuclear RNA) genes lie upstream of the startpoint similar to other promoters. In both cases, the transcription factors

recognize individual promoter sequences, which in turn direct the binding of RNA poiymerase. Three types of RNA poiymerase III promoters are shown in Fig. 3.6.

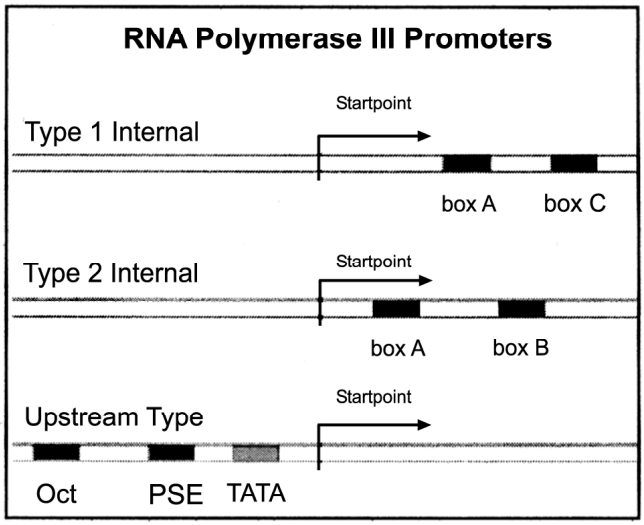


Fig 3.6 : RNA polymerase III Promoters

3.5.2 Internal promoters

The two internal type promoters have bipartite structure. Two short sequence elements remain separated by a variable sequence. The distance between boxA and boxB in a type 2 promoter vary extensively, but bringing the boxes too close inhibits function. Internal promoters bind to three different factors. TFIIIA (zinc finger proteins), TFIIIB (consists of TBP and two other proteins) and TFIIIC (large protein complex [>500 kD], has at least 5 subunits and the size is comparable to RNA poiymerase III itself). Most likely, at type 2 promoters, TFIIIC recognizes boxB, but binds to a more extensive region including both boxes A and B. At type 1 promoters, TFIIIA binds to a sequence that includes boxC, and this is required to enable TFIIIC to bind. In both cases, the binding of TFIIIC in turn enables TFIIIB to bind to a sequence surrounding the startpoint.

TFIIIA and TFIIIC removal of from the promoter by high salt concentration in vitro but allowing the presence of TFIIIB in the vicinity of the startpoint is sufficient to allow RNA poiymerase III to bind at the startpoint. So TFIIIB appears to be the only true initiation factor required by RNA poiymerase III and TFIIIA and TFIIIC are assembly factors, whose role is to assist the binding of TFIIIB at the right location. This sequence of events explains how the promoter boxes downstream can cause RNA poiymerase III to bind at the startpoint, farther upstream. TFIIIB includes the same protein, TBP that is present in SL1; this could be the subunit of TFIIIB that interacts directly with RNA poiymerase III. Any alteration in the upstream the internal promoter region can alter the efficiency of transcription.

Genes having upstream promoters has the TATA element which confer specificity for type III polymerases. Interestingly, some snRNAs are transcribed by poiymerase II while others are transcribed by poiymerase III. In both the cases,

the same type of promoters is present. Initiation commences at TATA site but the presence of PSE (Proximal Sequence Element) and Oct elements along with their factors increases the transcription efficiency. The PSE element may be essential at promoters used by RNA polymerase II, whereas it is stimulatory in promoters used by RNA polymerase IH

RNA polymerase III terminates transcription with U's immediately after a GC-rich region but there is no fonnation of stem-loop structure. Termination usually occurs at the second U within a run of four U's, but some molecules terminate with 3 or even 4 U's.

3.5.3 Eukaryotic promoters

Most eukaryotic genes have conserved DNA sequences at -25 to -30 from the start point of RNA transcription called TATA box or Hogness box that specify a particular start point during RNA transcription.. By convention, the sequence is given on the non-template strand. The TATA box has a consensus sequence 5' -TATAAAA- 3'. Mutation at this region affects the transcriptional process. Promoters that lack TATA box, there is no definite initiation point but appears to be controlled by a CT-rich area, called the initiator element (Inr) having a consensus sequence TCA, at +1 coupled with a down stream promoter element (DPE) at about +28 to +34.

Further upstream, many promoters have a CAAT box, found approximately at -75 positions of many genes but the position may vary. The consensus sequence is 5' -GGCCAATGT- 3'. Alteration of CAAT box markedly reduces transcriptional rate.

Additional sequences like GC box (5' -GGGCGG- 3') is found at -90 positions and often there are more than one copy of GC elements. This sequence may function in either orientation i.e. 5' -GGGCGG- 3' or 5'GGCGGG-3'. Interestingly, promoter of one gene may vary considerably from the other gene and no element is essential for all promoters.

3.5.4 Enhancer & Silencer

Enhancer sequences influences the transcriptional rate of a gene. Enhancer may function in either orientation and is usually located far away from the actual initiation point of transcription, sometime 1000 bp apart from the promoter sequence, usually upstream the start point. In animal cells, enhancers can be located down stream from the initiation point.

Similarly there are sequences that have the same properties like that of the enhancers but they repress rather that activate gene transcription. The elements

are called silencer elements. Silencers are less common than enhancers. There are no consensus sequences for eukaryotic enhancer or silencer element and their exact mechanism of action is still not clear. It is proposed that specific protein factors interact with the enhancer elements and folds DNA in a way so that enhancer elements interact with the transcriptional factors and regulators in the promoter region and subsequently activate (enhances) or repress (silencer) RNA transcription.

3.5.5 Initiation of transcription by RNA polymerse II

Robert Roder (1979) discovered that eukaryotic RNA polymerase II could not initiate transcription unless some other protein factors are added to the reaction mixture *in vitro.* Biochemical analysis revealed the existence of specific proteins called **Transcription** factors that are required by RNA polymerase II to start RNA synthesis. Some of these factors bind directly with the DNA while others appear to bind to the RNA polymerase. Transcription factors are usually designated by letters TF (for transcription factor) followed by Roman numeral - I or II or III to indicate the type of polymerase they bind and finally followed by a letter characterizing the type of factor, e.g. TFIIB, TFIIE etc.

Two general types of transcription factors are involved:

1. Those proteins that are required by all polymerase II and are called **Basal Transcription factors.** At least five basal transcription factors are required for initiation of transcription ob RNA polymerase II *in vitro* system.

2. Additional transcription factors that bind to DNA sequences and control the expression of specific genes and thus responsible for regulation of gene expression.

The first step in formation of a initiation complex is the binding of a complex called **TFIID.** One of the subunit of TFIID recognizes and binds to TATA box and this subunit is called **TATA-binding protein (TBP).** In essence, TFIID appears to be similar to the sigma factor in RNA polymerase (Fig. 3).

TBP is a 30 kd protein that binds to the minor groove of the DNA at the TATA promoter sequence. Binding is 10^5 times more tightly to TATA than with other sequences. TBP is saddle shaped protein with two similar domains. The TATA box binds to the concave surface of TBP. This binding induces large conformational changes. The minor

groove widens from 5Å to 9Å bat does not break the hydrogen bonds. This substantial unwinding of the minor groove enables extensive contact with the anti-parallel â strands on the concave side of the TBP. Immediately outside the TATA box, classical B-DNA resumes. This complex is distinctly asymmetric. The asymmetry is crucial for specifying a unique start site and ensuring that transcription proceeds unidirectionally.

The surface of the TBP saddle provides docking sites for the bindings of other transcription factors.

DNA bound TBP of TFIID first recruits **TFIIA** which further enhances the binding of the TBP.

Then **TFIIB** is recruited forming a complex **TBP-TFIIB** at the promoter region.

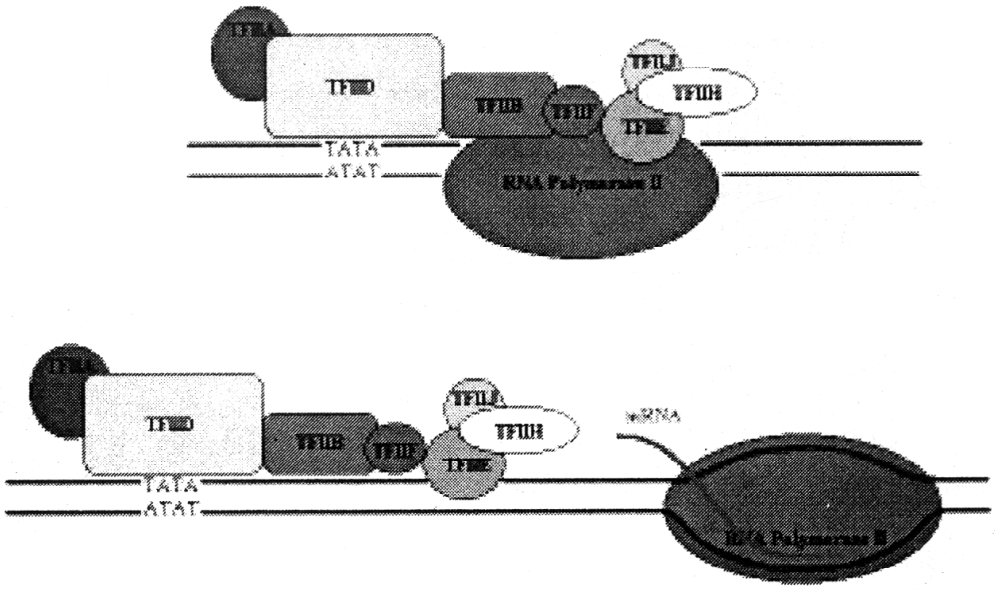
Binding of TFIIB sets the stage for the binding of RNA polymerase, which binds to the TBP-TFIIB complex in association with a third factor, **TFIIF.**

Two additional factors - **TFIIE** and **TFIIH** bind to the initiation complex and appear to be necessary for initiation of transcription. TFIIH is multi-subunit factor. First two subunits has hehcases activity, which my unwind DNA around the initiation site. Another subunit of TFIIH is a protein kinase that phosphorylates repeated sequences present in the C-terminal domain of the largest subunit of RNA polymerase II.

Phosphorylation of the C-terminal domain of the RNA polymerase II releases the enzyme from its association with the initiation complex, allowing it to proceed along the template as it elongates the growing RNA chain.

The various transcription factors described here represents the minimal system required for transcription *in vitro;* additional factors may be needed within the cell. Furthermore, RNA polymerase II appears to remain associated with some transcription factors *in vivo* prior to the assembly of a transcription complex on DNA. Such preformed holoenzyme complexes probably are recruited to a promoter via direct interaction with TFIID. What actually occurs within the eukaryotic cell

during RNA transcription still needs to be worked out. Moreover, the functions of many of the basal transcription factors are still unknown and unanswered.



3.5.6 Elongation

Placement of the first ribonucleotide with its corresponding deoxyribonucleotide in the DNA each new ribonucleotide attaches to the 3′ -OH group of the previous ribonucleotide. RNA synthesis is continuous and proceeds in the 5′ to 3′ direction. Energy is derived from the cleaving ofihe two phosphate group of the new incoming ribonucleotide that pairs with the DNA and bjnds to the 3′ -OH of the previous ribonucleotide residing within the RNApolymerase. The double stranded RNA-DNA hybrid is very transient. As RNA polymerase moves forward, the nascent RNA separates from the DNA. At any given time, the number of nucleotides of RNA that remain paired with the DNA template may be as many as 2 to 12. The unfolded DNA rapidly rewinds after RNA synthesis is over at that point and therefore nicking of the DNA is not required to release the tension of unwinding. Moreover, only 18 bases are unwounded at any time. The rate of synthesis is not always consistent as other proteins present in the cell often influence the rate of RNA synthesis.

Occasionally, RNApolymerase II stalls synthesis, may be due to some configuration change in the polymerase or my encounter a nucleotide sequence that causes the polymerase to stall on the DNA. When this happens, **TFIIS** causes the RNA polymerase to move backward, and then **TFIIS** removes the 3′ end of the RNA, permitting the RNA polymerase to attempt elongation again over the point

where the stall occurred. It may be noted that eukaryotic RNA polymerases have 3′ — 5′ exonuclease activity.

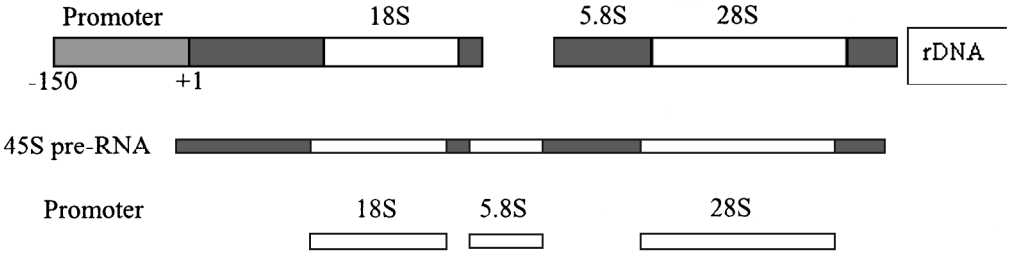
3.5.7 Termination

The termination of transcription in eukaryotic genes is less well understood than in bacterial genes. The three eukaryotic RNA polymerases use different mechanisms for termination. RNA polymerase I requires a termination factor, like the rho factor utilized in termination of some bacterial genes. Unlike rho, which binds to the newly transcribed RNA molecule, the termination factor for RNA polymerase I binds to a DNA sequence downstream of the termination site. RNA polymerase III ends transcription after transcribing a terminator sequence that produces a string of Us in the RNA molecule, like that produced by the rho-independent terminators of bacteria. Unlike rho-independent terminators in bacterial cells, RNA polymerase III does not requre that a hairpin structure precede the string of Us. In many of the genes transcribed by RNA polymerase II, transcription can end at multiple sites located within a span of hundreds or thousands of base pairs.

3.5.8 Transcription by RNA polymerse I

RNA polymerase I & III apply the same basic mechanism of transcription and only differ in the recruitment of specialized transcription factors that recognizes and associate with appropriate promoter sequences.

RNA polymerase I is solely responsible for the transcription of ribosomal RNA genes, which are present in tandem repeats. Transcription of these genes yields a large **45S** pre-RNA, which is then processed to yield the 28S, **18S,** and **5.8S** rRNAs.



The promoter sequence of ribosomal RNA gene is recognized by two transcription factors, **UBF** (upstream binding factor) and **SLl** (selectivity factor 1), which together bind to the promoter site and then recruit polymerase I to form the initiation complex. The SLl is composed of four protein subunits one of which is TBR The promoter here lacks the TATA box, and TBP therefore do not bind to

specific promoter sequences. Instead, the association of TBP with ribosomal RNA genes is mediated by the binding of other proteins in the SL1 complex to the promoter, situation similar to the association of TBP with the Inr sequences of polymerase II genes that lack TATA boxes.

The genes for tRNAs, 5S rRNA and some of the small RNAs involved in splicing and protein transport are transcribed by polymerase III. Interestingly, the promoters of these genes lie within, rather than upstream of the transcribed sequence. The well studied 5S RNA of *Xenopus* revealed that at first TFIIIA binds to the promoter followed by TFIIIC, TFIIIB and then the polymerase. In case of tRNA, the promoter sequence is recognized by TFIIIC and not TFIIIA. The multimeric TFIIIB protein appears to be the most common factor for all the polymerases to initiate transcription.

Ribosomal RNA (rRNA)

There are 4 kinds. In eukaryotes, these are

18S rRNA. One of these molecules, along with some 30 different protein molecules, is used to make the **small subunit** of the ribosome.

28S, 5.8S, and 5S rRNA. One each of these molecules, along with some 45 different proteins, are used to make the **large subunit** of the ribosome.

The S number given for each type of rRNA reflects the rate at which the molecules sediment in the ultracentrifuge. The greater the number, the larger the molecule (but not proportionally).

The 28S, 18S, and 5.8S molecules are produced by the processing of a single primary transcript from a cluster of identical copies of a single gene. The 5S molecules are produced from a different cluster of identical genes.

Transfer RNA (tRNA)

There are some 32 different kinds of tRNA in a typical eukaryotic cell.

Each is the product of a separate gene.

They are small (~4S), containing 73-93 nucleotides.

Many of the bases in the chain pair with each other forming sections of double helix.

The unpaired regions form 3 loops.

Each kind of tRNA carries (at its 3' end) one of the 20 **amino acids** (thus most amino acids have more than one tRNA responsible for them).

At one loop, 3 unpaired bases form an **anticodon.**

Base pairing between the anticodon and the complementary codon on a mRNA molecule brings the correct amino acid into the growing poiypeptide chain. Further details of this process are described in the discussion of translation.

Messenger RNA (mRNA)

Messenger RNA comes in a wide range of sizes reflecting the size of the poiypeptide it encodes. Most cells produce small amounts of thousands of different mRNA molecules, each to be translated into a peptide needed by the cell.

Many mRNAs are common to most cells, encoding "housekeeping" proteins needed by all cells (e.g. the enzymes of glycolysis). Other mRNAs are specific for only certain types of cells. These encode proteins needed for the function of that particular cell (e.g., the mRNA for hemoglobin in the precursors of red blood cells).

Small Nuclear RNA (snRNA)

Approximately a dozen different genes for snRNAs, each present in multiple copies, have been identified. The snRNAs have various roles in the processing of the other classes of RNA. For example, several snRNAs are part of the spliceosome that participates in converting pre-mRNA into mRNA by excising the intrbns and splicing the exons.

Small Nucleolar RNA (snoRNA)

As the name suggests, these RNAs (there are probably over 100 of them) are found in the nucleolus where they are responsible for several functions:

Some participate in making ribosomes by helping to cut up the large RNA precursor of the 28S, 18S, and 5.8S molecules.

Others chemically modify many of the nucleotides in these molecules, e.g., by adding methyl groups to ribose.

Still others serve as the template for the synthesize of telomeres.

In vertebrates, the snoRNAs are made from **introns** removed during RNA processing.

3.6 Transcription factors

Transcription factor is, a protein that works in concert with other proteins to either promote or'suppress the transcription of genes. More specifically, transcription factors regulate gene expression. They bind to specific sequences of DNA upstream or downstream to the gene they regulate and then either enhance or repress transcription of these genes by assisting or blocking RNA polymerase binding respectively.

A defining characteristic of transcription factors is that they contain a **DNA binding domain** (DBD) which bind to gene specific regulatory sites *{e.g.,* promoter sequences). In addition, transcription factors often contain a second domain that sense external signals and in response transmit these signals to the rest of the transcription complex resulting in up or down regulation of gene expression. In some cases the DBD and signal sensing domains reside on separate proteins that associate within the transcription complex to regulate gene expression.

Other proteins such as coactivators, chromatin remodelers, histone acetylases, deacetylases, kinases, and methylases, also playing crucial roles in gene regulation but they lack DNA binding domains, and therefore are not classified as transcription factors.

3.6.1 Regulation

Gene regulation is a highly complex process as it is dependent upon a number of factors. *In vitro* experiments suggested that the assembly of transcription factors dictated by the DNA sequence. However, epigenetic information present on DNA appears to play an important role in transcriptional activation.

3.6.2 Classes of transcription factors

There are three classes of transcription factors: a) Upstream transcription factors are proteins that bind somewhere upstream of the initiation site to stimulate or repress transcription, b) Inducible transcription factors are similar to upstream transcription factors but require activation or inhibition, c) General transcription factors

General transcription factors are involved in the formation of a preinitiation complex that participate in the transcription of class II genes to mRNA templates.Tata binding protein, (TBP) is a general transcription factor, that binds to the TATAAbox, the motif that resides upstream from the coding region in all genes. TBP is responsible for the recruitment of the RNA Pol II holoenzyme, the final event in transcription initiation. The most common general transcription factors are TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH. They are ubiquitous and interact with tire core promoter region surrounding the transcvription start site(s) of all class II genes.

TFIIA

TFIIA consists of two subunits in yeast and three in humans and drosophila (two subunits are derived from a precursor protein). TFIIA binds directly to TBP and stabilizes its binding to DNA, perhaps through direct contact with the DNA. TFIIA binding does not preclude TFIIB binding or other components of the transcription complex. However, binding of TFIIA to TBP is mutually exclusive with binding of some negative regulatory proteins. TFIIA acts as an anti-repressor, stabilizing TFIID binding by blocking repressors of transcription that inhibit binding of other transcription factors or that remove TBP from the DNA. Activation of transcription may be dependent on this TFIIA function.

TFIIB :

It is **a** single subunit measuring 35 to 40 kd. **The protein possesses a** zinc finger domain at the N-terminus and a direct repeat in a proteolytically stable C-terminal domain. TFIIB binds directly to TBP, recruits RNA polymerase II, in part through an interaction with the small subunit of TFIIF. Several acidic activators can bind TFIIB in vitro. **The protein probably** stabilizes TBP binding to TATA element and is required for association of RNA polymerase II to the initiation complex. TATA-Binding Protein is shown in green, TFIIB in red (Fig. 3.7).

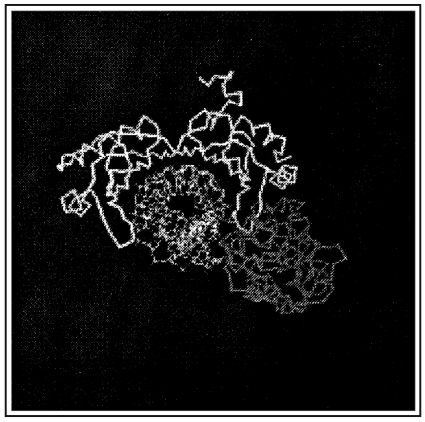


Fig. 3.7: 3D structure of TFIIB

TTFIID is a multi-component (>5 subunits) transcription factor that recognizes and binds to the promoter DNA. TFIID consists of a DNA binding subunit that recognizes the TATA element and is therefore designated TATA-binding protein (or TBP), as well as several TBP-associated factors (or TAFs). TFIID helps in recruiting the rest of the factors through a direct interaction with TFIIB. The TBP subunit of TFIID is sufficient for TATA element binding and TFIIB interaction, and can support basal transcription. However, this basal transcription reaction does not respond to upstream transcription activators. Many of these regulatory factors interact with TBP or TAFs in various in vitro assays. TBP also interacts directly with TFIIA.

Features: TBP consists of a 180 ammo acid domain that is sufficient for activity. This domain is made up of an imperfectly repeated sequence, and the repeats are reflected in the symmetry of the molecule (see picture below). The protein resembles a saddle, with the inner surface contacting DNA and the outer surface presumable making protein-protein contacts. TFIID binding is thought to be the first step in transcription initiation. Some of the TAFs also bind to initiator elements. TBP is also a component of the RNA polymerase I and RNA polymerase III transcription complexes.

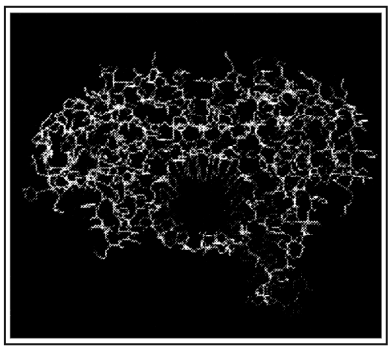
TFIIE: It has two subunits, probably a tetramer consisting of two molecules of each subunit. The large subunit has a zinc finger domain. TFIIE modulates the helicase and kinase activities of TFIIH and the two factors show species-specific interactions. It recruits TFIIH to the initiation complex and modulates TFIIH kinase and helicase activities. Appears to be required for escape of the RNA polymerase into elongation mode (promoter clearance).

TFIIF: The molecule has two subunits. TFIIF binds directly to RNA polymerase II. TFIIF is necessary for RNA polymerase II to stably associate with the TFIIF-TFIIB-promoter complex. There is a protein interaction between the small subunit and TFIIB in vitro and a genetic interaction between the large subunit and TFIIB. It Helps recruit RNA polymerase II to the initiation complex in collaboration with TFIIB. TFIIF is a component of the yeast holoenzyme and mediator complexes. Promotes transcription elongation, may remain associated with the elongating polymerase.

TFIIH: Mammalian and yeast TFIIHs have at least six subunits. Most subunits are now cloned, although not all are published.Yeast subunits: SSL2(RAD25), RAD3, SSL1, TFB1, TFB2, TFB3, TFB4. In addition to the TFIIH subunits, there is an associated complex consisting of a CDC-like kinase and cyclin-like subunit. This kinase complex is sometimes referred to as TFIIK. Yeast subunits: KIN28 and CCL1. The two largest TFIIH subunits are ATP-dependent helicases of opposite polarity. Two of the smaller subunits have possible zinc finger domains. TFIIH appears to be dependent upon TFIIE for incorporation into the initiation complex. The associated kinase (TFIIK) complex can phosphorylate the C-terminal domain of the pol II largest subunit. TFIIH is essential for promoter melting (separation of the two DNA strands) and/or promoter clearance (i.e. for pol II to break free of the initiation complex into elongation mode). Surprisingly, TFIIH also is essential for Nucleotide Excision Repair (NER) of damaged DNA. The relationship between TFIIH's transcription and repair functions is not understood yet.

SPT16 **:** Subunit of the heterodimeric FACT complex (Sptl6p-Pob3p), facilitates RNA Polymerase II transcription elongation through nucleosomes by destabilizing and then reassembling nucleosome structure of the DNA. *Sptl6p* has been found to physically interact with *PobSp* (which has sequence similarity to some HMG chro-matin-associated proteins) and the catalytic subunit of DNA polymerase alpha, *PolJp.* Some of the *Sptl6p/Pob3p* complex in the cell is chromatin associated, and some copurifies with the DNA polymerase alpha-primase complex. The N-terminal third of *Sptl6p* is necessary for the maintenance of chromatin repression, but not for activation of genes. Homologs of *SPT16* have been found in *K. lactis* and human

TBP (TATA binding protein))is a DNA binding protein that binds sequence specifically to the TATA box. It is vital for all eukaryote transcription, and will in some cases be forced to bind non sequence specifically. It is involved in DNA melting (double strand separation) and bends the DNA by 80° (the AT-rich sequence to which it binds facilitates easy melting). The TBP is an unusual protein in that it binds the minor groove using a â sheet. TBP is â subunit of the eukaryotic transcription factor TFIID. TFIID is the first protein to bind to DNA during the formation of the pre-initiation transcription complex of RNA polymerase II (RNA Pol II). Binding of TFIID to the TATA box in the promoter region of the gene initiates the recruitment of other factors required for RNA Pol II to begin transcription. Each of these transcription factors are formed from the interaction of many protein subunits, indicating that transcription is a heavily regulated process. TBP is also a necessary component of RNA polymerase I and RNA polymerase III, and is perhaps the only common subunit required by all three of the RNA polymerases.



Yeast TBP bound to DNA.

When TBP binds to a TATA box within the DNA, it distorts the DNA by creating a nearly 90 degree bend. The distortion is accomplished through a great amount of surface contact between the protein and DNA. TBP binds with the negatively charged phosphates in the DNA backbone through positively charged lysine and arginine amino acid residues. The sharp bend in the DNA is produced through projection of four bulky phenylalanine residues into the minor groove. As the DNA bends, its contact with TBP increases, thus enhancing the DNA-protein interaction.

The strain imposed on the DNA through this interaction initiates melting, or separation, of the strands. Because this region of DNA is rich in adenine and thymine residues, which base pair through only two hydrogen bonds, the DNA strands are more easily separated. Separation of the two strands exposes the bases and allows RNA polymerase II to begin transcription of the gene.

FACT complex (Facilitates chromatin transcription complex): An abundant nuclear complex, which was originally identified in mammalian systems as a factor required for transcription elongation on chromatin templates. The FACT complex has been shown to destablilize the interaction between the H2A/H2B dimer and the H3/ H4 tetramer of the nucleosome, thus reorganizing the structure of the nucleosome. In this way, the FACT complex may play a role in DNA replication and other processes that traverse the chromatin, as well as in transcription elongation. FACT is composed of two proteins that are evolutionarily conserved in all eukaryotes and homologous to mammalian Sptl6 and SSRP1. In metazoans, the SSRP1 homolog contains an HMG domain; however in fungi and protists, it does not. For example, in S. cerevisiae the Pob3 protein is homologous to SSRP1, but lacks the HMG chromatin binding domain. Instead, the yFACT complex of Sptl6p and Pob3p, binds to nucleosomes where multiple copies of the HMG-domain con^{+}aining protein Nhp6p have already bound, but Nhp6p does not form a stable complex with the Sptl6p/Pob3p heterodimer.

3.6.3 Structural binding motifs

DNA-Binding motifs

In general, transcription activators have two domains: one domain binds specifically to DNA and the second domain interact with other transcription factors. Four different types of transcription activators have been identified and their DNA binding motif characterized.

1. **Zinc finger domains** : have repeats of cystein and histidine residues, which interact with Zn++ ions to fold in a fingerlike fashion to grasp the DNA. E.g. TFIIIA, Steroid hormone receptors also have Zinc finger domains, which regulate gene transcription in response to hormones like estrogen & testosterone.

2. **Helix-turn helix :** In this type of activators, one helix makes most of the contact with DNA while the other helix lies across the complex to stabilize the interaction. E.g. Homeodomain protein, catabolic activator protein (CAP) in *Drosophila*
   Note: Homeodomain proteins play critical roles in the regulation of gene expression during embryonic development.

1. **Leucine Zipper:** contain DNA binding domains formed by dimerization of two polypeptide chains. The leucine zipper contains four or five leucine residues spaced at intervals of seven residues resulting in the exposure of hydrophobic side chains at one side of the helical region. This hydrophobic region of the domains serves as the point of dimerization of the two domains of the transcription activator, thereby interlocking the DNA strand within it by the interaction of positively charged lysine and arginine with the DNA.

2. **Helix loop helix:** In the type, the amino acid sequence is similar to leucine zipper domains, except that their dimerization domains are formed by two other domains separated by a loop.
   An interesting feature is that both leucine zipper and helix loop helix transcription factors is that different members of these families can dimerize with each other. Such dimerization gives rise to formation of an array of transcription activators that differ in DNA specificity and also binding to other transcription factors. These two transcription activators play an important role in regulation tissue specific gene expression.
   The activation domains are not well characterized as their DNA binding domains. Some of these domains are acidic, some basic. But they somehow interact with basal transcription factors like TFIID, TFIIB and facilitate the initiation of transcription of specific genes.

3. **MADS box** is a conserved sequence element found in a family of transcription factor encoding genes. The length of the MADS-box is defined differently by various authors, but typical lengths suggested are 168 base pairs or 180 base pairs. The element encodes the MADS-domain that have DNA-binding properties. In plants, MADS-box genes have undergone a substantial radiation. In Arabidopsis the MADS box genes SOC and FLC have been shown to have an important role in the integration of molecular flowering time pathways. These genes are essential for the correct timing of flowering, and help to ensure that fertilization occurs at the time of maximal reproductive potential.

3.6.4 Protein-binding motifs

STAT

The Signal Transducers and Activator of Transcription (STAT) proteins regulate many aspects of cell growth, survival and differentiation. The transcription,factors of this family are activated by the Janus Kinase JAK and dysregulation of this pathway is frequently observed in primary tumors and leads

to increased angiogenesis and enhanced survival of tumors. Knockout studies have provided evidence that STAT proteins are involved in the development and function of the immune system and play a role in maintaining immune tolerance and tumor surveillance.

Function of STAT proteins

STAT proteins were originally described as latent cytoplasmic transcription factors that require phosphorylation for nuclear retention. The unphosphorylated STAT proteins shuttle between the cytosol and the nucleus waiting for its activation signal. Once the activated transcription factors reach the nucleus, they bind to a consensus DNA-recognition motif called gamma activated sites (GAS) in the promoter region of cytokine-inducible genes and activate transcription of these genes.

Activation of STAT proteins

Extracellular binding of Cytokines induces activation of the intracellular Janus kinase that phosphorylates a specific tyrosine residue in the STAT protein which promotes the dimerization of STAT monomers via their SH2 domain. The phosphorylated dimer is then actively transported in the nucleus via importin a/b and RanGDP complex. Once inside the nucleus the active STAT dimer binds to cytokine inducible promoter regions of genes containing gamma activated site (GAS) motif and activate transcription of this proteins. The STAT protein can be dephosphorylated by nuclear phosphatases which leads to inactivation of STAT and the transcription factor becomes transported out of the nucleus by exportin crml/RanGTP.

Unit 4 Post Transcriptional Modification of RNA

Structure

4.1 Introduction

In prokaryotes, tRNA and rRNA undergo modification after being transcribed but mRNA do not get the opportunity to undergo modification as the transcription and translation processes are coupled. In Eukaryotes, all the three types of RNAs (mRNA, tRNA and rRNA) undergo post transcriptional modifications.

4.2 Post transcription modification of r RNA

Prokaryotes have three ribosomal rRNA (23S, 16S and 5S) equivalent to the eukaryotic 28S, 18S and 5S rRNAs of eukaryotic cell. In both the cell types, the processing of a single pre-RNA transcript produces different rRNAs. In eukaryotes, only the 5S RNA does not undergo much modification as they are synthesized from a separate gene. The steps of cleavage to remove the introns and obtaining the final product are shown in the figure 4.1.

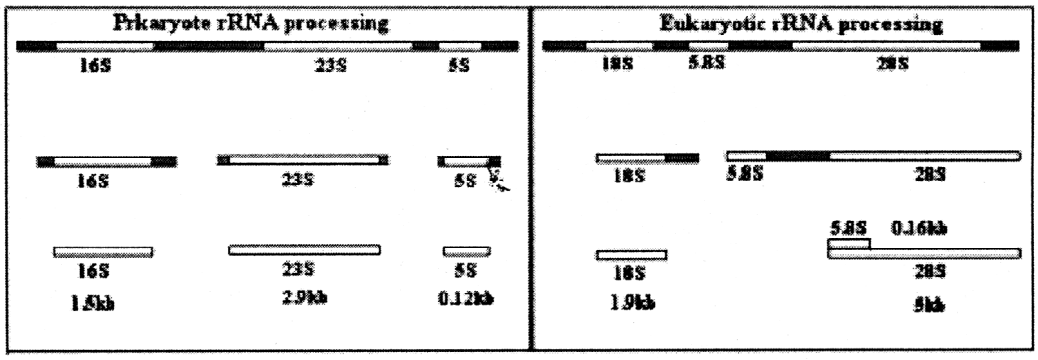


Fig. 4.1: Processing of newly synthesized rRNA

In eukaryotes, after the 5.8S RNA is produced, it is hydrogen bounded to 28S RNA. Further, rRNA processing involves the addition of methyl groups and sugar moieties to specific nucleotides, but the function of these modifications is unknown.

4.3 Post transcriptionaJ modification of tRNA

Prokaryotes and eukaryotes synthesize tRNA as precursor molecules. Some pre t-RNA transcripts have several tRNA sequences, which are cleaved and modified to obtain mature functional tRNA. Some tRNA sequences are present within the pre-RNA transcripts.

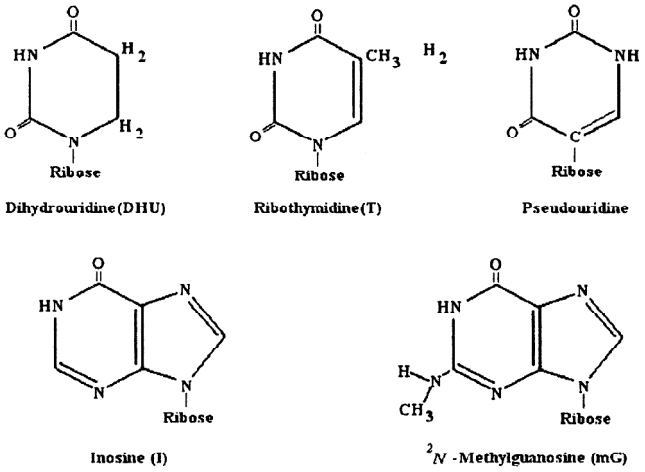
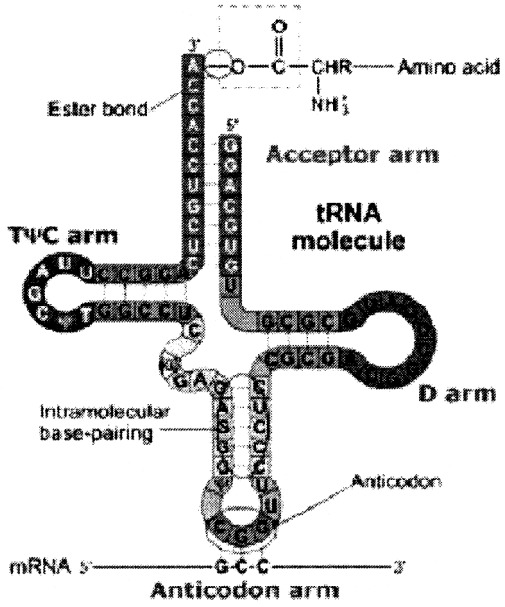


Fig. 4.2a : Modified bases



The 5′ end of all tRNAs is modified by RNase P (ribozyme). Conventional RNase modify the 3′ of end of tRNA which is the by followed by addition of -CCA nucleotides. Some tRNA has the information of-CCA already encoded in the DNA. All tRNAs have -CCA sequence at their 3' end. Moreover, 10% of the bases in tRNAs are altered to yield a variety of modified nucleotides (Fig. 4.2a & 4.2b) at specific positions in tRNA molecules but their exact function is still not clear.

4.4 Transcriptional modification of mRNA

mRNA in all organisms can be distinguished into three primary regions (Fig. 4.3).

a) The 5′ untranslated region (5′ UTR) also called the leader sequence do not code for any amino acids but carry vital information for subsequent mRNA modifications and translation. In bacterial mRNA, this region contains a consensus sequence called the Shine-Dalgarno sequence, which serves as the ribosomebinding site during translation. It is found approximately seven nucleotides upstream of the first codon translated into an amino acid (called the start codon). Eukaryotic mRNA has no equivalent consensus sequence in its 5′ untranslated region, ribosomes bind to a modified 5′ end of mRNA.

b) The next section of mRNA is the **protein-coding region,** which comprises the codons that specify the amino acid sequence of the protein. The protein-coding region begins with a start codon and ends with a stop codon.

c) The last region of mRNA is the **3′untranslated region** (3′ UTR), a sequence of nucleotides at the 3′ end of mRNA that is not translated into protein. The 3′ untranslated region affects the stability of mRNA and the translation of the mRNA protein-coding sequence.

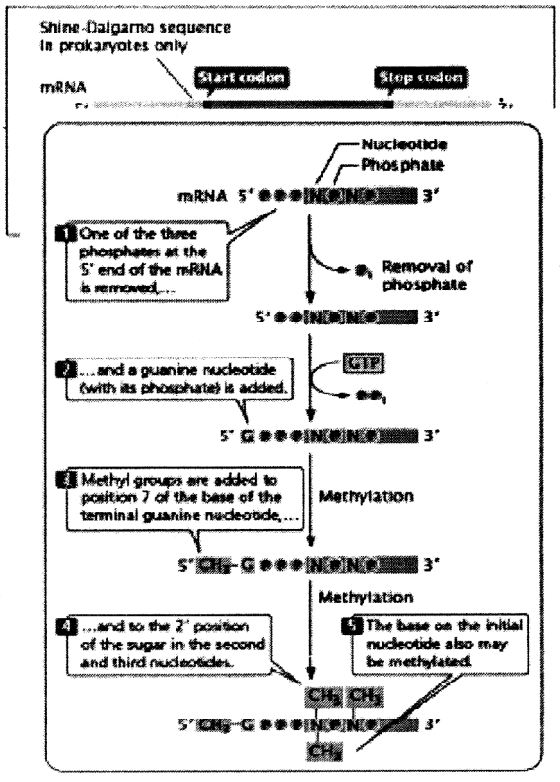


Fig. 4.3: Structure of mRNA molecule

Processing of mRNA

In eukaryotes, transcription and translation are separated in both time and space and this separation provides an opportunity for eukaryotic RNA to be modified before it is translated: The initial transcript of protein-encoding genes of eukaryotic cells is called pre-mRNA Eukaryotic mRNA undergoes extensive alteration after transcription. Changes are made at both the 5' end, the 3' end, and also in the protein-coding section of the RNA molecule

The addition of the 5′Cap

The 5′ end of all eukaryotic pre-mRNAs are modified by the addition of a extra nucleotide, followed by methylation (addition of CH_3 group) to the 2′-OH group of the sugar of one or more nucleotides at the 5′ end - a process termed as capping and the structure is called 5′ **cap** (Fig. 4.4).

Capping occurs a few moments after the commencement of transcription. Transcription starts with a nucleoside triphosphate which is usually a purine (A or G). The capping process involves the addition of GTP molecule in the reverse direction to the 5′ terminal residue. Addition of the 5′ terminal G is catalyzed by a nuclear enzyme, guanylyl transferase. Subsequently, methyl groups are added to the N7of GTP molecule.

The next step is to add another methyl group, to the 2′ -O position of the penultimate base (which was actually the original first base of the transcript before any modifications were made). This reaction is catalyzed by another enzyme (2′ -O-methyl-transferase). A cap that possesses this single methyl group is known as a cap 0 and is found in unicellular animals.

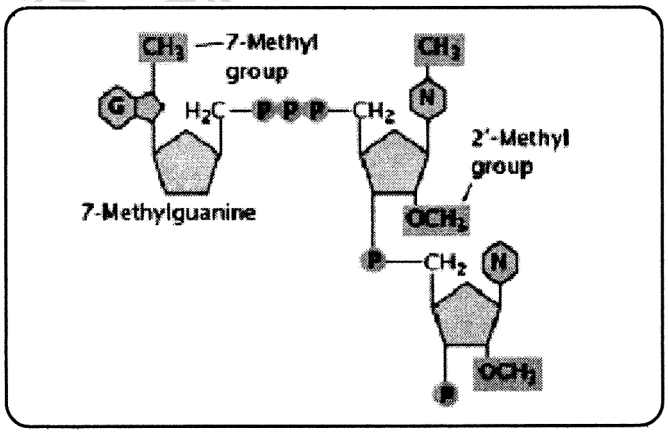


Fig 4.4 : Structure of 5' cap

A cap with the two methyi groups is called cap 1. This is the predominant type of cap in all Fig. 4,4: Structure of 5' cap eukaryotes except unicellular organisms. Methylation may also occur at second and third positions and is called cap 2. If the second nucleotide is adenine, then the methylation occurs at the N^6 position since it already has a methyl group at 2′ position. The third-base modification is always a 2′-O ribose methylation. This cap usually represents less than 10-15% of the total capped population.

In addition to the methylation involved in capping, a low frequency of internal methylation occurs in the mRNA only of higher eukaryotes. This is accomplished by the generation of N^6 methyladenine residues at a frequency of about one modification per 1000 bases. There are 1-2 methyladenines in a typical higher eukaryotic mRNA, although their presence is not obligatory, since some mRNAs do not have any.

The cap blocks the 5′ end of mRNA. The 5' cap helps to align mRNAs on the ribosome during translation initiation. Cap binding proteins recognize the cap and attach to it; a ribosome then binds to these proteins and moves downstream along the mRNA until the start codon is reached and translation begins. The presence of a 5′ cap also increases the stability of mRNA and influences the removal of introns.

4.5 The addition of the Poly(A) tail on mRNA

Most mature eukaryotic mRNAs have from 50 to 250 adenine nucleotides at the 3′ end [a **poiy (A) tail].** These nucleotides are not encoded in the DNA but are added after transcription **(Fig.** 4.5) in a process termed polyadenylation. Polymerase II encoded genes transcribes well beyond the end of the coding sequence (more than 1000) at the 3′ end which is then cleaved and the poly(A) tail is added.

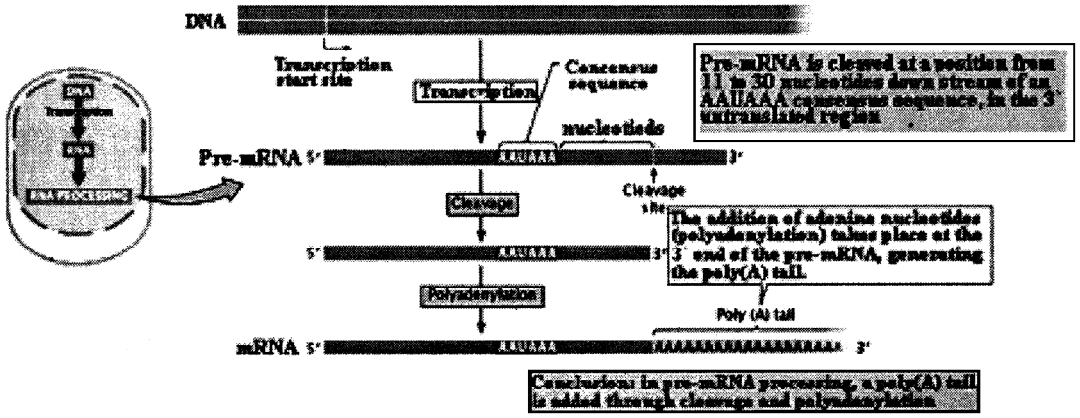


Fig. 4.5: Addition of poly(A) tail at 3′ end of pre-mRNA

The site of cleavage at the 3′ end of pre-mRNA is determined by specific upstream and downstream sequences **(Fig.** 6). Upstream consensus is usually AAUAAA that determines the site of the cleavage and resides 11 to 30 nucleotides

upstream of the cleavage site. A sequence rich in Us (or Gs and Us) is typically present down-stream of the cleavage site. In mammals, 3′cleavage and the addition of the poly(A) tail requires a complex consisting of several proteins:

a) Cleavage and polyadenylation specificity factor (CPSF);

b) Cleavage stimulation factor (CstF);

c) At least two cleavage factors (CFI and CFII) ;

d) Polyadenylate polymerase (PAP).

CPSF binds to the upstream AAUAAA consensus sequence, whereas CstF binds to the downstream sequence (Fig. 4.6). CstF after cleaving the pre-mRNA leave the complex. The cleaved 3′ end of the pre-mRNA is then degraded. CFSF and PAP remain bound to the pre-mRNA and carry out polyadenylation. After the addition of approximately 10 adenine nucleotides, a poly(A)-bindirig protein (PABH) attaches to the poly(A) tail and increases the rate of polyadenylation. As more of the tail is synthesized, additional molecules of PABII attach to it.

The poly(A) tail confers stability to many mRNAs, increasing its half life, making it available for longer time for the translational process, before it is degraded by cellular enzymes. The stability conferred by the poly(A) tail is dependent on the proteins that attached to the tail.

Eukaryotic mRNAs that lack a poly (A) tail depend on a different mechanism for 3′ cleavage. It requires the formation of a hairpin structure with the aid of a small ribonucleoprotein particle (snRNP)

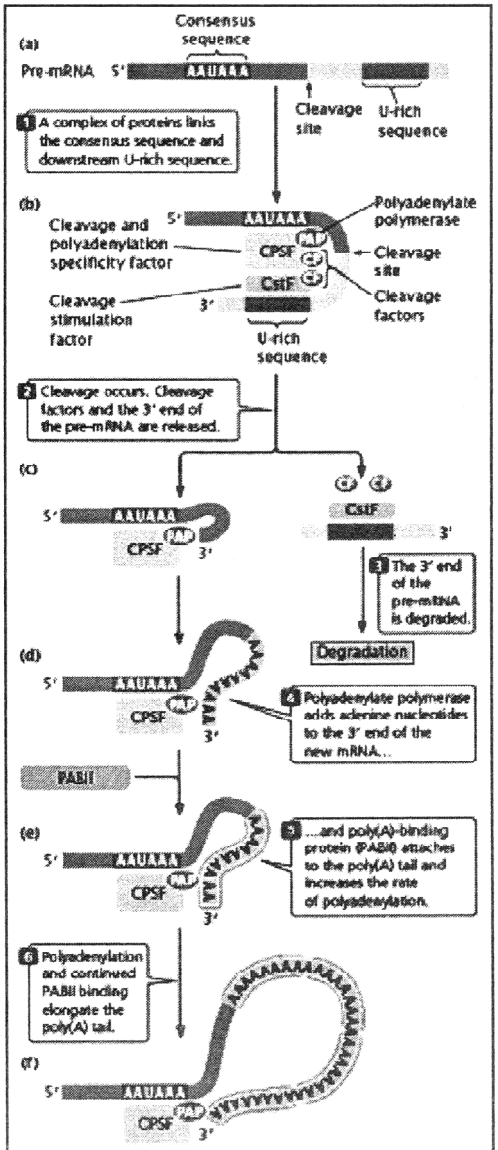


Fig. 4.6 : Process of binding of CPSF and Cstf

called U7 (Fig. 4.7). U7 contains an snRNA with nucleotides that are complementary to a sequence on the pre-rnRNA just downstream of the cleavage site. U7 most likely binds to this complementary sequence. A hairpin-binding protein binds to the hairpin structure and stabilizes the binding of U7 to the complementary sequence on the pre-mRNA and cleave the 3' non coding sequences

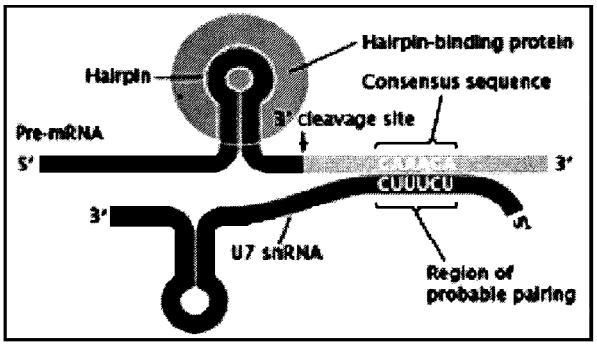


Fig. 4.7 : Ribonucleoprotein particle

4.6 RNA Splicing

4.6.1 Intron Removal

Most eukaryotic genes are interrupted by non-coding sequences called introns. The intron consists of GU at 5′ end and AG at 3′ end, while a branch site (A) in the middle and a (py)n, meaning a stretch of primidine near the 3′ end. During mRNA processing, the introns are precisely excised from the mature mRNA. At first, a protein complex called spliceosomes cleaves the 5′ end of the intron. In the second step, the 5′ end of the intron is joined to an adenine residue within the intron at the 2′-OH group of the adenine nucleotide to form a 2′, 5′ -phosphodiester linkage, which is quite unusual bond. The resulting intermediate is a lariat like structure. Next the 3 end of the intron is spliced followed by Hgation of two exon units (Fig. 4.8).

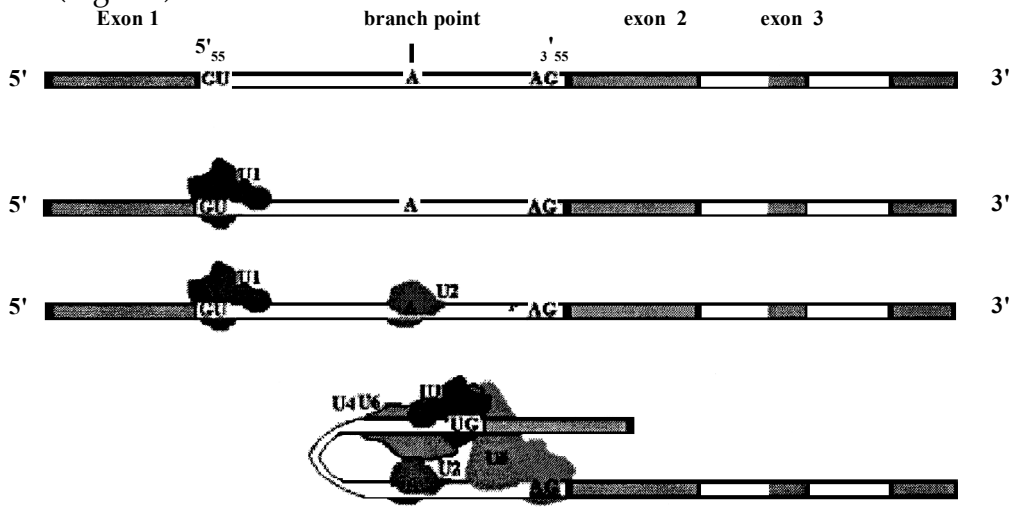


Fig. 4.8: Schematic diagram of splicing mechanism

The entire process is defined by three critical elements in the mRNA:

a. Sequences at the 5′ splice site of the intron (GU)
b. Sequence at the 3′ splice site of the intron (AG)
c. Sequences within the intron at the branch point where the 5' end joins.

4.6.2 Splicesosome

Spliesome are protein and snRNA complexes.

The RNAs are snRNAs called U_1 U_3 U_4 U_5 U_6 Their size varies from 50-200 nucleotides.

U_1 U_2 and U. along with specific proteins exist as independent units.

U_4 and U_6h along with their proteins are grouped together as a single unit.

U, snRNP first recognizes the 5′ splice site sequence and bind to it by complementary base pairing

U_2 then binds at the branch site with 'A' within the intron

A preformed, complex consisting of U_4U_6 and U_5 snRNPs is then incorporated into the forming splisosome

The U_5 component binds with both the 5′ and 3′ splice sites.

The snRNPs then catalyses the reaction by first cleaving the 5 splice site, then joining the 5′ terminal nucleotide with a specific adenine residue within the intron followed by cleavage at 3′ splice site and joining of the exons.

The snRNAs in the snRNPs actually catalyzes the reaction.

Note: RNAs are capable of self splicing- i.e. remove their own introns. Eg. 28S rRNA in *Tetrahymena.* Self-splicing RNAs are also present in mitochondria, chloroplast and bacteria. On the basis of the catalytic activity of self-splicing RNAs, they have been grouped into two classes.

Class I: In this type, rRNA first cleaves itself at the 5′ end of the intron. The 3′ end of the exon then catalyses the reaction at the 3′ end of the intron followed by joining of the two exons.

Class II: In this type, self-splicing rRNAs exhibit characteristics of the reaction as observed with mRNA described above.

Alternative splicing: Most pre-mRNAs have multiple introns and exones which can be arranged in alternative ways by splicing of the same mRNA can produce different mRNA- a novel means of controlling gene expression. This process is known as **alternative splicing** and occurs frequently in genes of eukaryotes that provide an important mechanism for tissue-specific and developmental regulation of gene expression. The regulation and selection of splice sites is done by Serine/Argi nine-residue proteins, or **SR proteins** (Fig. 4.9). The use of alternative splicing factors leads to a modification of the definition of a "gene".

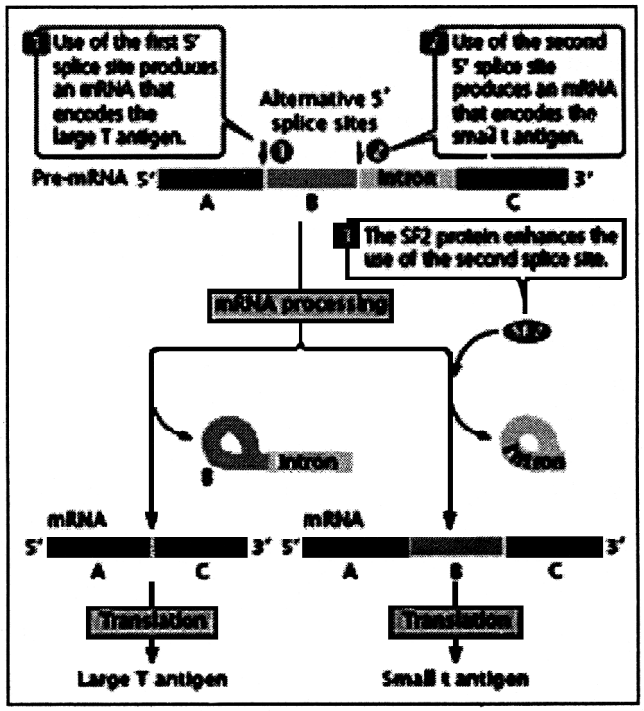


Fig. 4.9: SR Proteins

There are four known modes of alternative splicing:

Alternative selection of promoters: this is the only method of splicing which can produce an alternative N-terminus domain in proteins. In this case, different sets of promoters can be spliced with certain sets of other exons.

Alternative selection of cleavage/polyadenylation sites: this is the only method of splicing which can produce an alternative C-terminus domain in proteins. In this case, different sets of polyadenylation sites can be spliced with the other exons.

Intron retaining mode: in this case, instead of splicing out an intron, the intron is retained in the mRNA transcript. However, the intron must be properly encoding for amino acids. The intron's code must be properly expressible, otherwise a stop codon or a shift in the reading frame will cause the protein to be non-functional.

Exon cassette mode: in this case, certain exons are spliced out to alter the sequence of amino acids in the expressed protein.

E.g. 1. Transcriptional activators consist of two distinct domains: a DNA binding domain and an activation domain. These domains are generally encoded

in separate exons, so alternative splicing allows them to be reassorted into different combinations, thereby enabling the production of activators and repressors from the same gene.

2. In *Drosophila,* alternative splicing of the same pre-mRNA determines whether a fly will be a male or female.

Patterns of alternative splicing can vary in different tissues. Several protein factors have been isolated but the mechanism by which the correct splice sites are selected in pre-mRNA is not known. Variations in the expression of such splicing factors in different cell types may result in tissue specific patterns of alternative splicing, there by contributing to the regulation of gene expression during development and differentiation.

4.6.3 Significance of alternative splicing

Alternative splicing is of great importance to genetics - it invalidates the old theory of one DNA sequence coding for one polypeptide. External information provide the clue of alternative splicing. Since the methods of regulation are inherited, the interpretation of a mutation may be changed.

Alternative splicing allows more information to be stored much more economically in a limited space. It has been noted that it is unnecessary to change the DNA of a gene for the evolution of a new protein. Instead, a new way of regulation could lead to the same effect, but leaving the code for the established proteins unharmed. Another speculation is that new proteins could be allowed to evolve much faster than in prokaryotes. This machanism may allow for a higher probability for a functional new protein. Therefore the adaptation to new environments can be much faster - with fewer generations - than in prokaryotes. This might have been one very important step for multicellular organisms with a longer life cycle.

Trans-splicing :

In this type, exons from two different pre-mRNAs are joined to form a single mRNA. For example, in trypanosomes, all mRNAs have an identical spliced leader sequence of 35 nucleotides. This leader sequence is present the 5' end of a 137 nucleotide RNA chain. This leader sequence is then spliced and added to all the 5' end of mRNAs by trans-splicing reactions. Trans-splicing machinery also exists in mammals as mammalian cells are capable of carrying out trans-splicing reactions with the nematode spliced leader RNAs (Fig. 4.10).

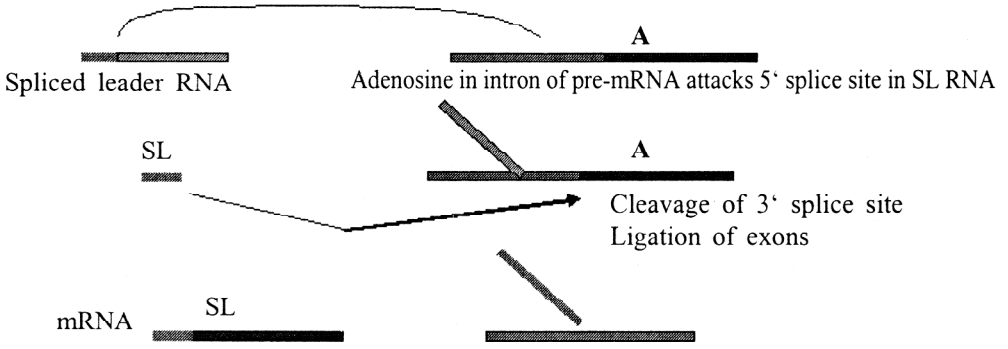


Fig. 4.10: Trans-splicing reactions with the nematode spliced leader RNAs.

RNA Editing

RNA editing refers to RNA processing events (other than splicing events) that alter the protein coding sequences of some mRNAs. In trypanosomes and related protozoans, addition of'U' and deletions of some nucleotides occur in some mRNAs. The information required for editing is encoded in "guide" RNAs, which are complementary to edited portions of the mature mRNA. The guide RNA contains poly-U tail, which donate the 'Us' during editing. Sometimes the editing is so extensive that half the nucleotide sequences are altered.

In mammalian cells similar events occur in mitochondria and also in the somatic cells. For example, in human body two types of apolipoproteins are found, Aop-BlOO (4536 aa: synthesized in liver is unedited form and transports lipids in the circulation) & Apo-B48 (2152 aa: synthesized in the intestine, is edited form and helps in the absorption of dietary lipids in the intestine). In the intestine, CAA codon is changed to a stop codon by enzymatic conversion of C to U by removal of the cytosine amino group at specific site in the mRNA.

4.7 Nuclear export of mRNA

4.7.1 Introduction

Compartmentahzation of the eukaryotic genome by the nuclear membrane was probably a necessity to have a greater control over the functioning of the genome and also to avoid unnecessary alterations in the genomic constituents by exposing it to the bustling biochemical activity that occur in the cytoplasm. This compartmentahzation further ensures the presence of specialized environments for different stages of gene expression, such as transcription and protein production.

Trafficking of materials between the nucleus and cytoplasm primarily rely

on transport receptors in the **importin-**β superfamily. However, export of mRNA uti- I lizes distinct soluble machinery. In yeast, it has been observed that protein-protein interaction is required for the export of mRNA from the nucleus to the cytoplasm. For example, in yeast Mex67p interact with Mtr2p and facilitates the export of I po!y(A)$^+$ RNA. In metazoans and in humans, TAP was confirmed to be the orthologue of Mex67p, redesignated as **NXFl** (nuclear export factor 1) which interacts with pl5/ NXT1 in the nucleus for transportation of mRNA. Mtr2p and pl5 share no sequence similarity but the Mex67p-Mtr2p complex displays similar structural architecture to I the NXFl-p 15 heterodimer. However, the distinction between mRNA export and the importin-β family/Ran network is not absolute, as an importin-13 family member has recently been implicated in mRNA export as well.

4.7.2 mRNA export is coupled to splicing

NXFl does not bind directly to cellular mRNA. Experiments with *Xenopus* oocytes demonstrated that the process of splicing can contribute to the efficiency of mRNA export; the spliced product from adenovirus major late (Adml) mRNA was shown to export more efficiently than an identical mRNA engineered to lack an intron. On the mRNAs, exon junction complexes (EJC) are formed after splicing. EJC complex include several components like REF (nuclear export factor), SRml60, RNPS1, DEK, Y14, and later its protein partner Magoh. Recruitment of a unique set of proteins to the spliced mRNA may promote export competency of mRNA.

The notion that EJC deposition leads to recruitment of NXFl is an attractive model to explain the stimulatory effect of splicing on export. Direct interactions between REF and NXFl have been observed in both human and yeast systems. REF also shuttles between nucleus and cytoplasm and enhances mRNA export when injected into *Xenopus* oocyte nuclei as a recombinant protein. The enhanced placement of REF onto mRNA in a splicing-dependent fashion, as well as its association with NXFl, made REF a prime candidate for recruiting NXF1 onto mRNA.

A connection between splicing and mRNA export was further solidified with the characterization of a novel role for the putative RNA helicase UAP56 [56-kDa U2AF(65)-associated protein]. Recruitment of REF to spliced mRNA is dependent upon its interaction with UAP56. From these data, a very simple yet elegant mode of coupling splicing with mRNA export became evident. Namely, REF is recruited to splice mRNA through direct interactions with UAP56, and consequently, REF (and the EJC in general) recruits the export factor NXFl to promote exit from the nucleus by mediating docking and presumably movement through the pore.

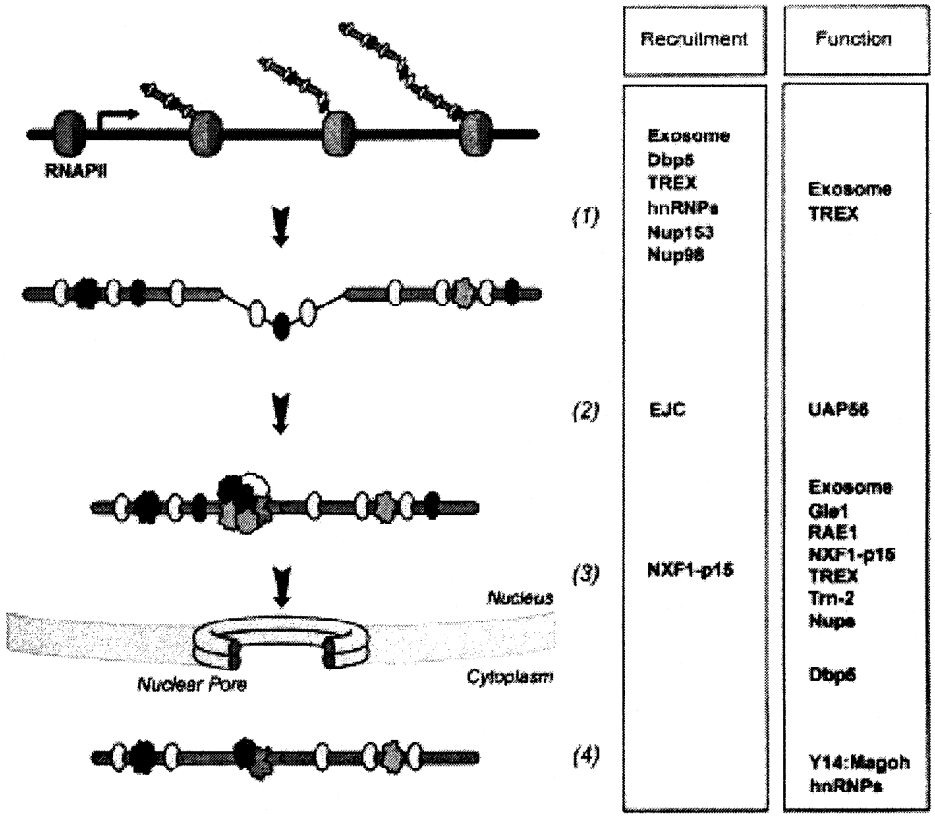


Fig. 4.11: A schematic illustration of mRNA biogenesis is depicted, with the proposed times of recruitment and functions for specific proteins indicated in the boxes.

(1) Transcription: Much evidence points to cotranscriptional loading of factors involved in RNA processing, export, and quality control to the nascent transcript. The mobile pore proteins, Nupl53 and Nup98, are candidates (indicated in red text) for loading onto mRNA early in its biogenesis, although this is yet to be demonstrated.

(2) Splicing. The splicing factor UAP56 interacts with REF, a component of EJC that is deposited on mRNA during splicing. Loading of certain transport factors, such as REF, can also occur independent of splicing as a part of TREX or if the RNA is sufficiently long.

(3) Remodeling and export. NXFl-pl5 is recruited to the mRNA via protein-protein interactions, readying export-competent mRNA for mobilization out of the nucleus. At this step, other proteins, such as Glel, RAE1, Trn-2 (transportin-2),

and TREX components, are also thought to function. Certain hnRNPs and EJC components are shed from the mRNP prior to export, and proper mRNP formation appears to be monitored at this stage by the exosome. Specific pore proteins, or Nups, are implicated in moving mRNA cargo through the pore (Fig. 4.12). Although loaded onto the transcript early in biogenesis, Dbp5 may play a late role in remodeling and/ or moving the mRNP complex. (4) Cytoplasmic function. Factors remaining on the mRNA, such as Y14 and Magoh, influence translation and localization of mature mRNPs.

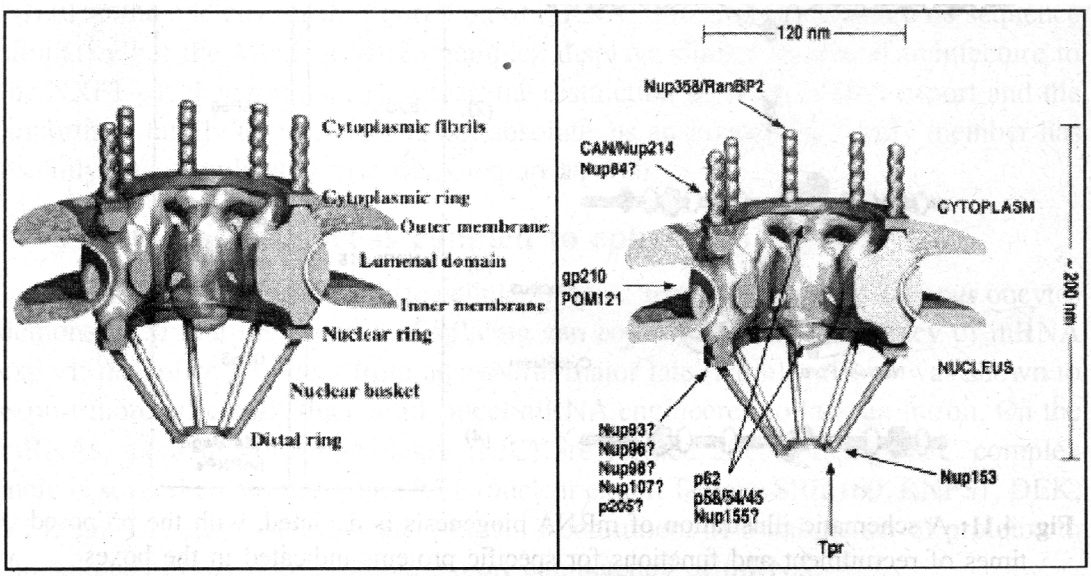


Fig. 4.12. Distribution and dynamics of pore proteins and associated factors implicated in mRNA export. Nupl53 and Nup98 are both dynamically associated with the nuclear pore in a manner dependent on ongoing transcription. However, in the case of Nupl 53, there is also evidence for a stable population, which is schematically illustrated here proximal to the inner nuclear membrane. The presence of distinct populations of"Nupl 53 is consistent with epitope exposure of this protein: different regions are exposed at the distal and proximal ends of the pore basket. Regardless of exactly how Nupl53 is arranged at the pore basket, there is evidence that the C-terminal region of this pore protein can extend into the cytoplasm, although Nupl53 does not appear to be released from this face of the pore. In contrast, Nup98 exists in equilibrium with a cytoplasmic pool and is known to interact with components of both sides of the pore. RAE1/Gle2 is a partner protein of Nup98. Both Nupl53 and Nup98 associate with the Nupl07-160 complex, a stable component of the pore. Tpr is also a component of the nuclear pore basket and relies on interaction with Nupl53 for correct localization. CAN/Nup214 is localized to fibrils extending from the cytoplasmic ring of the pore and is a docking site for the dynamic DEAD box helicase, Dbp5.

Despite the observation that splicing promotes export of mRNA, such processing is not the only or even the major route for export factor recruitment. Recent functional analysis of mammalian cells also suggested that splicing does not always have a major effect on mRNA export per se. Splicing and export factor recruitment that has been documented may represent only one way that export factors load onto mRNA and indeed may make a significant contribution at this step only when the RNA is particularly short

4.7.3 Coordinating mRNA export with transcription and turnover

The addition of a 5' cap, splicing, polyadenylation, and cleavage events occur in close connection to transcription. Concurrent with the processing events, mRNAs are also packaged with a number of proteins specific to this class of RNA (Fig. 1). A significant subset of such proteins was originally classified as hnRNPsfor their ability to associate with heterogeneous nuclear RNA. hnRNP Al a component of hnRNPs was initially implicated in mRNA transport, but it's exact role in transport mechanism remains to be elucidated.

Other proteins, not necessarily classified originally in the hnRNP category but loaded onto mRNA, have been functionally connected to both export and transcription. Yralp and Sub2p display both genetic and physical interactions with all members of the yeast THO complex, a protein complex identified originally for a role in transcription elongation. REF and UAP56 along with the vertebrate counterparts of THO and a new protein of unknown function, Texl, make up the TREX (transcription and export) complex. In yeast, specific TREX components associate with genes during transcription and, individually, their deletion results in nuclear poly(A)$^+$ accumulation. Together, this suggests that TREX proteins may be important in mediating cotranscriptional recruitment of factors important in export. For example, one protein in the TREX complex, Hprlp, is required for efficient targeting of Yralp and Sub2p to genes undergoing active transcription. Therefore, cotranscriptional recruitment and splicing-dependent recruitment represent two broad mechanisms by which mRNA export factors can associate with RNA cargo. For instance, REF maybe efficiently loaded onto specific RNA cargos via cotranscriptional targeting of the TREX complex and/or through splicing-dependent deposition of the EJC.

Another example of the connection between transcription and export is found in the DEAD box helicase Dbp5. Dbp5 is localized at steady state to the cytoplasmic fibrils of the nuclear pore complex (NPC) and has been hypothesized to be involved in a terminal step of mRNA release from the NPC, possibly acting in a remodeling

step to unwind mRNPs entering the cytoplasm. Interaction of Dbp5p with TFIIH during mRNA transcription and its shuttling between nucleoplasm and cytoplasm in S. cerevisiae is an indication that Dbp5 may load onto mRNA cargo very early in biogenesis that enables transport and remodeling of mRNA. Overall, much evidence is arising to support a link between mRNA synthesis and the effective recruitment of export factors to the nascent transcript.

4.7.4 Moving on to the nuclear pore

Nuclear protein complexes (NPC) span the nuclear envelope and serve as gateways of communication individual nuclear pore proteins or nucleoporins (Nups) present several times, creating octagonal symmetry. The pore also has asymmetric features on its nuclear and cytoplasmic faces. Although much of the process of mRNA export is being deciphered, there is still little known about how mRNPs interface with pore machinery. Some recent studies have focused on the roles of proteins that are closely associated with the pore, such as Glel and RAE1/ Gle2. Glel is essential for mRNA export in both yeast and human cells, and hGlel is a dynamic factor that shuttles between nucleus and cytoplasm. The shuttling domain of hGlel acts as a dominant-negative export inhibitor of both buIkpoly(A)$^+$ RNA and specific mRNA transcripts. Docking of hGlel at the NPC was recently shown to depend on an interaction with the pore protein Nupl55.

Murine RAE1 is essential; however, cells from mice bearing targeted disruption of RAE1 do not have a detectable defect on bulk mRNA export. In contrast, RAE1 deletion in yeast results in nuclear accumulation of po!y(A)$^+$ RNA. Although there appear to be redundant factors in vertebrates, hRAEl interacts with NXF1 and the nucleoporin Nup98, as well as with mRNA itself, and has been speculated to be involved in delivering mRNA cargo-receptor complexes to Nup98. Nup98, in turn, has been implicated through antibody inhibition studies in the export of mRNA as well as other classes of RNA. Nup98 shares similarity with yeast nuclear pore proteins Nupl45, Nupll6, and NuplOO. Deletion of yNup 145 causes the nuclear accumulation of poly(A)$^+$ RNA.

Vertebrate pore proteins have not been exhaustively screened and individually tested for roles in mRNA export. However, along with Nup98, five other vertebrate pore proteins, Nupl53, Nupl60, Nupl33, Tpr, and CAN/Nup214, have so far been implicated in the export of mRNA (Fig. 4.13). Mouse embryos deficient in CAN/ Nup214not only show arrest in the G_2 phase of the cell cycle but also demonstrate nuclear accumulation of poly(A)$^+$ RNA. Nupl59/Rat7,the

yeast orthologue of CAN/ Nup214, is similarly implicated in mRNA export, with a temperature-sensitive mutation causing very rapid onset of accumulation of poly(A)$^+$ RNA in the nucleus. CAN/Nup214 associates with the mRNA export factor Dbp5, an interaction conserved from yeast to vertebrates. In addition, CAN/ Nup214 is the only vertebrate nucleoporin with a steady-state localization exclusively on the cytoplasmic side of the pore that has been implicated in mRNA export thus far. Although much remains to be elucidated,Nup98 and Nupl53 are prime candidates for coupling the production of mRNA to its transport into the cytoplasm.

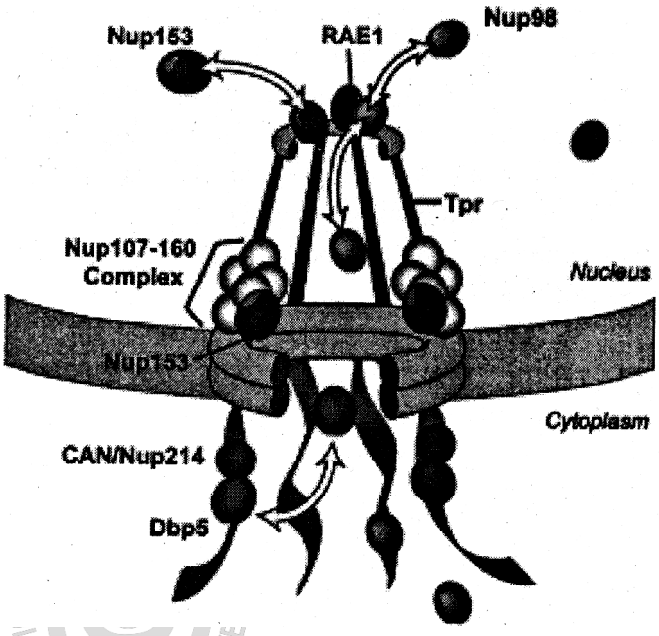


Fig. 4.13 : Different Pore Proteins

4.7.5 Transport through to pore : putting individual components into context

In recent years, several models to explain the mechanism of movement through the pore have been proposed. In each, the phenyalanine-glycine repeat motif (FG repeat) regions found in several nucleoporins play a prominent role, both in contributing to an exclusion barrier as well as in serving as binding sites for cargo-receptor complexes. Consistent with this, NXF1 directly interacts with the FG repeat domains of several nucleoporins in vitro. It is presumed that during translocation of mRNAs, associated large heterogenous mRNPs undergo remodeling in conjunction with transport. Elegant immunoelectron microscopy studies have gone on to illustrate that certain proteins are shed from the mRNP, while others accompany the RNA through the pore. Complicating things further, the nuclearpore basket itself has been observed to adopt different conformations when the Balbiani ring mRNA is traversing the pore. The dynamic nature of specific pore basket components themselves (Nup98 and Nupl53),as well as the sensitivity of such mobility to the transcriptional status of the cell, suggests that basket

remodeling may normally be ongoing in a manner linked to RNA trafficking. Recent information suggests that the native NPCs distal ring of the pore basket is not an open hole but rather a dense structure and thus has little scope of remodeling. An alternative point of entry into the pore, in between the fibers of the pore basket, has also been proposed. Much work is still needed to understand how mRNA enters and translocates through the pore. Future approaches that provide high-resolution real-time imaging as well as more precise functional assays are sure to yield a very interesting story about how the complicated network of mRNA biogenesis connects with translocation through the nuclear pore and the downstream fate of the mRNA.

4.7.6 Exportins and importins

The traffic through the nuclear envelope is mediated by a protein family which can be divided into **exportins** and **importins.** Binding of a molecule (a "cargo") to exportins facilitates its export to the cytoplasm. Importins facilitate import into the nucleus.

The function of exportins and importins is regulated by a G protein called **"Ran".** There are two types of G proteins: heterotrimeric G proteins and monomeric G proteins (or small G proteins). The latter includes Ras, Ran, Rho, Rab, etc. Like other G proteins, Ran can switch between GTP-bound and GDP-bound states. Transition from the GTP-bound to the GDP-bound state is catalyzed by a **GTPase-activating protein (GAP)** which induces hydrolysis of the bound GTR The reverse transition is catalyzed by **guanine nucleotide exchange factor (GEF)** which induces exchange between the bound GDP and the cellular GTP.

The GEF of Ran (denoted by RanGEF) is located predominantly in the nucleus while RanGAP is located almost exclusively in the cytoplasm. Therefore, **in the nucleus Ran will be mainly in the GTP-bound state** due to the action of RanGEF while **cytoplasmic Ran will be mainly loaded with GDP.** This asymmetric distribution has led to the following model for the function of exportins and importins.

It is thought that binding between an exportin or importin and its cargo depends on their interaction with Ran: **RanGTP enhances binding between an exportin and its cargo but stimulates release of importing cargo; RanGDT has the opposite effect,** namely, it stimulates the release of exportin's cargo, but enhances the binding between an importin and its cargo. Therefore, the exportin and its cargo may move together with RanGTP inside the nucleus, but the cargo will be released as soon as the complex moves into the cytoplasm (through nuclear pores), since RanGTP will be converted to RanGDP in the cytoplasm. By contrast,

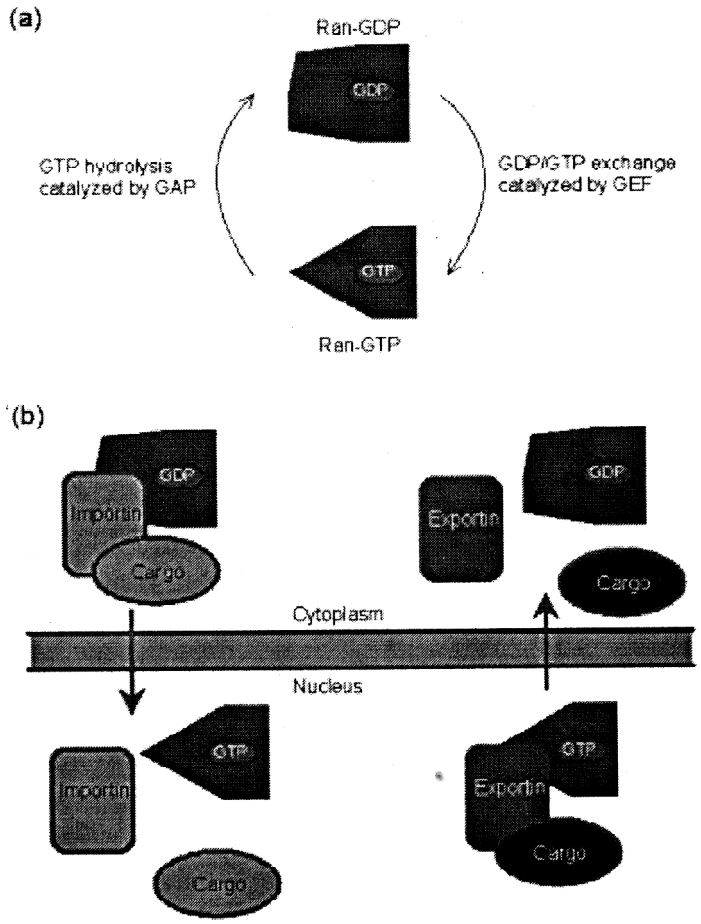


Fig. 4.14: Ran, importin and exportin. **(a)** The two states of Ran: GTP-bound and GDP-bound. **(b)** General function of importins and exportins

the importin and its cargo may move together with RanGDP in the cytoplasm, but the cargo will be released in the nucleus since RanGDP will be converted to RanGTP in the nucleus.

Unit 5 Translation

Structure

5.1 Genetic code

5.1.1 Introduction

A code is a system of symbols that equates information in one language with information in another e.g. Morse code. In living organisms, the hereditary information is written in the language of four nucleic acids, A, G, C, and TAJ. and the language of proteins are written in amino acids. As there are 20 amino acids that are genetically encoded by DNA or RNA sequences, the first question that comes in mind is that how many nucleotides are necessary to code for one amino acid. *Now* ***we know that genetic codes are triplet codons, each of which represents a single amino acid.***

Initially, the number of nucleotides necessary to code for one amino acid was derived by reasoning and later confirmed by experimentation. Scientists reasoned if one nucleotide represented one amino acid, only four nucleotides could be coded. If two nucleotides represented each amino acid, there would be 4^2 =16 possible combination of couplets, still not sufficient to code for 20 amino acids. If the code consisted of groups containing one and two nucleotides, it would have 4 + 16-20, just sufficient for all the 20 amino acids but would fail to recognize the pause signal between two genes. Groups of three nucleotides in a row would provide 4^3 = 64 different triplet combinations, more than enough to code for **all** the amino acids. Theoretically, the above logic appeared to be simple but had to await experimental evidence that proved beyond doubt that groups of three nucleotides indeed are necessary to code for a single amino acid. Each triplet nucleotides in genetics are referred to as **codon.**

5.1.2 Experimental evidences

Crick and his coworkers tested acridine induced mutations in the B cistron of rll locus of T4 phage (acridine mutation is produced by addition or deletion of a

nucle-otide). The mutations acted on the principle of addition or deletion of a single nucleotide pair in DNA molecule. Arbitrarily the mutations were designed as + or -on the basis of their suppression effect on other mutations.

For eg. :

In DNA : TAC TCC CGA ACG ATA CCA GAG

In RNA : AUG AGG GCU UGC UAU GGU CUC

Protein : 

Mutations induced by acridine treatment (arbitrarily designated as 4- mutations)

Point mutation

↓

In DNA : TAC GCC CGA ACG ATA CCA GAG (point mutation; a base is replced)

In RNA : AUG CGG GCU UGC UAU GGU CUC [NO FRAME SHIFT MUTATION]

Protein : 

Second Mutations that suppress the first imitation (designated as - mutations)

Insertion Deletion

↓ ↓ ↓

In DNA : TAC GCC CCGA CGA TAC CAG AG

In RNA : AUG CGG GGU GCU AUG GUC UC

Protein :  FRAME SHIFT MUTATION

Three + Mutations can restore the oriainal mutation

Insertion **[Partial restoration of original]**

In DNA : TAC GCC CCA TCG ATA CCA GAG CCA [three positive mutation restores the

In RNA : AUG CGG CGU ACU UAU GGU CUC GGU original FRAME SHIFT mutation]

Protein : 

Observations:

1. A single + or - mutation is sufficient to alter the wild type trait into mutant
2. The effect of a single + mutation is suppressed either by a - mutation or by 2 further + mutations. Similarly - mutatios can by suppressed by either by a + mutation or by 2 further - mutations.

Explanations & Deductions:

1. Due to deletion or insertion of a nucleotide, codon constitution of reading frame gets altered after the point of such change. For this, the polypeptide product will be different resulting in a mutant trait.
2. If it is insertion mutation, two further insertion mutations can restore the original trait. Similarly, if the mutation is caused by deletion, two further deletion mutations can restore the original trait.

So, three changes of the similar kind or multiples of 3 are necessary to restore to wild type. This can only be possible if the codons are triplet i.e. consists of 3 nucleotide.

5.1.3 Nucleotide sequence is collinear with a polypeptide's amino acid sequence

The nucleotides in RNA or DNA are arranged in a linear order. As DNA codes for proteins, it was assumed that the amino acids in the polypeptide chain must also be arranged in a linear fashion. Although the proteins have a highly complex three dimensional structure under normal circumstances, analysis has revealed that the amino acids in any polypeptide chain are arranged one after another, have definite polarities and show no branching. Thus the linear arrangements of nucleotides and the amino acids led to suggested that there must be one to one corresponding relationship of nucleotides and amino acids during protein synthesis.

5.1.4 Overlapping vs. non-overlapping nature of codons

Logical derivation of 3 nucleotides constituting a codon however could not provide clue as to how the codon are arranged i.e. either overlapping or non-overlapping. Point mutations with mutagens like nitrous acids or provalin were used to decipher the arrangement of codons.

a) If Overlapping, a sequence of 9 nucleotides will code for 7 amino acids as shown below. Alteration of a single nucleotide will produce changes in minimum three codons.

b) If non overlapping, 9 nucleotide will code for only three amino acids and alteration of a single nucleotide will alter only one codon.

Original sequence	***Mutant sequence***
CAGAGCUCA	CAGAACUCA
CAG.....codon 1	CAG.....codon 1
AGA.....codon 2	AGA.....codon 2
GAG.....codon 3	GAG.....codon 3
AGC.....codon 4	AGC.....codon 4
GCU.....codon 5	GCU.....codon 5
CUC.....codon 6	CUC.....codon 6
UCA...codon 7	UCA...codon 7

b) If non overlapping, 9 nucleotide will code for only three amino acids and alteration of a single nucleotide will alter only one codon.

Original sequence			***Mutant sequence***		
CAGAGCUCA			CAGAACUCA		
CAG	AGC	UCA	CAG	AAC	UCA
codon 1	codon 2	codon 3	codon 1	codon 2	codon 3

Experiments revealed that a single-nucleotide substitution mutation caused an amino acid substitution of leucine for proline. The two adjacent amino acids were unchanged by the mutation providing evidence against overlapping codons (Tsugita and Fraenkel, 1960).

	U	C	A	G	
U	UUU, UUC: Phe; UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr; UAA, UAG: Stop	UGU, UGC: Cys; UGA: Stop; UGG: Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His; CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G
A	AUU, AUC, AUA: Ile; AUG: Met	ACU, ACC, ACA, ACG: Thr	AAU, AAC: Asn; AAA, AAG: Lys	AGU, AGC: Ser; AGA, AGG: Arg	U C A G
G	GUU, GUC, GUA, GUG: Val	GCU, GCC, GCA, GCG: Ala	GAU, GAC: Asp; GAA, GAG: Glu	GGU, GGC, GGA, GGG: Gly	U C A G

5.1.5 Properties of genetic code

1. Each Genetic Code is a **Triplet Codon** each of which specifies an amino acid.
2. The code is **non-overlapping.** E.g. in mRNA 9 nucleotide sequence 5 GAAGUUGAA3 , will be translated to 3 amino acid sequence corresponding to GAA, GUU and GAA and no more.
3. The **code is degenerate,** which means that in many cases more than one codon specifies the same amino acids.
4. The **code is coma less.** A coma less code means that no punctuations are needed between any two words i.e. after one amino acid is coded; the second amino acid will be automatically coded by the next three nucleotides.
5. The code includes **three stop, or nonsense codons:** UAA, UAG, and UGA. These three codons do not code for any amino acids, rather they terminate translation.
6. The code is **non-ambigious.** Non-ambiguous code means that there is no ambiguity about a particular codon. A particular codon will always code for the same amino acid wherever it is found. *Only exception lies with AUG and GUG at the start point where both code for methionine although GUG actually code for valine.*
7. The **code is universal.** Almost all micro and macro organisms use the same genetic code with a few exceptions. For example, a different code exists in mitochondria of some eukaryotes, so that in cytoplasm and mitochondria same codon may code for different amino acids.
8. During translation, the code is read from 5′ to 3′ direction. Moving from the 5 to the 3 end of an mRNA, each successive codon is sequentially interpreted into an amino acid (N-terminus to C-terminus of a polypeptide).
9. There exists a fixed reading frame for any gene that includes the initiation codon. The initiation codon specifies the first amino acid to be translated, which is usually AUG that codes for methionine. Mutations may modify the message encoded in sequence in three ways:
    a) **Frameshift mutations** where nucleotide insertions or deletions alter the genetic instruction for polypeptide by changing the reading frame.
    b) **Missence mutations** change a codon for one amino acid to a codon for a different amino acid.
    c) **Nonsense mutations** change a codon for an amino acid to stop codon.

Deciphering the Genetic Code

Nirenberg and Matthaei developed the technique in the laboratory of Khorana in 1961 to decipher the genetic code. They artificially synthesized mRNA having only 'Uracil' as the component. Next, they used this mRNA to synthesize polypeptide in cell free extracts of E. coli. When poly U was used, polypeptides consisting entirely of phenylalanine were produced indicating that UUU must code for phenylalanine. Similarly poly 'C RNAs produced polypeptides made entirely of proline, meaning that CCC must code for proline. This method was used to decipher the code of almost all amino acids.

Exceptions:

- In few single celled eukaryotic protozoans known as ciliates, UAA and UAG, which are nonsense codons in most organisms, specify the amino acid glutamine.

- In mitochondria of yeast, CUA specifies threonine instead of leucine.

mRNA codon	virus/pro & Eukaryotes	Mitochondria		
		Yeast	Drosophila	Mammal
AUA	Isoluecine	Methionine	Methionine	Methionine
AGA, AGG	Arginine	Arginine	Serine	Stop
CUA	Leucine	Threonine	Leucine	Leucine
UGA	Stop	Tryptophan	Tryptophan	Tryptophan

5.2 The translation machinery

5.2.1 tRNAs

About 15% of total RNA in a cell is tRNA. tRNAs plays a significant role by serving as adopter molecules that recognize the right enzyme activated amino acid and the anticodon on mRNA. Robert Holly (1965) first elucidated the base sequence of alanine tRNA from yeast. Later, more that 100 tRNA was identified and se~ quenced. All known tRNA share common structural features probably because tRNA molecules must be able to interact in nearly the same way with ribosomes, mRNA and elongation factors. Common features are as follows:

1. All tRNAs are single chains containing 73 to 93 ribonucleotide (~ 25kd)

2. tRNAs possess some unusual bases like inosine, pseudouridine, dihydrouridine, ribothymidine and methylated or derivatives of AUCG. (Methylation prevents the formation of base pairing, rendering them inaccessible for pairing with other base pairs or other type of interaction and also imparts hydrophobic character, important for interactions with synthetase and ribosomal proteins and for folding).

3. The 5′ end of tRNAs is phosphorylated and usually p-Guanine.

4. The base sequence at 3′ end of mature tRNA is always -CCA. Activated amino acid binds to the 3′ -OH group of the terminal adenosine.

5. About half the nucleotides in tRNAs are base paired to form double helix. Five groups are not base paired:

   a) 3′ CCA terminal

   b) T→ C loop (ribothymine-pseudouracil-cytosine)

   c) Extra arm (may have variable no. of residues) (present in only class II tRNA= serene and leucine

   d) Anticondon loop

   e) DHU loop (contains several dihydrouracil)

6. The anticodon loop (Fig. 5.1) consists of seven bases, with the following sequence -5′ Pyrimidine- Pyrimidine-**X-Y-Z**-modified purine- variable base 3′

7. X-ray crystallography study of phenylalanine tRNA by Alexander and Aaron (1974) showed that tRNA is a L-shaped structure. There are two segment of double helix; each having about 10 bases pairs in each turn, in accordance with the cloverleaf mode. Bases in the non-helical region participate in unusual bondings (e.g. GG, AA, CC). Moreover, 2′-OH of the ribose phosphate backbone acts as hydrogen donor and interacts with each other. Most bases are stacked on one another and the hydrophobic interactions between the aromatic rings help to stabilize the architecture of the molecule. Subsequent analyzing of all tRNAs showed that they follow the same basic plan.

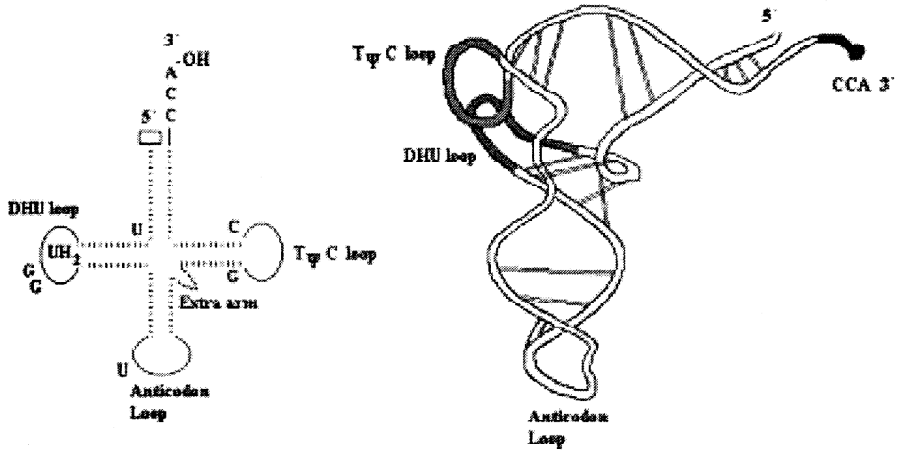


Fig. 5.1: The anticodon loop

In both prokaryotes and eukaryotes tRNA are synthesized as precursor molecules. Some pre t-RNA transcripts have several tRNA sequences, which are cleaved and modified to obtain mature functional tRNA. Some tRNA sequences are present within the pre-RNA transcripts.

5.2.2 Ribozyme

Ribozyme is an enzyme in which RNA component of a protein-RNA complex is responsible for the catalytic activity rather than the protein (Sidney Altman, 1983).

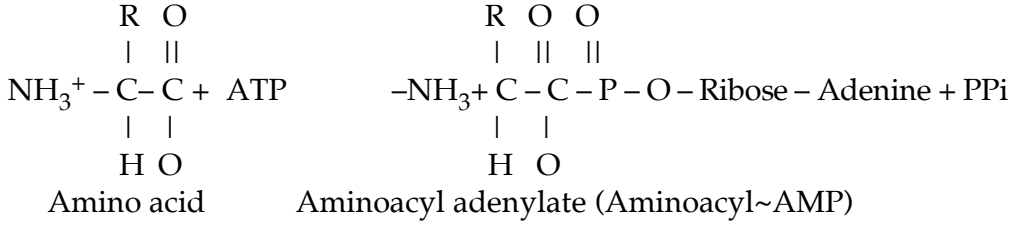
The 5′ end of tRNA is modified by RNase P (ribozyme). Conventional RNase modify the 3′ of end of tRNA which is the by followed by addition of -CCA nucleotides. Some tRNA has the information of-CCA already encoded in the DNA. All tRNAs have -CCA sequence at their 3' end.

In *E. coli,* there are 60 genes arranged in 25 units that synthesize tRNA molecules. One unit transcribes multimeric precursor for 7 tRNAs that are then cleaved by RNase P, RNase D to produce mature tRNA ($tRNA_{leu}$ 2 × $tRNA_{met,}$ 2 x $tRNA_{gtu,}$ 2 x tRNA $_{glu2}$)

5.2.3 Amino acid activation & linkage to specific tRNA by specific synthetase

Thermo dynamically, formation of peptide bond between -NH_2 of one amino acid and -COOH group of another amino acid is not favored. The barrier is overcome by activating the -COOH group of the precursor amino acid. Activation takes place by the following mechanism: -

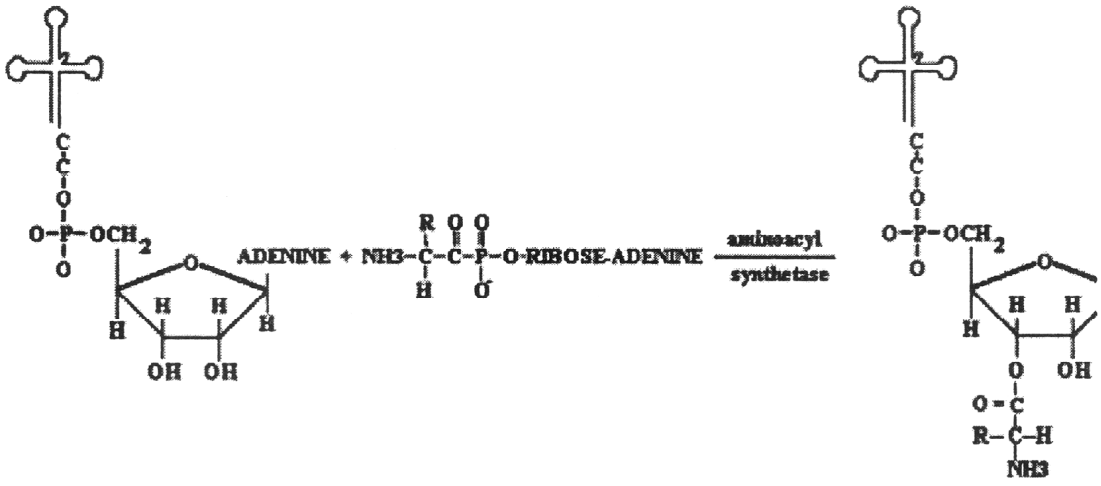
Step 1.



In the first step, an amino acid is linked to the phosphoryl group of AMP and therefore is known as aminoacyl~AMP

Step 2.

In the second step, aminoacyl group of aminoacyl-AMP is transferred to either 2′ or 3′ hydroxyl group of the ribose unit at the 3′ end of tRNA to form aminoacyl tRNA.



The attachment of amino acid to a tRNA is important not only because it activates the carboxyl group but also because amino acids by themselves cannot recognize the codons on mRNA. tRNA serves as an adopter molecule by recognizing the codons on one hand and by bringing specific amino acids, represented by the codons at the site of protein synthesis.

5.2.4 Aminoacyl tRNA synthetase

The amino acid is attached at the 3′ end of the tRNA to either the 2′ hydroxyl or the 3′ hydroxyl.

1. ***Class I*** *amino-acyl tRNA synthetases attach their associated amino acids to the tRNA* ***2' hydroxyl*** (NOTE: typically the hydrophobic amino acids)

2. ***Class II*** *amino-acyl tRNA synthetases attach their associated amino acids to the tRNA* ***3' hydroxyl*** (NOTE: typically hydrophilic amino acids)

For each amino acid, there exists a specific amonoacyl tRNA synthetase. These enzymes have been grouped into two classes on the basis of short signature sequences. In Class I there are 10 aminoacyl synthetases that recognize the larger amino acids and more hydrophobic enzymes. Class II (ancient type) aminoacyl tRNA synthetase recognizes the smaller amino acids. Class I enzymes acetylated the 2 -OH group and have a parallel ? domain (the classical dinucleotide binding fold) while class II enzymes acetylate the 3 -OH group (except phe) on the ribose and have an anti parallel ? domain (aminoacyl activating domain).

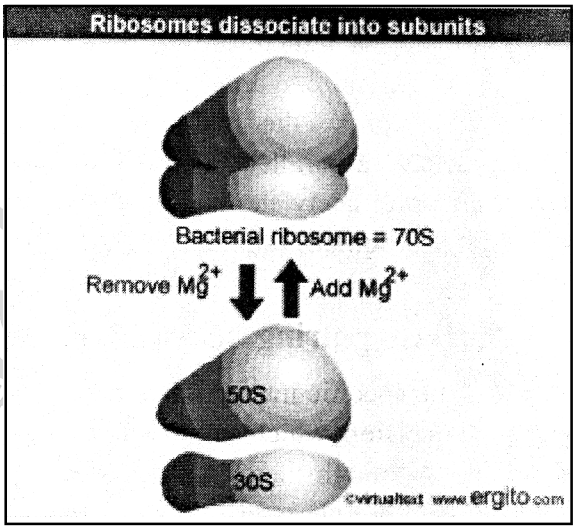
5.2.5 Base pairing between an mRNA codon and tRNA anticodon

It is the specific interaction between tRNA's anticodon and mRNA's codon that makes the decision, which amino acid will be incorporated into the growing polypep-tide chain. Although there is at least one kind of tRNA for each of the 20 amino acids, cells do not necessarily carry tRNAs with anticodons complementary to all of the 61 possible codon triplets in the degenerate genetic code. For e.g. in *E coli,* makes 79 different tRNAs containing 42 different anticodons. Obviously 19 (61-42=19) potential anticodons are not represented. Thus 19 mRNA codons will not find a complementary anticodon in E coli collection of tRNAs although such codons are present and are being coded into polypeptide chains. Therefore, there must be some tRNAs that recognize more than one codon for a particular amino acid. That is, the anticodons of these tRNAs can interact with more than one codon for the same amino acid. Although the exact codon-anticodon interaction of mRNA and tRNA is not very clear, Rrancis Crick, by analyzing the genetic code concluded that $3^{'}$ nucleotide in may codons adds nothing to the specificity of the codon. For example $5^{'}$ GGU$3^{'}$, 5′GGC3′, 5′GGA3′ and 5′GGG3' all encode glycine. It does not matter whether the anticodon on $tRNA^{gly}$ has a complementary base pare for the last codon at 3′ end provided the first two nucleotides are matched properly the tRNA will add glycine to the growing polypeptide chain. The same is true for other amino acids that are encoded by more than one codon. Thus the $5^{'}$ nucleotide

of tRNA's anticodon can often pair with more than one kind of nucleotide in the 3' position of an mRNA's codon. A tRNA charged with a particular amino acid can thus recognize several or even all of the codons for that amino acid. This flexibility in base pairing between the 3′ nucleotide in the codon and 5′ nucleotide in the anticodon is known as **wobble.**

5.2.6 Ribosome

Ribosomes are assembly of rRNA molecules and numerous proteins that synthesizes proteins under the direction from mRNA template. The bacterial system have ribosomes that sediment at 70S and the eukaryotic ribosomes sediment at 80S. Each ribosome can dissociate into two units: 50S and 30S in bacteria, while in eukaryotes it is 60S and 40S. The smaller subunit binds with mRNA to initiate translation.



Fi8- 5-2: Ribosomes dissociate into subunits

The actual number of rRNAs and number of proteins vary among species. For example, mammalian ribosomes have 4 rRNAs and 80 proteins while in *E. coli* the ribosomes have 3 rRNAs and 52 proteins. Although the number of proteins exceeds the number of rRNAs, the rRNAs constitute the major portion of the ribosomes and generally account for over 60% of the mass of the ribosome. The general structures of prokaryotic and eukaryotic ribosomes are more or less similar. Prior to translation, the ribosomes exist as two separate units- small subunit and larger subunit.

		rRNA	No. of proteins
Small subunit	Prokaryotes (30S)	16S	21 proteins
	Eukaryotes (40S)	18S	~30 proteins
Large subunit	Prokaryotes(50S)	23S and 5S	34 proteins
	Eukaryotes (60S)	28S, 5.8S, and 5S	~45 proteins

Each subunit has a specific three-dimensional shape that allows the two subunits to interlock with each other. The complete ribosome in prokaryotes and eukaryotes has a Svedberg value 70S and 80S respectively. There are two sites within the ribosome that can hold tRNAs: A site (aminoacyl or entry site) and the **P site** (peptidyl or donor site). The ribosome assembles on the mRNA with the A site oriented toward the 3′ end of the mRNA and the P site toward the 5′ end. At initiation, the three nucleotides of the initiation codon (AUG) align in the P site, where they pair with the anticodon in the initiator tRNA. This situates the A site over the second codon, which is now ready to receive the appropriate charged tRNA to continue translation.

During translation, several ribosomes may attach one after the other onto an mRNA and proceed through translation as chain of ribosomes. When this happens, the assembled ribosomes and the mRNA together are called polysome.

5.2.7 Ribozyme

Ribozymes are naturally occurring catalytic RNA molecules that have separate catalytic and substrate binding domains. The substrate binding domain binds to specific sequences of substrate RNA molecules by nucleotide complementarity and the catalytic domain cleaves the target RNA at a specific site.

The substrate binding domain can be engineered to bind to any target RNA and thus can be utilized a therapeutic agent. However, susceptibility of RNA molecules to enzymatic degradation in target cells and difficulty associated with the production of large scale synthetic RNA molecules has been solved by synthesizing an oligodeoxynucleotide with a ribozyme catalytic domain (~20 nucleotide) flanked by sequences that hybridize to the target mRNA after transcription. Such synthetic oligonucleotides are amplified and is cloned into eukaryotic expression vector. Transfection of target cells by engineered vectors will produce ribozymes that can cleaves the target mRNA, thereby suppressing the translation of the protein that is responsible for the disorder. Various cancers and viral diseases could be treated with genetically engineered riboszymes.

5.3 Prokaryotic and eukaryotic translation

Eukaryotic mRNA generally encodes a single polypeptide chain but prokaryotic are sometimes polysistronic. A mature mRNA of both prokaryotes and eukaryotes has coding and non-coding sequences.

The 5′ non-coding sequences are referred to as **'5′ untranslated region' (5'UTR).**

In prokaryotes, 5′ UTR have specific sequences called **'Shine Delgarno'** sequence, which just precedes the coding sequence. This sequence aligns the mRNA on 3′ end of 16S rRNA present in small unit of the ribosome.

.....**AGGAGGU**UUGACCU**AUG**..... pro-mRNA

.....**UCCUCCA**..... 16S rRNA

In Eukaryotes, the ribosomes recognize the 7-methylguanosine at the 5′ end. The ribosomes then scan downstream of the 5′ cap until they encounter an initiation codon i.e. AUG.

In both prokaryotic and eukaryotic cells, translation always initiates with the amino acid methionine, usually encoded by AUG. Alternative initiation codons, such as GUG that normally code for valine or CUG arginine are used by bacteria to code for N-fromyl-methionine at the initiation point.

The translation process is a complex mechanism that operates in the cytosol and requires the presence of mRNA, aminoacyl tRNA, ribosomes and many different protein factors. Translation takes place on ribosomes; which can be conceived as a moving protein-synthesizing machine. The entire translation process in both eukaryotes and prokaryotes can be distinguished into **initiation, elongation** and **termination.** Because more is known about translation in bacteria, the process described here will primarily focus bacterial translation. In most respects, eukaryotic translation is similar, although there are some significant differences that will be noted as we proceed through the stages of translation.

5.3.1 Initiation of translation in prokayotes

Initiation steps involves the association of specific **methionyl tRNA** (*f*-met-tRNA), **mRNA** and **ribosome subunits, initiation factors, guanosine triphosphate (GTP)** and recognition of the first codon, which in most cases is **AUG.**

- **IF1** facilitates the separation of the two ribosomal subunits
- At first the small **30S** subunit of the ribosome binds to the protein initiation factor **IF3** (Fig. 5.3a)

- **IF3** and **30S** subunit complex then binds to the Shine-Dalgarno sequence (AGGAGG) present on the mRNA at the 5′ end, approximately 7 nucleotide upstream the start codon **AUG** (AGGAGG: is complementary to the six nucleotides 3′ UCCUCC5′ on the **16SrRNA** at the 3end) (more than one Shine-Dalgarno sequence may be present in a single mRNA and therefore in most cases they are polycistronic).

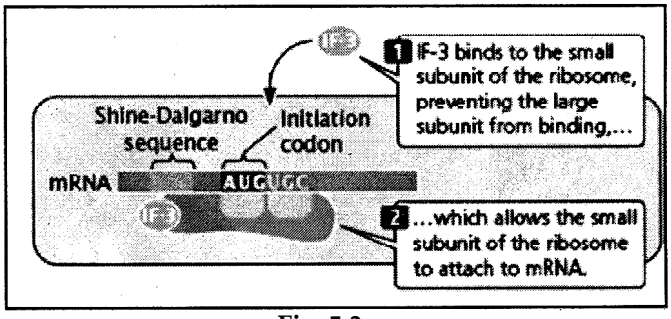


Fig. 5.3a

- Another initiation factor **IF2** that has a GTP bound to it, specifically recognizes and binds to the **initiator *f*-met-tRNA** and facilitates the binding of the *f*-met-tRNA to mRNA 3OS complex (Fig. 5.3b).

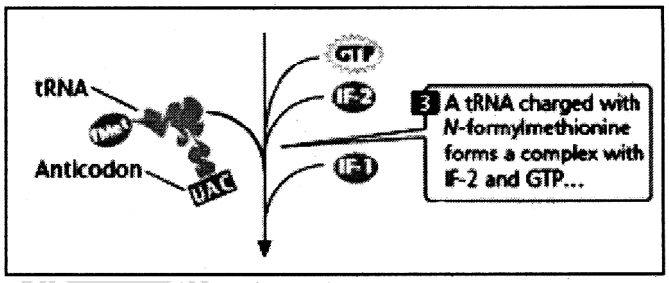


Fig. 5.3b

- The anticodon pairs with the initiation codon AUG on mRNA but lying within the 30S complex
- Binding of the *f*-met-tRNA releases the **IF3** from the initiation complex
- Release of **IF3** allows the 50S ribosomal subunit to bind to the complex
- 50S ribosome then triggers the hydrolysis of the GTP molecule bound to **IF2**
- Hydrolysis of the GTP results in the formation of **70S** complex
- Formation of **70S** ribosome and

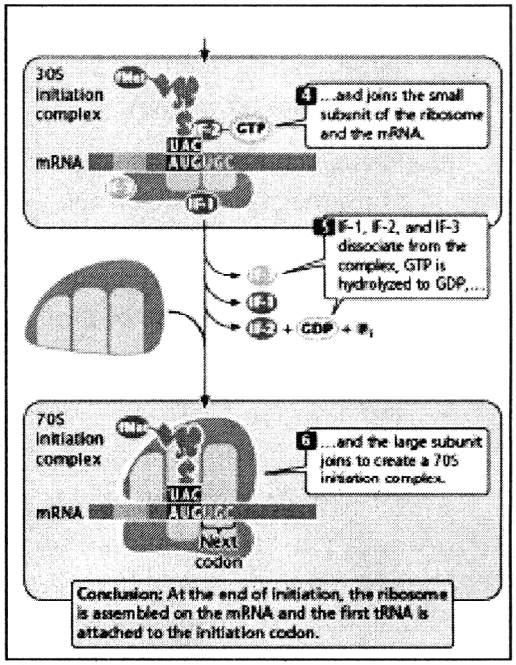


Fig. 5.3c

the binding of *f*-met-tRNA to mRNA at the initiation codon complete the initiation process that is now ready to begin peptide bond formation during elongation. (Fig. 5.3c)

- At initiation, the three nucleotides of the initiation codon **AUG** on mRNA align in the P site of ribosome, where they pair with the anticodon in the initiator tRNA. This arrangement of AUG in ribosome places the second codon in the A site, which is now ready to receive the appropriate charged tRNA to continue with the process of elongation of the polypeptide chain.

5.3.2 Initiation of translation in eukaryotes

In eukaryotes, Shine-Dalgarno sequences are absent and therefore the initiation occurs in a different way.

First, a eukaryotic initiation factor **elF4A,** a multimerie protein has a cap binding protein (CAP), that binds to the cap at the 5′ end of the mRNA. There exists several other factors with helicase activity that helps to unwind the secondary structures that may exists on mRNA.

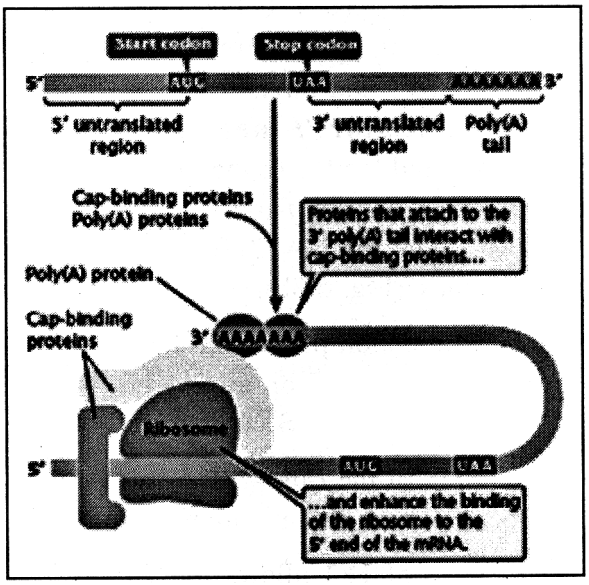


Fig. 5.4: Binding *of* the ribosome to the 5′ end of the mRNA

Then, a complex of the 40S ribosomal subunit with the initiator Met-tRNA.Met, along with several elF binds at the methylated cap of the 5′ end of the mRNA

The ribosomal subunit the scans down the mRNA to locate the initator codon AUG in the concensus in the Kozak sequence **ACC(AUG)G.**

The poly(A) tail at the 3′ end of eukaryotic mRNA also plays a role in the initiation of translation. Proteins that attach to the poly(A) tail interact with proteins that bind to the 5′ cap, enhancing the binding of the small subunit of the ribosome to the 5′ end of the mRNA. This interaction between the 5′ cap and the 3′ tail suggests that the mRNA bends backward during the

initiation of translation, forming a circular structure **(FIG** 2) A few eukaryotic mRNAs contain internal ribosome entry sites, where ribosomes can bind directly without first attaching to the 5' cap.

Once AUG is located, the 40S ribosome firmly binds with it and then 60S ribosomal subunit binds by displacing the elFs, producing the 80S initiation complex.

At initiation, the three nucleotides of the initiation codon (AUG) on mRNA align in the P site of ribosome, where they pair with the anticodon in the initiator tRNA. This arrangement of AUG in ribosome places the second codon in the A site, which is now ready to receive the appropriate charged tRNA to continue with the process of elongation of the polypeptide chain.

5.3.3 Elongation of the polypeptide chain

After the initiation process is over, the elongation of the polypeptide chain begins by joining charged amino acids. The process requires:

(1) the 70S complex;
(2) tRNAs charged with their amino acids;
(3) several elongation factors (EF-Ts, EF-Tu, and EF-G); and
(4) GTP.

A ribosome has three sites that can be occupied by tRNAs; the **aminoacyl,** or A, **site,** the **peptidyl,** or **P, site,** and the **exit,** or **E, site** (Fig. **5.5).**

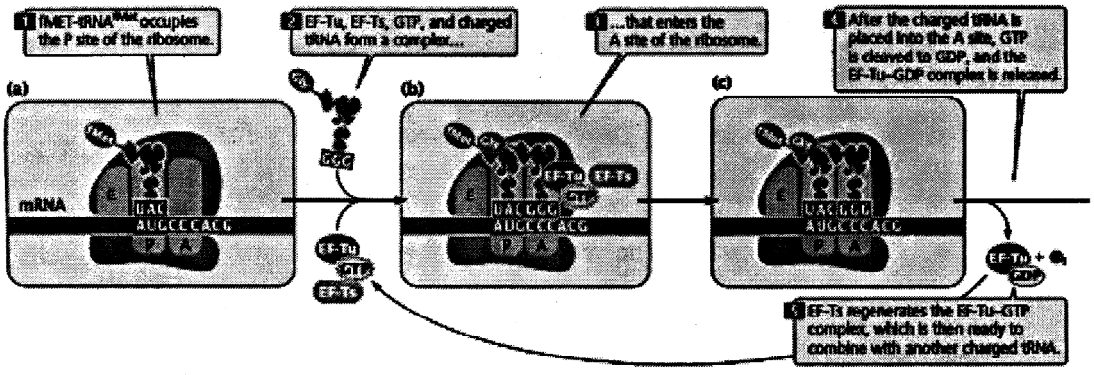


Fig. 5.5 : Sites of a ribosome

At first the ribosome occupies a position on mRNA in such a way that the P site is positioned over AUG and the adjacent A site is unoccupied (Fig 5.5a). The initiator tRNA immediately occupies the P site (the only site to which the fMet-tRNAfMet is capable of binding), but all other tRNAs first enter the A site.

Elongation occurs in three steps.

1. The first step is the delivery of a charged tRNA with its amino acid attached to the A site. This requires the presence of elongation factors **EF-Tu, EF-Ts,** and **GTP** (Fig. 5.5b).

   EF-Tu first joins with GTP and then binds to a charged tRNA to form a three-part complex.

   This three-part complex enters the A site of the ribosome, where the anticodon on the tRNA pairs with the codon on the mRNA.

   After the charged tRNA is in the A site, GTP is cleaved to GDP, and the EF-Tu-GDP complex is released (Fig. 5.5c)

   Factor EF-Ts regenerates EF-Tu-GDP to EF-Tu-GTP.

   In eukaryotic cells, a similar set of reactions delivers the charged tRNA to the A site.

2. The second step of elongation is the creation of a peptide bond between the amino acids that are attached to tRNAs in the P and A sites **(Fig. 5.6**). The

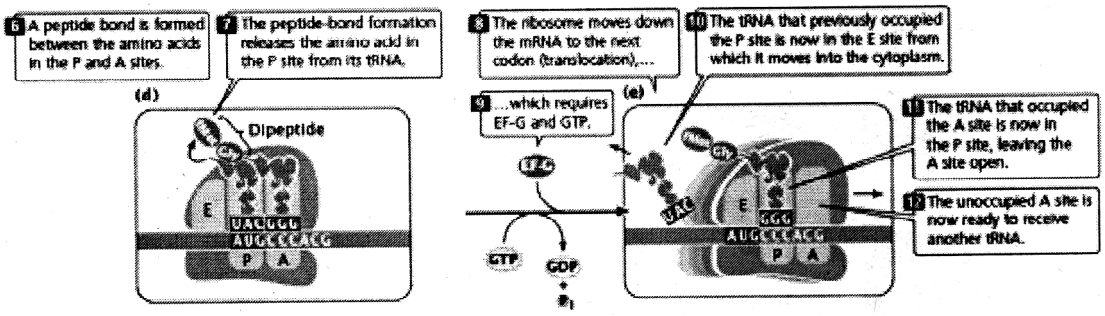


Fig. 5.6: The second step of elongation

formation of this peptide bond releases the amino acid in the P site from its tRNA. The activity responsible for peptidebond formation in the ribosome is referred to as **peptidyl transferase.** The peptidyl bond formation is catalyzed by the rRNA of the large subunit of the ribosome.

The third step in elongation is **translocation, (Fig. 5.7)** where the ribosome moves down the mRNA in the 5′—3′ direction. The A site of the ribosome moves

forward to occupy the next codon. The process requires **EF-G** (elongation factor) and the hydrolysis of **GTP** to **GDP.** Because the tRNAs in the P and A site are still attached to the mRNA through codon- anticodon pairing, they do not move with the ribosome as it translocates. Consequently, the tRNA that previously was occupying the P site now occupies the E site, from which it moves into the cytoplasm. tRNA that occupied the A site now occupies the P site, leaving the A site open for the next incoming tRNA specified by the mRNA's codon sequence. Thus, the progress of each tRNA through the ribosome during elongation can be summarized as follows:

Cytoplasm → A site → P site → E site → cytoplasm

Throughout the cycle, the polypeptide chain remains attached to the tRNA in the P site. The ribosome moves down the mRNA in the 5′ —3′ direction, adding amino acids one at a time.

Elongation in eukaryotic cells takes place in a similar manner.

5.3.4 Termination

Addition of new amino acids stops when the A site of ribosome translocates to a termination codon. Because there are no tRNAs with anticodons complementary to the termination codons, no tRNA enters the A site of the ribosome when a termination codon is encountered (**Fig. 5.7a**). Instead, proteins called **release factors** bind to the ribosome (**Fig. 5.7b**). *E. coli* has three release factors—RF1, RF2, and RF3. Release factor 1 recognizes the termination codons UAA and UAG, and RF2 recognizes UGA and UAA. Release factor 3 forms a complex with GTP and binds to the ribosome. The release factors then promote the cleavage of the tRNA in the P site from the polypeptide chain; in the process, the GTP that is

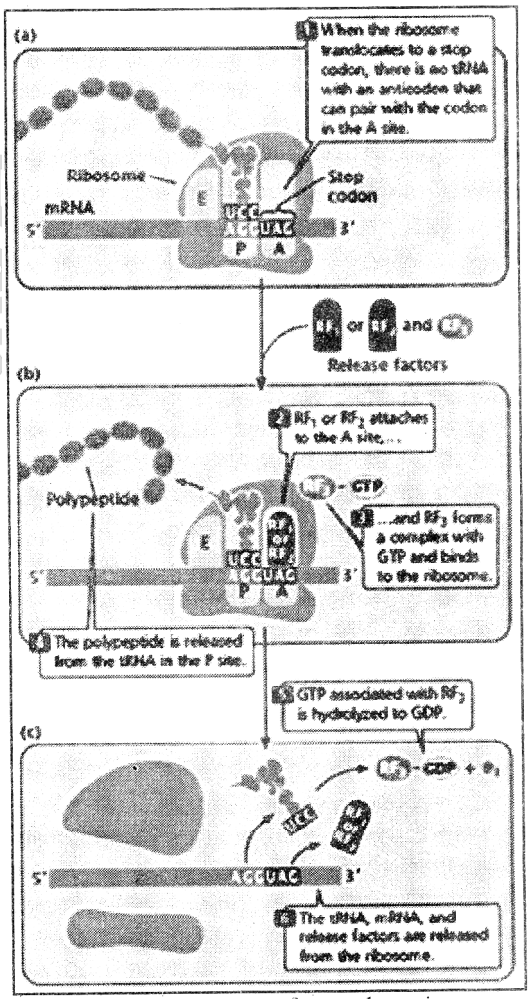


Fig. 5.7: Steps of translocation

complexed to RF3 is hydrolyzed to GDP. Additional factors help bring about the release of the tRNA from the P site, the release of the mRNA from the ribosome, and the dissociation of the ribosome **(Fig. 5.7c).** Findings from recent studies suggest that the release factors bring about the termination of translation by completing a final elongation cycle of protein synthesis. In this model, RF1 and RF2 are similar in size and shape to tRNAs and occupy the A site of the ribosome, just as the amino acid-tRNA-EF-Tu-GTP complex does during an elongation cycle. Release factor 3 is structurally similar to EF-G; it then translocates RF1 and RF2 to the P site, as well as the last tRNA to the E site, in a way similar to that in which EF-G brings about trans location. When both the A site and the P site of the ribosome are cleared of tRNAs, the ribosome can dissociate. Research findings also indicate that some of the sequences in the rRNA play a role in the recognition of termination codons.

Translation in eukaryotic cells terminates in a similar way, except that there are two release factors: eRFl, which recognizes all three termination codons, and eRF2, which binds GTP and stimulates the release of the polypeptide from the ribosome.

5.4 Regulation of translation

Regulation of gene expression refers to the cellular control of the amount and timing of changes to the appearance of the functional product of a gene. Gene regulation gives the cell control over its structure and function, and is the basis for cellular differentiation, morphogenesis and the versatility and adaptability of any organism. Prokaryotic and eukaryotic gene expressions are regulated at several different levels of flow of information from gene to their final product. Any step of gene expression may be modulated, from the DNA-RNA transcription step to post-translational modification of a protein. Stages where gene expression is regulated are:

Regulate gene expression by chemical and structural modification of DNA or chromatin (e.g. by methylation, phosphorylation, acetylation or structural changes)

By controlling when and how often a given gene is transcribed (during transcription e.g. by repressor or activator proteins, attenuation etc.)

By modifying the primary RNA transcript (post transcriptional modifications)

Selection of transcribed and processed RNA (transport from nucleus to cytoplasm)

During translation of mRNA (e.g. by an antisense RNA)

Post-translational control (e.g. by proteolysis or modification of the gene product)

5.4.1 DNA modification

1. Chemical modification of DNA

Methyiation of DNA refers to addition of a methyl group to the number 5 carbon of the cytosine pyrimidine ring by the enzyme methyl transferase. Generally methyiation occurs on the cytosine in the CpG dinucleotide sequence (CpG islands). DNA methyiation pattern can be inherited without changing the DNA sequence. As such, it is part of the epigenetic code and is the most characterized epigenetic (changes in phenotype without any alteration in the genomic material) mechanism.

DNA methyiation has been found in all vertebrate. In humans, approximately 1% of DNA bases undergo DNA methyiation (~60-70% of all CpGs are methylated mainly 5′ regulatory regions). In adult somatic tissues, DNA methyiation typically occurs in a CpG dinucleotide context; non-CpG methyiation is prevalent in embryonic stem cells. In plants, cytosines are methylated both symmetrically (CpG or CpNpG) and asymmetrically (CpNpNp), where N can be any nucleotide.

DNA methyiation may impact the transcription of genes in two ways.

1. First, the methyiation of DNA may itself physically impede the binding of transcriptional proteins to the gene, thus blocking transcription.

2. Second, and likely more important, methylated DNA may be bound by proteins known as Methyl-CpG-binding domain proteins' (MBDs). MBD proteins then recruit additional proteins to the locus, such as histone deacetylases and other chromatin remodelling proteins that can modify histones, thereby forming compact, inactive chromatin termed silent chromatin.

This link between DNA methyiation and chromatin structure is very important. . In many disease processes such as cancer, gene promoter CpG islands acquire abnormal hypermelhylation, which results in heritable transcriptional silencing In particular, loss of Methyl-CpG-binding Protein 2 (MeCP2) has been

implicated in Rett syndrome and Methyl-CpG binding domain protein 2 (MBD2) mediates the tran-scriptional silencing of hypermethylated genes in cancer

Analysis of the pattern of methylation in a given region of DNA (generally a promoter) can be achieved through a method called bisulfite mapping. Methylated cytosine residues are unchanged by the treatment, whereas unmethylated ones are changed to uracil. The differences are analyzed **in** sequencing gels, Abnormal methylation patterns are thought to be involved in carcinogenesis.

2. Structural modification of DNA

Transcription of DNA is highly dependent on the secondary structure of DNA molecule. Histone proteins, responsible for supercoiling of DNA can modify the structure temporarily or more permanently depending on the phosphorylation or methylation of the histone proteins respectively. Such modifications influences the level of gene expression. In general, the density of its packing is indicative of the frequency of transcription.

Histone acetylation is also an important process in transcription. Histone acetyltransferase enzymes (HATs) such as CREB-binding protein also dissociate the DNA from the histone complex, allowing transcription to proceed. Often, DNA methylation and histone acetylation work together in gene silencing. The combination of the two seems to be a signal for DNA to be packed more densely, lowering gene expression.

5.4.2 Regulation of transcription

Transcription is an important level of control in eukaryotic cells. Transcription regulation of a gene by RNA polymerase can be regulated by at least five mechanisms:

Specificity factors alter the binding specificity of RNA polymerase for a given promoter or set of promoters, making it more or less likely to bind to them (i.e. sigma factors used in prokaryotic transcription).

Repressors bind to non-coding sequences on the DNA strand that are close to or overlapping the promoter region, impeding RNA polymerase's progress along the strand, thus impeding the expression of the gene.

Basal factors These transcription factors position RNA polymerase at the start of a protein-coding sequence and then release the polymerase to transcribe the mRNA. Recruitment of these proteins at the promoter region affects the RNA polymerase activity.

Activators enhance the interaction between RNA polymerase and a particular promoter, encouraging the expression of the gene. Activators do this by increasing the attraction of RNA polymerase for the promoter, through interactions with subunits of the RNA polymerase or indirectly by changing the structure of the DNA.

Enhancers are sites on the DNA helix that are bound to by activators in order to loop the DNA bringing a specific promoter to the initiation complex.

Examples :

The $ó^{32}$ subunit of RNA polymerase changes itself in such a way that the enzyme binds to a specialized set of promoters when E. coli bacteria are subjected to heat stress producing heat-shock response proteins.

When there is excess tryptophan in the cell, the amino acid binds to a specialized repressor protein, changing the structural conformity of the repressor such that it binds to the operator region for the operon that synthesizes tryptophan, preventing their expression and thus suspending production which also represents a form of negative feedback mechanism.

In bacteria, the lac repressor protein blocks the synthesis of enzymes that digest lactose when there is no lactose to feed on. When lactose is present, it binds to the repressor, causing it to detach from the DNA strand.

5.4.3 Gene regulation can be summarized as how they respond

Inducible systems - An inducible system is off unless there is the presence of some molecule (called an inducer) that allows for gene expression. The molecule is said to "induce expression".

Repressible systems - A repressible system is on except in the presence of some molecule (called a corepressor) that suppresses gene expression. The molecule is said to "repress expression". In both the cases, the control mechanism varies in prokaryotic and eukaryotic cells.

5.4.4 Post-transcriptional regulation

The cells regulate the posttranscriptional activity by several way to check how much the mRNA should be translated into proteins. Cells do this by Capping, Splicing, and the addition of a Poly(A) Tail. These processes occur only in

eukaryotes because in prokaryotes, the transcription and translation is coupled.

Capping changes the five prime end of the mRNA to a three prime end by 5'-5' linkage, which protects the mRNA from 5' exonuclease, which degrades foreign RNA. The cap also helps in ribosomal binding.

Splicing removes the introns, noncoding regions that are transcribed into RNA, in order to make the mRNA able to create proteins. Cells do this by spliceosome's binding on either side of an intron, looping the intron into a circle and then cleaving it off. The two ends of the extrons are then joined together.

Addition of poly(A) tail a poly(A) tail is just junk RNA added to the 3' end in order to slowly be degraded by a 3' exonuclease in order to increase the half life of mRNA.

Unit 6 Antisense and Ribozyme Technology

Structure

6.1 Antisense molecules and their mechanism of action

6.1.1 Introduction

Expression of some genes may be regulated or suppressed with the aid of small RNA molecules - a process termed as RNA silencing, also known as RNA interference or posttranscriptional gene silencing. Although many of the details of this mechanism are still poorly understood, it appears to be widespread, existing in fungi, plants, and animals. It may also prove to be a powerful tool for artificially regulating gene expression in genetically engineered organisms.

The discovery of antisense RNA was preceded first by observation of transcriptional inhibition by antisense RNA expressed in transgenic plants. In an attempt to alter flower colors in petunias, researchers introduced additional copies of a gene encoding chalcone synthase, a key enzyme for flower pigmentation into petunia plants of normally pink or violet flower color. The scientists' goal was to produce petunia plants with improved flower colors but instead produced less pigmented, fully or partially white flowers, indicating that the activity of chalcone synthase had been substantially decreased. Similar suppression of gene activity was also observed in fungus *Neurospora crassa.* This phenomenon was called *co-suppression of gene expression,* but the molecular mechanism remained unknown. Sometime later, plant virologists noted that plants carrying only short transgenic non-coding regions of viral RNA sequences would showed enhanced levels of protection. The reverse experiment, in which short sequences of plant genes were introduced into viruses, showed that the targeted gene was suppressed in an infected plant. This phenomenon was labeled "virus-induced gene silencing" (VIGS), and the set of such phenomena were collectively called post transcriptional gene silencing. Craig C. Mello and Andrew Fire's 1998 published in *Nature* the potent gene silencing effect after injecting double stranded RNA into C. *elegans.* In investigating the regulation of muscle protein production, they observed that neither mRNA nor antisense RNA injections had an effect on protein production,

but double-stranded RNA successfully silenced the targeted gene. As a result of this work they were awarded Nobel Prize in the year 2006.

6.1.2 Antisense RNA molecules

Antisense molecules are nucleotide sequences that interact with complementary strands of nucleic acids and modify expression of genes. For example, Antisense RNA is single-stranded RNA that is complementary to an mRNA strand transcribed within a cell and capable of blocking the translation machinery. Antisense molecules occur naturally. For example, in both mice and humans, the gene for the insulin-like growth factor 2 receptor (Igf2r) that is inherited from the father synthesizes an antisense RNA that appears to block synthesis of the mRNA for Igf2r.

Historically, the effects of antisense RNA have often been confused with the effects of RNA interference, a related process in which double-stranded RNA fragments called small interfering RNAs trigger catalytically mediated gene silencing, most typically by targeting the RNA-induced silencing complex (RISC) to bind to and degrade the mRNA.

6.1.3 RNA interference (RNAi)

In the course of working with artificially synthesized single stranded antisense RNA molecules, it was discovered that double stranded RNA (dsRNA) molecule can also act as a powerful suppressant of genes expression. dsRNA may arise in several ways: by the transcription of inverted repeats in DNA into a single RNA molecule that base pairs with itself; by the simultaneous transcription of two different RNA molecules that are complementary to one another and pair; or by the replication of double-stranded RNA viruses (Fig 6.1).

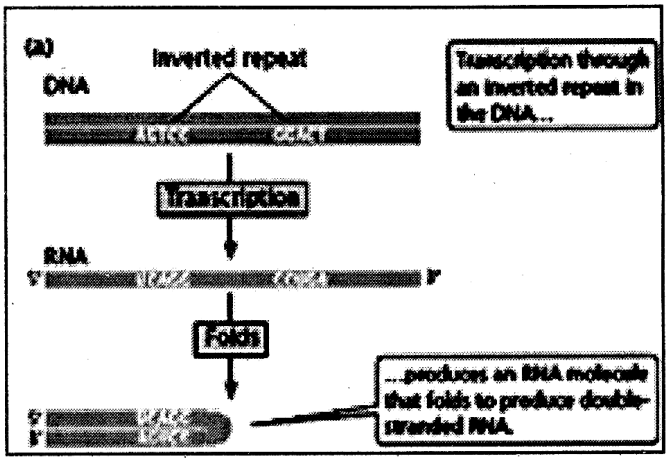


Fig. 6.1: Replication of double-stranded RNA viruses

Drosophila, an enzyme called Dicer cleaves and processes the double-stranded RNA to produce In fact, the suppressive effect of antisense RNA probably also depends on its ability to form dsRNA. The ability of dsRNA to suppress the

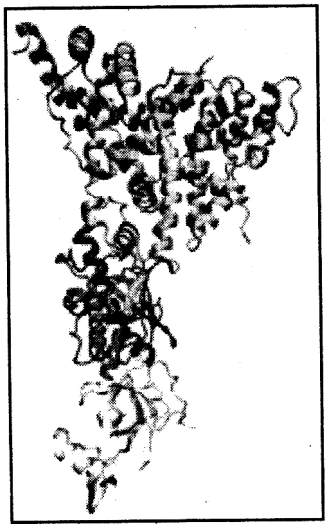


Fig 6.1 a.

expression of a gene corresponding to its own sequence is called RNA interference (RNAi). It is also called post-transcriptional gene silencing or PTGS.

6.1.4 Mechanism of RNAi

Normally single-stranded RNA molecules are found in the cytoplasm of a cell. If double-stranded RNA (dsRNA) formation occurs in the cell, the enzyme called Dicer to cleave the dsRNA into fragments containing 19 base pairs (~2 turns of a double helix) with two additional nucleotides at the opposite end of each strand. The two strands of each fragment then separate — releasing the antisense strand.

Dicer *is a ribonuclease in the RNase III family that cleaves double-stranded RNA (dsRNA) and pre-microRNA (miRNA) into short double-stranded RNA fragments called small interfering RNA (siRNA) about 20-25 nucleotides long, usually with a two-base overhang on the 3' end. Dicer contains two RNase III domains and one PAZ domain; the distance between these two domains of the molecule is determined by the length and angle of the connector helix and determines the length of the siRNAs it produces. Dicer catalyzes the first step in the RNA interference pathway and initiates formation of the RNAAnduced silencing complex (RISC), whose catalytic component argonaute is an endonuclease capable of degrading messenger RNA (mRNA) whose sequence is complementary to that of the siRNA guide strand.*

With the aid of a protein, it binds to a complementary sense sequence on a molecule of mRNA. If the base-pairing is exact, the mRNA is destroyed. Because of their action, these fragments of RNA have been named "short (or small)

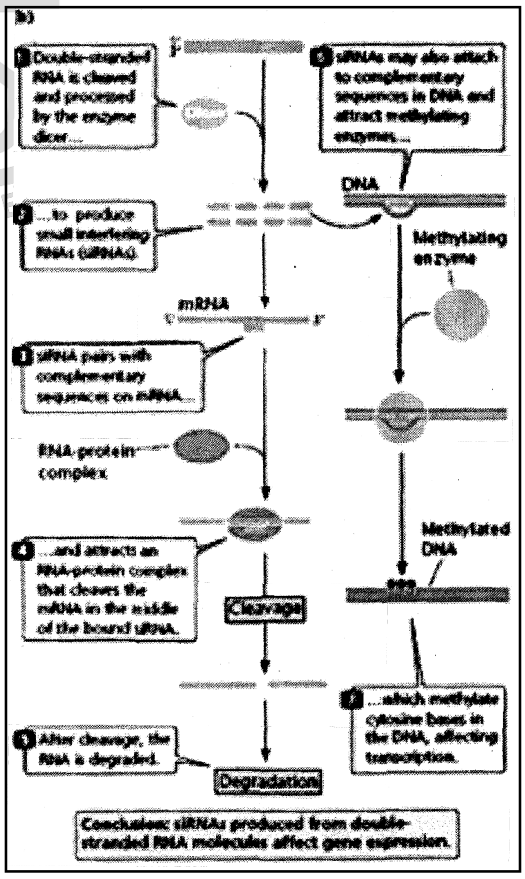


Fig. 6.2b:

interfering RNA" **(siRNA).** The complex of siRNA and protein is called the "**R**NA-induced silencing complex" **(RISC)** (Fig. 6.2b).

In fission yeast, evidences indicate that siRNAs can also inhibit the transcription of genes perhaps by binding to complementary sequences on DNA or by binding to the nascent RNA transcript as it is being formed. Synthetic siRNA molecules that bind to gene promoters can repress transcription by methylation of the DNA in the promoter and, perhaps, methylation of histones in the vicinity. The siRNA forms a complex called the RITS complex ("RNA-induced initiation of transcriptional gene silencing") with at least three different proteins. How these siRNAs, synthesized in the cytosol gain access to the DNA in the nucleus is unknown.

Example :

The rice plant of strain LGC-1 produces abnormally low levels of proteins called glutelins although there are several glutelin genes. Interestingly, it was observed that two closely-similar glutelin genes are located back to back on the same chromosome and a deletion has occurred in the 3' region of the first glutelin gene that has removed the stop signal. As a consequence, RNA polymerase II transcribes right past the first gene and on into the second. The result is a messenger RNA with almost-identical sequences running in opposite directions. Such a composition of the mRNA molecule allows it to fold up into a molecule of double-stranded RNA (dsRNA). A Dicer-like enzyme cuts up the dsRNA into small interfering RNAs (siRNAs) that suppress further transcription of those genes as well as other glutelin genes

Some related. RNA molecules produced through the cleavage of double-stranded RNA bind to complementary sequences in the 3′UTR of mRNA and inhibit their translation. RNA silencing is thought to have evolved as a defense against RNA viruses and transposable elements that move through an RNA intermediate (see Chapter 20). The extent to which it contributes to normal gene regulation is uncertain, but dramatic phenotypic effects result from some mutations that occur in the enzymes that carry out RNA silencing.

Amplification of RNAi : In C. elegans, plants, and Neurospora, the introduction of a few molecules of dsRNA has a potent and long-lasting effect. In plants, the gene silencing spreads to adjacent cells (through plasmodesmata) and even to other parts of the plant (through the phloem). RNAi within a cell can continue after mitosis in the progeny of that cell. Triggering of RNAi in C. elegans can even pass through the germline into its descendants. Such amplification of an initial trigger signal suggests a catalytic effect. It turns out that these organisms have RNA-dependent RNA poly-merases (RdRPs) that uses the mRNA targeted

by the initial antisense siRNA as a template for the synthesis of more siRNAs. Synthesis of these "secondary siRNAs even occurs in adjacent regions of the mRNA. So not only can these secondary siRNAs target additional areas of the original mRNA, but they are potentially able to silence mRNAs of other genes that may carry the same sequence of nucleotides. This phenomenon, called **"transitive RNAi".**

In mammalian cells, introducing dsRNA fragments only reduces gene expression temporarily. However, Brummelkamp *et. al.* report in the 19 April 2002 issue of **Science** that they have succeeded in introducing into (mammalian) cells a **DNA vector** that can continuously synthesize a siRNA corresponding to the gene that they want to suppress. Two months later the cells still failed to manufacture the protein whose gene had been turned off by RNAi.

6.1.5 MicroRNAs (miRNAs)

MicroRNAs were first described by Lee, et al. in 1993. **MicroRNAs (miRNA)** are single-stranded RNA molecules of about 21-23 nucleotides in length thought to regulate the expression of other genes. miRNAs are encoded by genes that are transcribed from DNA but not translated into protein (non-coding RNA); instead they are processed from primary transcripts known as *pri-miRNA* to short stem-loop structures called *pre-miRNA* and finally to functional miRNA, Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules, and they function to downregulate gene expression.

The genes encoding miRNAs are much longer than the processed miRNA molecule. miRNAs are first transcribed as primary transcripts or pri-miRNA and processed to short, 70-nucleotide stem-loop structures known as pre-miRNA in the cell nucleus by protein complex known as the Microprocessor complex, consisting of the nuclease Drosha and the double-stranded RNA binding protein Pasha. These pre-miRNAs are then processed to mature miRNAs in the cytoplasm by interaction with the endonuclease Dicer, which also initiates the formation of the RNA-induced silencing complex (RISC). This complex is responsible for the gene silencing observed due to miRNA expression and RNA interference. The pathway in plants varies slightly due to their lack of Drosha homologs; instead, Dicer homologs alone effect several processing steps.

In *C. elegans,* successful development through its larval stages and on to the adult requires the presence of at least two **"microRNAs"** ("miRNAs") — single-stranded RNA molecules containing about 22 nucleotides and thus about the same size as siRNAs. These small single-stranded transcripts are generated by the cleavage of larger precursors using the C. elegans version of Dicer. The miRNA acts by either destroying or inhibiting translation of several messenger RNAs in

the worm (by binding to a region of complementary sequence in the 3' untranslated region [3^1-UTR] of the mRNA). miRNA genes have also been discovered in humans Dro-sophila, mice, frogs, fish, and plants (Arabidopsis) as well as in C. elegans.

6.1.6 Biological functions of antisense molecules

Immunity :

RNA interference provides immunity to plants especially against viruses and other foreign genetic materials. It is also suggested the RNA interference may also prevent self-propagation by transposons. Even before the RNAi pathway was fully understood, it was known that induced gene silencing in plants could spread throughout the plant in a systemic effect, and could be transferred from stock to scion plants via grafting. This phenomenon is a feature of the plants innate immune system, and allows the entire plant to respond to a virus after an initial localized encounter.

Although animals generally express fewer variants of the dicer enzyme than plants, RNAi in some animals has also been shown to produce an antiviral response. In both juvenile and adult *Drosophila*, RNA interference is important in antiviral innate immunity and is active against pathogens such as Drosophila X virus. A similar role in immunity may operate in *C. elegans*, as argonaute proteins are upregulated in response to viruses and worms that overexpress components of the RNAi pathway are resistant to viral infection.

The role of RNA interference in mammalian innate immunity is poorly understood and the hypothesis of RNAi-mediated immunity in mammals has been challenged as relatively little data is available. However, alternative functions for RNAi in mammalian viruses exist, such as miRNAs expressed by the herpes virus that may act as heterochromatin organization that triggers to mediate viral latency.

Genome maintenance

Components of the RNA interference pathway are used in many eukaryotes in the maintenance of the organisation and structure of their genomes. Modification of histones and associated induction of heterochromatin formation serves to downregulate genes pre-transcriptionally and this process is referred to as RNA-induced transcrip-tional silencing (RITS), and is carried out by a complex of proteins called the RITS complex. In fission yeast this complex contains argonaute, a chromodomain protein Chpl, and a protein called Tas3 of unknown function. As a consequence, the induction and spread of heterochromatic regions requires the argonaute and RdRP proteins. Indeed, deletion of these genes in the fission yeast *S. pombe* disrupts histone meth-ylation and centromere formation, causing slow or

stalled anaphase during cell division. Thus repression of gene expression by miRNAs appears to be a mechanism to ensure proper, coordinated gene expression as cells differentiate along particular paths. For example, when zygote. genes begin to be turned on in the zebrafish blas-tula, one of them encodes a miRNA that triggers the destruction of the maternal mRNAs that have been running things up to then. So miRNAs may play as important role as transcription factors in coordinating the expression of multiple genes in a particular type of cell at particular times.

6.2 Splicing repressors and activators control splicing at alternative sites

6.2.1 Introduction

Gene splicing mechanism exists in eukaryotes that facilitate the removal of intron from pre-mRNA. There are also evidences to suggest that alternative splicing enable the synthesis of two different proteins from the same peptide. For the alternative splicing mechanism to operate, it is believed that there exists splicing repressor proteins in the cells and function to produce alternative peptides from the same genes. For example, in *Drosophila* Sxl inhibit splicing at specifice sites, causing exons to be skipped, whereas Tra promotes splicing. As a consequence,

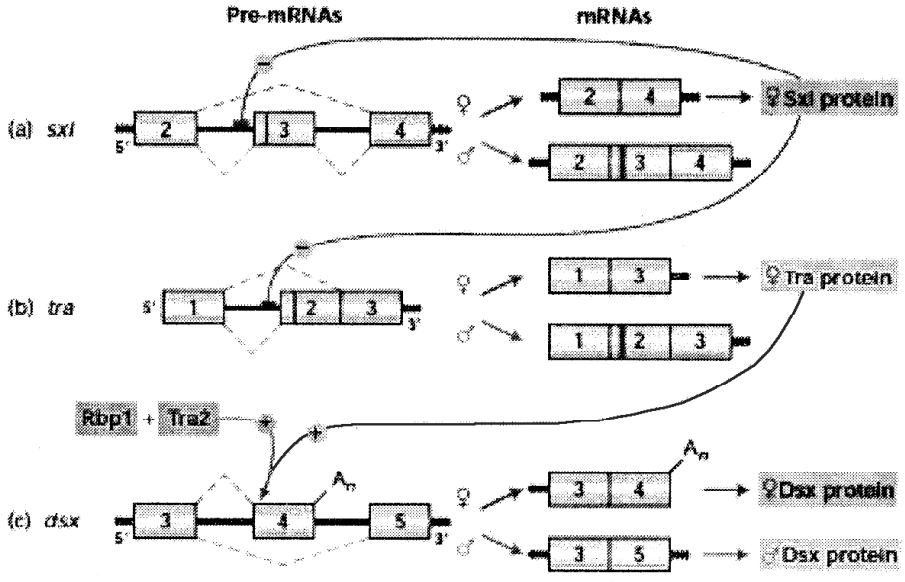


Fig. 6.3 : Alternative spilicing due to repression at splicing sites lead to the formation of alternative Dsx protein in Drosophila

distinct Dsx proteins are produced in female and male embryos leading to sexual differentiation (Fig. 6.3). Sxl-like splicing repressor expressed in hepatocytes might bind to splice sites for the EIIIA and EIIIB exons in the fibronectin pre-mRNA, causing them to be skipped during RNA splicing. Experimental examination in some systems has revealed that inclusion of an exon in some cell types versus skipping of the same exon in other cell types results from the combined influence of several splicing repressors and enhancers.

The action of similar proteins may explain the cell-type specific expression of fibronectin isoforms in humans. Splicing repressors expressed in hepatocytes might **bind** to splice sites for the EIIIA and EIIIB exons in the fibronectin pre-mRNA, causing them to be skipped during RNA splicing (Figure: 6.4). Alternatively, a Tra-like splicing activator expressed in fibroblasts might activate the splice sites associated with the fibronectin EIIIA and EIIIB exons, leading to inclusion of these exons in the mature mRNA.

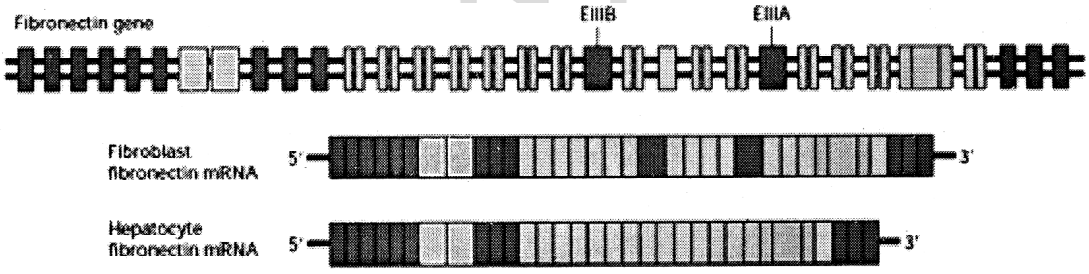


Fig.6.4: Alternative splicing fibronectin pre mRNA in human hepatocytes can produce different isoforms.

6.2.2 Repression of translation of mRNAs

Micro RNAs (miRNAs) are efficient tools for the sells to regulate mRNA translation. In C. elegans, the genes lin-4 and let-7 produce small RNA sequences 21 and 22 nucleotides long, respectively, that hybridize to the 3' untranslated regions of specific target mRNAs. For example, the *lin-4* miRNA, which is expressed early in embryogenesis, hybridizes

to the 3' untranslated regions of both the *lin-14* and *lin-28* mRNAs, thereby repressing translation of these mRNAs by an as yet unknown mechanism. Expression of *lin-4* miRNA ceases later in development, allowing translation of newly synthesized *lin-14* and *lin-28* mRNAs at that time.

In *C. elegans;* about 100 different miRNAs have been found in *C. elegans*, and

at least as many in humans. All miRNAs appear to be formed by processing of ~70-nucleotide precursor RNAs that form hairpin structures with a few base-pair mismatches in the stem of the hairpin. A ribonuclease called *Dicer*, cleaves the double-stranded RNA to produce miRNAs precursors. Interestingly, the base pairing between a miRNA and the 3' untranslated region of its target mRNAs is not precisely complementary and some base-pair mismatches occur in the hybridized region. This mismatching distinguishes miRNA-mediated translational repression from the related phenomenon of RNA interference, which we describe next.

6.2.3 Degradation of mRNAs in the cytoplasm

Concentration of any mRNAs in the cell depends not only on the rate of its transcription but also on the rate of its degradation. With more stable mRNAs, protein synthesis persists long after transcription of the gene is repressed. Usually, bacterial mRNAs are unstable and decay exponentially, probably because they need to switch genes rapidly in response to the change in environment. In multicellular organism, the cells reside in a more stable environment and do not require frequent adjustment to the changing environment. However, some proteins in eukaryotic cells are required only for short periods of time and must be degraded immediately. For example, during cell cycle activity, synthesis of cyclins occurs in burst at intervals. The mRNA of cyclins must be degraded very quickly. mRNA of other proteins like c-Fos and c-Jun, synthesized during S phase need to be degraded immediately after the function is over. mRNAs of such proteins have half life less than 30 minutes.

There are three main pathway that lead to the degradation of cytoplasmic mRNAs as shown in (Fig. 6.5).

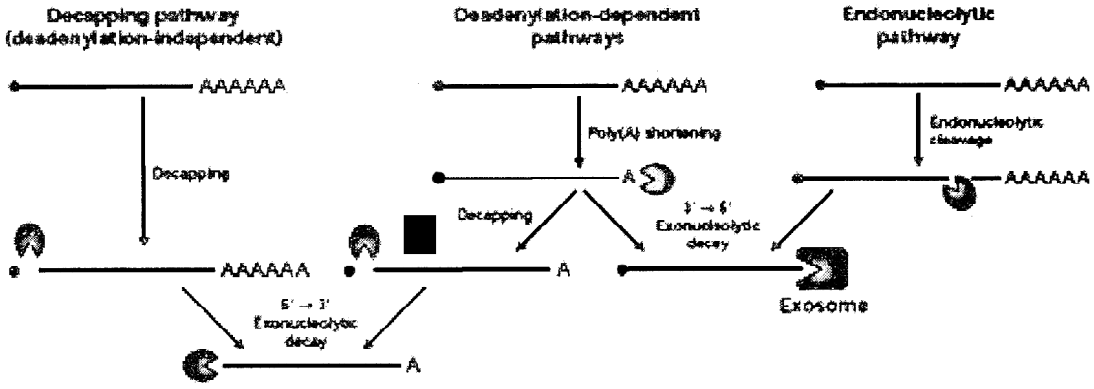


Fig.6.5: Main pathway that lead to degradation of cytoplasmic mRNAs

For most mRNAs, the length of the poly(A) tail gradually decreases with time through the action of a deadenylating nuclease. When it is shortened sufficiently, the PABPI molecules that bind during polyadenylation of mRNA can no longer bind and stabilize interaction of the 5′ cap and initiation factors (Fig. 6.6). The exposed cap then is removed by a decapping enzyme, and the unprotected mRNA is degraded by a 5′→3′ exonuclease. Removal of the poly(A) tail also makes mRNAs susceptible to degradation by cytoplasmic exosomes containing 3′→5′ exonucleases. The 5′→3′ exonucleases predominate in yeast, and the 3′→5′ exosome apparently predominates in mammalian cells. The rate of deadenylation determines the rate of degradation of mRNA. Recent experiments suggest that the bound proteins interact with a deadenylating enzyme and with the exosome, thereby promoting the rapid deadenylation and subsequent 3′→5′ degradation of these mRNAs. In this mechanism, the rate of mRNA degradation is uncoupled from the frequency of translation. Thus mRNAs containing the AUUUA sequence can be translated at high frequency, yet also degraded rapidly, allowing the encoded proteins to be expressed in short bursts.

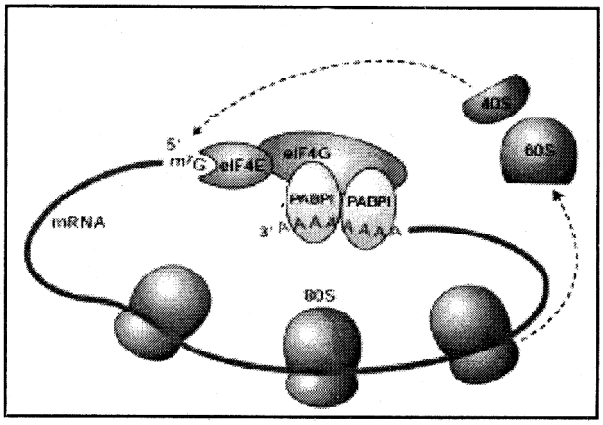


Fig. 6.6 : Initiation factors

Some mRNAs are degraded by decapping the mRNA before the deadenylation process is initiated. It appears that certain mRNA sequences make the cap sensitive to the decapping enzyme, but the precise mechanism is unclear.

In the other alternative pathway, mRNAs first are cleaved internally by endonucleases. The RNA-induced silencing complex (RISC) discussed earlier is an example of such an endonuclease. The fragments generated by internal cleavage then are degraded by exonucleases.

6.2.4 Regulation of mRNA translation and degradation

Ironresponse element-binding protein *(IRE-BP)* can be a classis example of protein that regulates the translation. Intracellular iron concentrations the protein in way that can regulate the translation of one mRNA and the degrade another. When the intracellular iron falls below the threshold level, IRE-BP proteins releases free irons in the system for the enzymes that require Fe to function. Again, when

the concentration of free iron increases within the system, the proteins bind to free Fe to prevent accumulation and toxicity.

Production of *ferritin*- an intracellular iron-binding protein is regulated by IRE-BP. The 5′ UTR of ferritin mRNA has a stem and loop structure containing a iron response element ***(IRE).*** The IRE-BP recognizes five specific bases in the IRE loop and the duplex nature of the stem. At low iron concentrations, IREBP is in active conformation and binds to the IREs. The bound IRE-BP blocks the 40S ribosomal subunit from scanning for the AUG start codon, thereby inhibiting translation initiation. The resulting decrease in ferritin means less iron is complexed with the ferritin and is therefore available to iron-requiring enzymes. At high iron concentrations, IRE-BP is in an inactive conformation that does not bind to the 5' IREs, so translation initiation can proceed. The newly synthesized ferritin then binds free iron ions, preventing their accumulation to harmful levels (Fig. 6.7a).

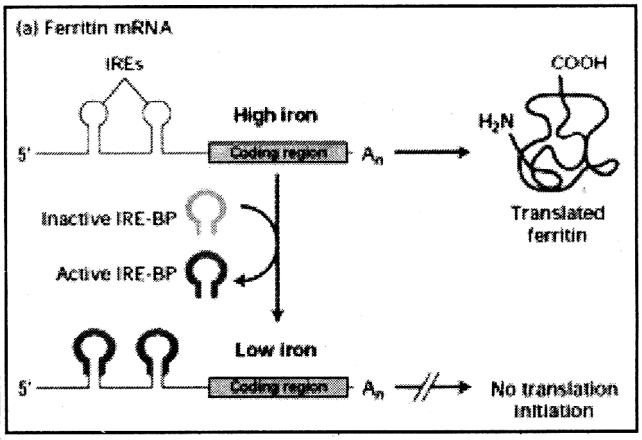


Fig. 6.7 a :

In vertebrates import of iron is also regulated. Ingested iron is carried through the circulation bound to a protein called *transferrin.* After binding to the transferrin receptor (TfR) in the plasma membrane, the transferrin-iron complex is brought into cells by receptor-mediated en-docytosis. The 3′-UTR of TfR mRNA contains IREs whose stems have AU′rich destabilizing sequences (Fig. 6.7b), At high iron concentrations, when the IRE-BP is in the inactive, nonbinding conformation, these AU-rich sequences are thought to promote degradation of TfR mRNA by the same

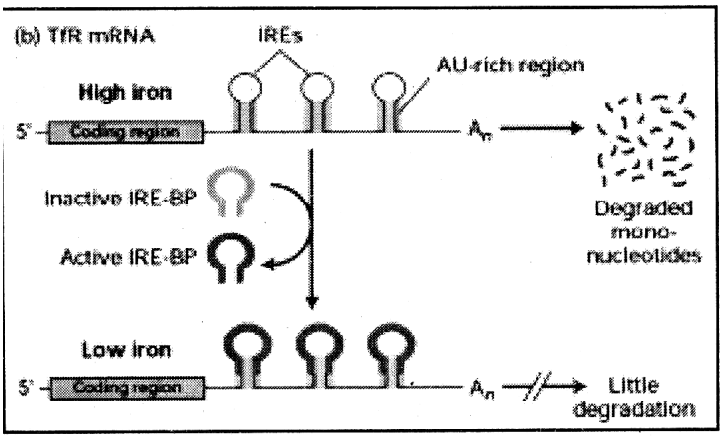


Fig. b.7b.

mechanism that leads to rapid degradation of other shortlived mRNAs, as described previously. The resulting decrease in production of the transferrin receptor quickly reduces iron import, thus protecting the cell. At low iron concentrations, however, IRE-BP can bind to the 3' IREs in TfR mRNA. The bound IRE-BP is thought to block recognition of the destabilizing AU-rich sequences by the proteins that would otherwise rapidly degrade the mRNAs. As a result, production of the transferrin receptor increases and more iron is brought into the cell.

6.2.5 Nonsense-mediated decay and other mRNA

Improperly processed mRNA cannot be translated and should be eliminated out of the system which otherwise can lead to production of an abnormal protein that interferes with functioning

of the normal protein. This effect is equivalent to dominant-negative mutations. Several mechanisms collectively termed **mRNA surveillance** help cells avoid the translation of improperly processed mRNA molecules.

Nonsense-mediated decay mechanism is another way how the cells get cleared of the wrongly processed mRNAs in which one or more exons have been skipped during splicing. During splicing, improper exon skipping sometime introduce stop codons. Nonsense-mediated decay results in the rapid degradation of mRNAs with stop codons that occur before the last splice junction in the mRNA. Analysis of yeast mutants suggests that some of the proteins in exon-junction complexes function in nonsensemediated decay.

6.3 Ribozymes

6.3.1 Concept of ribozymes

Until about 20 years ago, all known enzymes were proteins. But then it was discovered that some RNA molecules can act as enzymes; that is, catalyze covalent changes in the structure of substrates, RNA molecules that can catalyze a chemical reaction are called RIBOZYMES. The first ribozymes were discovered in the 1980s by Thomas R. Cech, who was studying RNA splicing in the ciliated protozoan *Tetrahymena therrnophila.* Subsequently, Sidney Altman, discovered bacterial RNase P complex. The ribozymes were found in the intron of an RNA **transcript,** which removed itself from the transcript and in the RNA component of the RNase P complex, which is involved in the maturation of pre-tRNAs. Many natural ribozymes catalyze either their own cleavage or the cleavage of other RNAs, but they have also been found to. catalyze the aminotransferase activity of the ribosome. Although most ribozymes are quite rare in the cell, their roles are sometimes essential to life.

Five classes of ribozymes have been described based on their unique characters in the sequences as well as three-dimensional structures (Bunnell, 1997). They are denoted as (1) the Tetrahymena group I intron, (2) RNase P, (3) the hammerhead ribozyme, (4) the hairpin ribozyme, and (5) the hepatitis delta virus ribozyme. They may catalyze self-cleavage as well as the cleavage of external substrates.

1. Group One Intron: The splicing reaction is self-contained; that is, the intron - with the help of associated proteins - splices itself out of the precursor RNA. the action is catalyzed by the RNA, only a single molecule of substrate is involved (unlike protein enzymes that repeatedly catalyze a reaction). **However, synthetic versions of Group I introns made in the laboratory can - *in vitro* - act repeatedly; that is, like true enzymes. The DNA of some Group I introns includes an open reading frame (ORF) that encodes a transposase-like protein that can make a copy of the intron and insert it elsewhere in the genome**

2. Ribonuclease P: All living things synthesize an enzyme - called Ribonuclease P - that cleaves the head (5′) end of the precursors of transfer RNA (tRNA) molecules. Ribonuclease P is a heterodimer containing a molecule of RNA and one protein. When the RNA is separated from the protein, the RNA retains its ability to catalyze the cleavage step (although less efficiently than the intact dimer), but the protein alone cannot do the job.

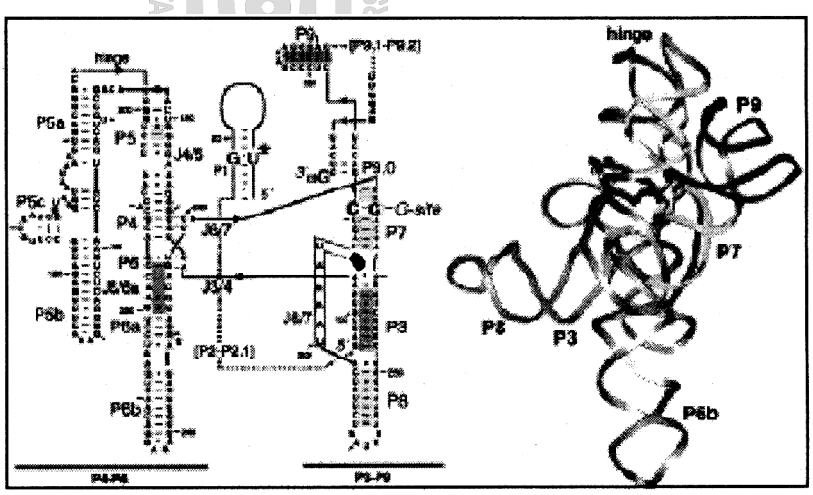


Fig. 6.8: Structure of Ribonuclease P

3. **Hammerhead ribozyme:** The name of hammerhead ribozyme is given by the similarity between its secondary structure and the shape of a hammerhead. They are the best understood subcategory of all ribozymes. As well as other

ribozymes, the hammerhead ribozyme is an antisense RNA. Some of the ribonucleotides within the sequence selectively form Watson-Crick base pairs with others to form a stem, while the rest stay in single stranded state called loop. These loops and stems can be predicted at the secondary structure level using conformational energy analysis, such as RNAdraw and mfold; and three dimensional structures were obtained mainly by X-ray crystallography

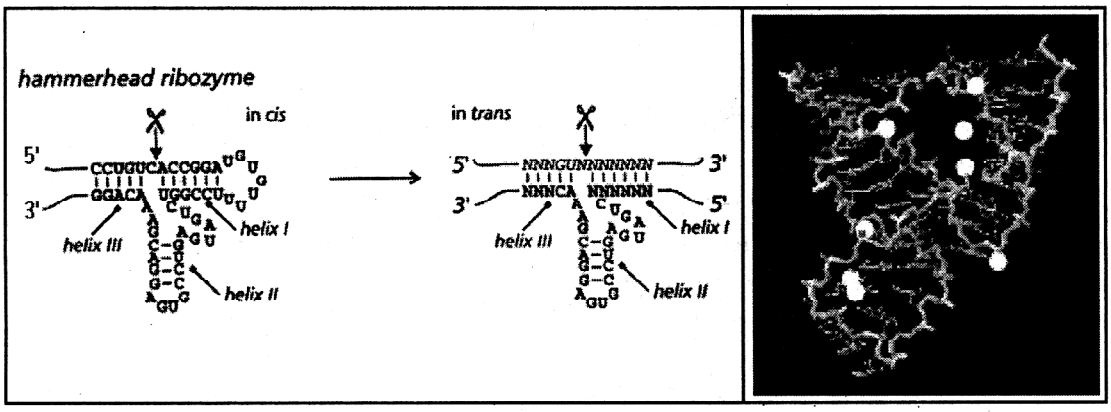


Fig. 6.9: Hammerheaded ribozyme

4. The hepatitis delta virus ribozyme : Some RNA viruses, such as the hepatitis delta virus, also include a ribozyme as part of their inherited RNA molecule. During replication of the viral RNA, long strands containing repeats of the RNA genome (viral genetic information) are synthesized. The ribozyme then cleaves the long multimeric molecules into pieces that contain one genome copy, and fits that RNA piece into a virus particle.

Since the discovery of ribozymes that exist in living organisms, there has been interest in the study of new synthetic ribozymes made in the laboratory. Ribozymes can be produced in the laboratory which, are capable of catalyzing their own synthesis under very specific conditions, such as an RNA polymerase ribozyme; although the polymerase activity is very limited. Such RNA polymerase ribozymes are able to add up to 14 nucleotides to a primer template in 24 hours until it is decomposed by hydrolysis of the phosphodiester bonds. Another artificially produced self cleaving RNAs ribozyme was produced by Tang and Breaker by in vitro selection of RNAs originating from random-sequence RNAs. Some of the synthetic ribozymes that were produced had novel structures, while some were similar to the naturally occurring hammerhead ribozyme.

The techniques used to discover synthetic ribozymes involve Darwinian evolution. This approach takes advantage of RNA's dual nature as both a catalyst and an informational polymer, making it easy for an investigator to produce vast

populations of RNA catalysts using polymerase enzymes. The ribozymes are mutated by reverse transcribing them with reverse transcriptase into various cDNA and amplified with mutagenic PCR. The selection parameters in these experiments often differ. One approach for selecting a ligase ribozyme involves using biotin tags, which are covalently linked to the substrate. If a molecule possesses the desired ligase activity, a streptavidin matrix can be used to recover the active molecules.

6.3.2 Ribozymes for human therapy

The application of ribozymes for gene therapy of autosomal dominant diseases has become popularized in recent years. Further this technology has widespread utility in the treatment of any disease, acquired or inherited, by inhibition of gene expression. The design of ribozymes is usually accomplished using computer assisted design programs, however they are not very useful in predicting the behavior of the ribozyme in the in vivo setting. To overcome this technical challenge, methods and strategy are being evolved to accurately assess the efficiency of ribozyme cleavage in vivo situations that significantly enhances the computer based design programs.

Already, a synthetic ribozyme that destroys the **mRNA** encoding a receptor of Vascular Endothelial Growth Factor (**VEGF**) is being readied for clinical trials. VEGF is a major stimulant of angiogenesis, and blocking its action may help starve cancers of their blood supply.

6.4 Antisense technology

6.4.1 Introduction

Over the last several decades, the knowledge of DNA/RNA physiology has been applied in a variety of ways. One of the more productive applications is the development of antisense technology. The basic idea is that if an oligonucleotide (a short RNA or DNA molecule complementary to a mRNA produced by a gene) can be introduced into a cell, it will specifically bind to its target mRNA through the exquisite specificity of complementary-based pairing—the same mechanism which guarantees the fidelity of DNA replication and of RNA transcription from the gene. This binding forms an RNA dimer in the cytoplasm and halts protein synthesis. This occurs because the mRNA no longer has access to the ribosome and because dimeric RNA is rapidly degraded in the cytoplasm by ribonuclease H. Therefore, the introduction of short chains of DNA complementary to mRNA will lead to a specific diminution, or blockage, of protein synthesis by a particular gene. In effect, the gene will be turned off.

The technical problems associated with the use of this technology are many. First, sufficient amounts of antisense oligonucleotide must be administered to the vicinity of target cells and, more importantly, must be taken up by those cells. Second, the antisense oligonucleotide should, ideally, have a long enough half-life within the cell to successfully impair mRNA translation into protein over a significant period of time. Finally, the oligonucleotide must also be nontoxic and sufficiently specific so as not to interfere with other cellular functions. In many applications, these hurdles have been overcome and antisense technology has developed into a productive branch of biology.

These technical challenges can be overcome in various ways depending on the specific application at hand. Oligonucleotides can be mixed with a variety of lipids to form complexes that are more easily incorporated by cell membranes, facilitating the entry of associated oligonucleotides into the cells. A number of other techniques have also been developed to facilitate the uptake of oligonucleotides by cells. Chemical modification of the antisense oligonucleotides can render them more stacble in cells and blood by increasing their resistance to ribonuclease digestion. Also, complementary DNAs or fragments of complementary DNAs can be incorporated in reverse sense in order to generate antisense RNA products in the host cell itself. This results in a long-term inhibition of the synthesis of the target protein.

6.4.2 Application of antisense technology *in vitro*

Antisense technology has been applied successfully in two general areas. The first is in fundamental research where the introduction of antisense oligonucleotides can help determine the role of a specific gene in a specific physiological process. For example, introduction of antisense oligonucleotides to inhibit the synthesis of angiotensinogen, the substrate from which cells make angiotensin II actually was found to stop the synthesis of angiotensinogen that resulted in a decline in cell growth. The introduction of angiotensin II to the cells restored this growth.

6.4.3 Therapeutic application of antisense technology

A second application of this technology, and one that is potentially of more immediate relevance to the practicing physician, is the use of this technology in therapy. In principle, antisense oligonucleotides complementary to viral RNAs can suppress a wide variety of viral infections; a tremendous amount of research is ongoing in this area. Similarly, antisense oligonucleotides directed towards the products of oncogenes can play a role in reducing the growth of cancer cells, and this lead is being hotly pursued.

Perhaps the most widely discussed application of antisense technology lies in its applications to gene therapy. In this case, a variety of vectors is used to introduce antisense-encoding genes into a large number of cells in a patient or animal to produce long-term inhibition of a protein. For example, in animal models the introduction of vectors encoding antisense angiotensin II receptor sequences results in long-term normotension in spontaneously hypertensive animals.

These are but a few of the possible applications of antisense technology. As familiarity with the relevant chemistry increases, it is likely that more effective oligonucleotides and gene vectors will be developed, thereby providing the ability to interfere at will with the translation of specific mRNAs.

6.4.4 Triplex antisense technology

In the face of all this progress, still newer technologies are being developed based on concepts related to antisense biology For example, it is known that oligonucleotides can, in certain instances, bind to duplex DNA molecules through an unusual kind of base pairing. In this triplex binding mode, oligonucleotides insert themselves into the major groove of the DNA double helix on a reasonably specific basis determined by the nucleotide sequence of the target DNA. This triplex technology provides the opportunity to reduce gene transcription jtself rather than to destroy mRNA once it is produced. Because the triplex oligonucleotides can be made to permanently alter the DNA after localizing to specific target sites, the technology actually has the potential to permanently silence genes.

6.4.5 RNA Inhibition

It has recently been shown that double-stranded RNA in the cytoplasm triggers an as yet poorly understood cascade of events leading to the suppression of the transcription of the gene producing the specific mRNA involved in the cytoplasmic RNA duplex. This could potentially lead to the development of new pharmacological agents.

Antisense technology is a formidable tool for investigating physiologic and pathologic processes. In addition, it is soon likely to become a mainstay of therapy, particularly in infectious diseases, with wider applications in the future as gene therapy techniques are developed further. Antisense Pharmaceuticals will soon be available for the routine care of patients and are expected to .prove to be effective, specific agents with favorable therapeutic profiles.

Unit 7 Recombination and Repair

Structure

7.1 Holiday junction in recombination

Recombination is a process or set of processes by which DNA molecules interact with one another to bring about a rearrangement of the genetic information or content in an organism. Although recombination as a process has been known for a hundred years, the real reason has not been appreciated until relatively recently. It now seems clear that recombination reactions exist to repair DNA. In bacteria, and probably also in eukaryotes, recombination mechanisms exist to repair stalled DNA replication forks. In the simplest sense, recombination is an exchange of both strands between two DNA molecules

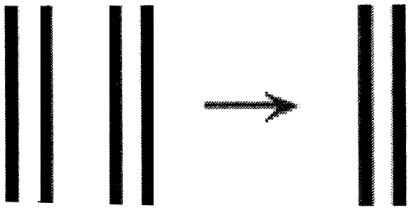


Fig. 7.1: Exchange of DNA strands between two homologous chromosomes.

In eukaryotic systems, you will be familiar with recombination as the process that is responsible for **crossing-over** during meiosis. Crossing-over has been well-documented genetically and is used to map the relative locations of genes on a chromosome.

7.2 The holiday model of genetic recombination

The model for recombination of two individual DNA strand was first proposed by Robin Holliday in 1964 and re-established by David Dressier and Huntington Potter in 1976 who demonstrated that the proposed physical intermediates existed.

In the most simplified explanation, two homologous DNA molecules align themselves which is followed by a nick at the same place on the two molecules as

shown in the figure below. This must happen in strands with the same polarity. The nicked strands then exchange themselves. The intermediate structure that is formed during such exchange is called a Holliday intermediate or Holliday structure. The shape of this intermediate in vivo is similar *to* that of the greek letter chi, hence this is also called a chi form.

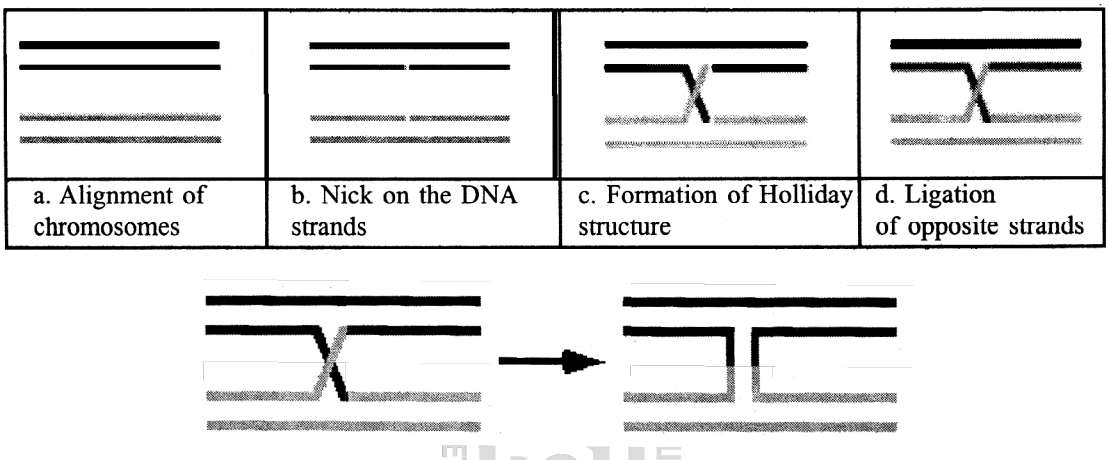


Fig. 7.2 : Initiation of recombination process and formation of Holiday structure

There are two ways in which the holiday structure can resolve itself to return back to its original conformation after the recombination process. If the same strands are cleaved a second time then the original two DNA molecules are generated (Fig. 7.3a). But, if the other strands are cleaved, then **recombinant** molecules are generated in a manner as shown in the figure below (Fig. 7.3b).

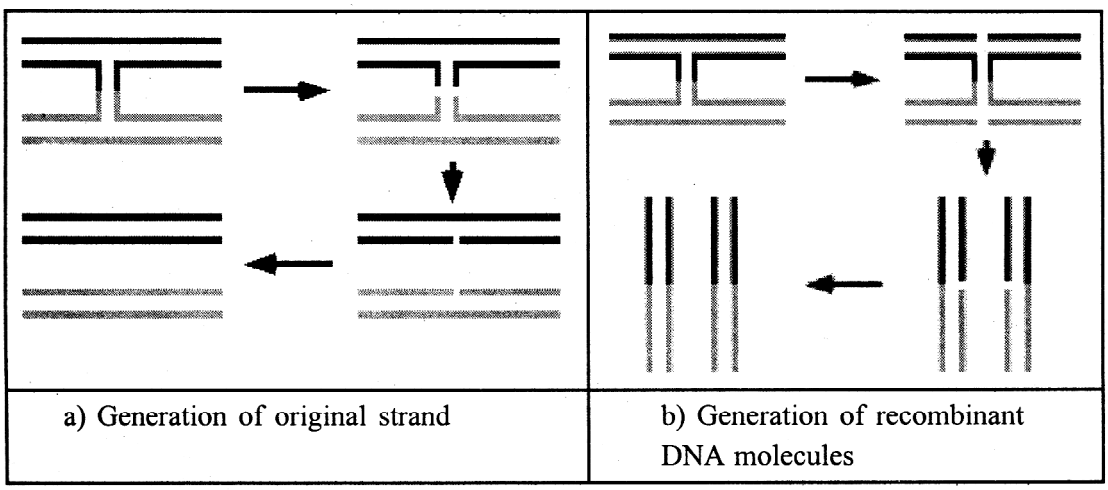


Fig. 7.3 : Two alternatives of recombination after the formation of holiday structure

In reality, a more complex mechanism operate to obtain different types of recom-bined strands as was observed in the Meselson-Weigle experiment where they located two different recombinant bacteriophages in a single plaque. These can be explained by modifying the above model slightly. As before, two homologous DNA molecules must be aligned and nicked at the same place. Following strand exchange the intermediate Holliday structure is formed. After the formation of the Holliday structure and ligation of the strands, the branch migrates, which can take place in either direction. The result is a physical transfer of part of one of the strands of one molecule with that of the other:

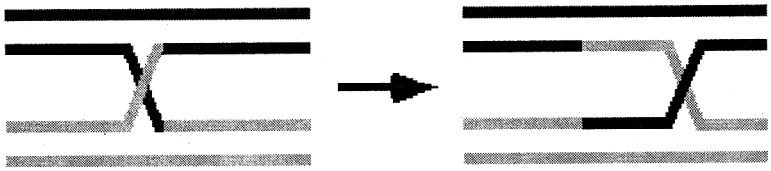


Fig. 7.4 : Two alternatives of recombination after the formation of holliday structure

For better understanding of the subsequent steps, one molecule is now rotated through 180° with respect to the other (Fig. 7.4).

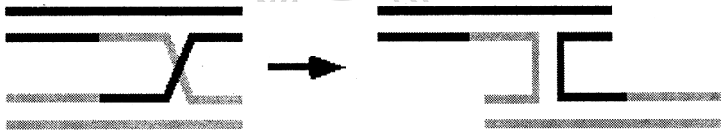
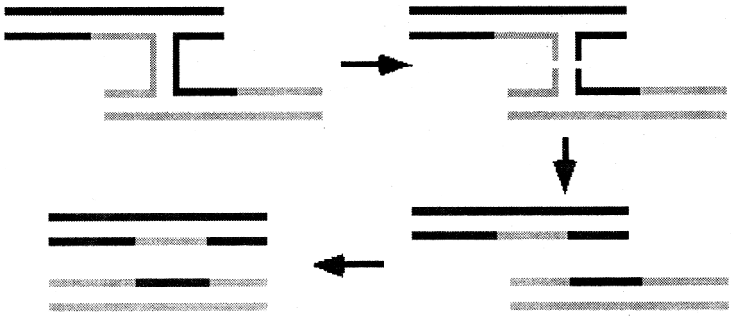


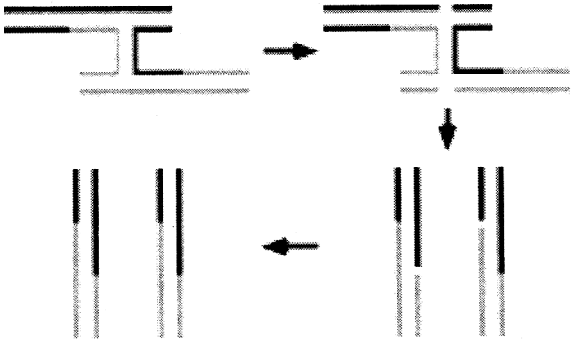
Fig. 7.5 : Rotation of one strand through 180° for better understanding

As described above, there are two possibilities of recombination which may result in two different consequences.

1. If the same strands are cleaved a second time then n on recombinant DNA molecules are generated but they each contain a region of heteroduplex DNA that spans the region of branch migration:



2. If the other strands are cleaved, then **recombinant** molecules are generated as before, however, each will also contain a region of **heteroduplex** DNA that spans the region of branch migration:



Double stranded nicks :

While the single-strand nick model provides a simple explanation for recombination, recent work in the yeast Sacchromyces cerevisiae shows that recombination is actually a response to Double-stranded breaks in the DNA molecule. The double-strand break model begins with the introduction of a double stranded break in one of the paired homologs (Fig. 7.6).

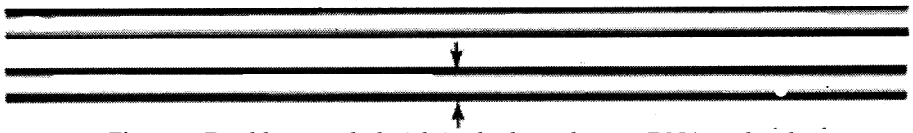


Fig. 7.6 : Double-stranded nick in the homologous DNA molecule

Introduction of the double-stranded break results in exonucleolytic degradation of the adjacent strand in a 5' to 3′ direction resulting in two single-stranded.whiskers (Fig. 7.7) which can now invade the paired homolog (via the action of a RecA like protein).

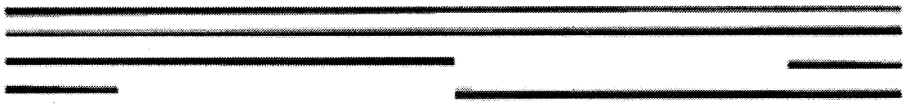


Fig. 7.7.: Exonucleolytic degradation of the adjacent strand in a 5' to 3′ direction

The DNA strand which was undisturbed serve as template and the invading 3' ends serve as primers for DNA synthesis and extends it self as shown below (Fig. 7.8).

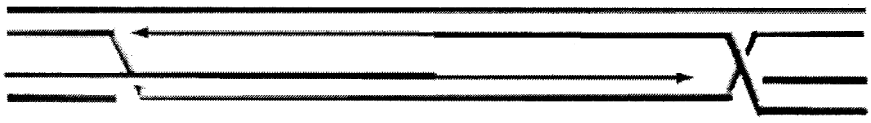


Fig. 7.8 : Extension of the overhands of the nicked strand

The 3' ends of the newly synthesized strand are then ligated with 5' ends of the degraded red homolog to form the **Double Holiday Junction** shown below (Fig. 7.9)

Fig. 7.9 : Formation of Double Holiday Junction

These Holiday Junctions are free to migrate as before to generate the heteroduplex regions as shown below (Fig. 7.10)

Fig. 7.10 : Free migration of the Holiday Junction

As before, each Holiday Junction can be resolved in two ways...

Resolution I: involves breaking and rejoining the two strands that cross between the two homologs.

Resolution II involves breaking and rejoining the two strands that do not cross between the two homologs

Since there are TWO Holiday Junctions, the exchange of flanking markers depends on how each Holiday Junction is resolved.

If both Holiday Junctions resolve via Resolution I, no exchange of flanking markers is observed as shown below. Note that the region between the two junctions involves heteroduplex which can be corrected to produce Gene Conversion (Fig. 7.11).

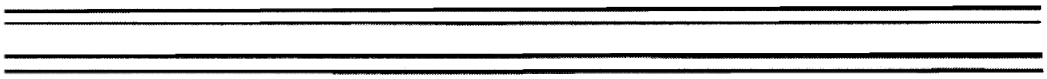


Fig. 7.11 : Formation of small region of heteroduplex if resolved through resolution 1

If Holiday Junction 1 undergoes Resolutation I and Holiday Junction 2 undergoes Resolutation II, or Holiday Junction 1 undergoes Resolution II and

Holiday Junction 2 undergoes Resolution I exchange of flanking markers is observed as shown below (Fig. 7.12).

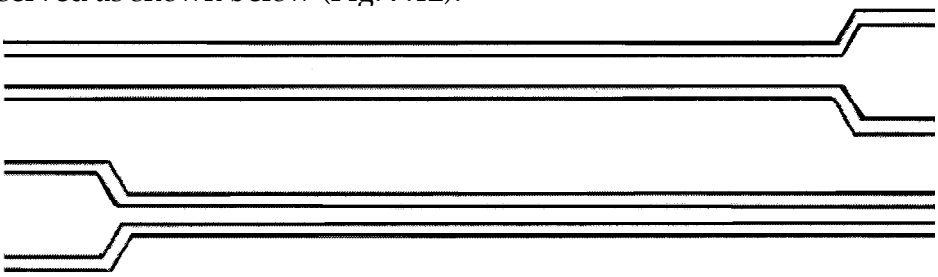


Fig. 7.12 : When both the Holliday junctions involve both type 1 and type 2 resolution, flanking markers are present

Finally, if both Holiday Junctions undergo Resolution II, no exchange of flankinj markers is observed.

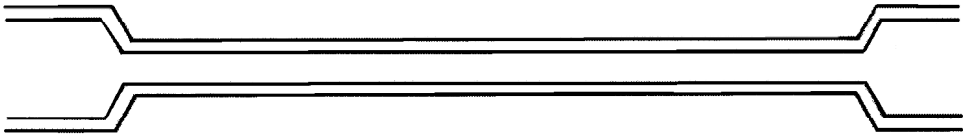


Fig. 7.13 : When both the Holliday junctions, undergo resolution 2 resolution no flanking markers is observed and the result is similar to fig. 11

7.3 Recombination proteins in *E. coli*

A number of the key proteins required for recombination. The structures and biochemical function of these proteins has been characterized which, is now helping us to understand details of the mechanism of recombination. The most important proteins are **RecA, RecBCD, RuvA, RuvB** and **RuvC.**

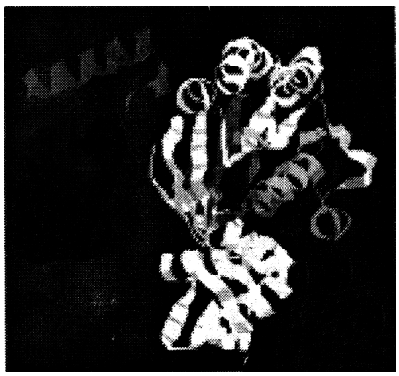


Fig. 7.14a :

RecA

The RecA protein is a multifunctional powerhouse! It has strand-exchange, ATPase and co-protease activities all packed into a compact 352 amino-acid, 38 kDa structure. It is required for all recombination pathways in *E. coli.* The RecA protein is a critical enzyme in this process, as it

catalyzes the pairing of ssDNA with complementary regions of dsDNA. The RecA monomers (Fig. 7.14a) first polymerize to form a helical filament around ssDNA, binding to a span of 4-6 nucleotides. Assembly of the nucleoprotein complex proceeds in a 5′→3′ direction. The complex is both fairly stable (half-life is 30 min) and is the active species that will promote strand exchange. RecA filament that forms is helical with a pitch of 82.7 Å and it consists of 6 monomer units per turn (Fig. 7.14b)

Fig. 7.14B :

During this process, RecA extends the ssDNA by 1.6 angstroms per axial base pair. Duplex DNA is then bound to the polymer. Bound dsDNA is partially unwound to facilitate base pairing between ssDNA and duplexed DNA. Once ssDNA has hybridized to a region of dsDNA, the duplexed DNA is further unwound to allow for branch migration. RecA has a binding site for ATP, the hydrolysis of which is required for release of the DNA strands from RecA filaments. ATP binding is also required for RecA-driven branch migration, but non-hydrolyzable analogs of ATP can be substituted for ATP in this process, suggesting that nucleotide binding alone can provide conformational changes in RecA filaments that promote branch migration. (Fig. 7.15)

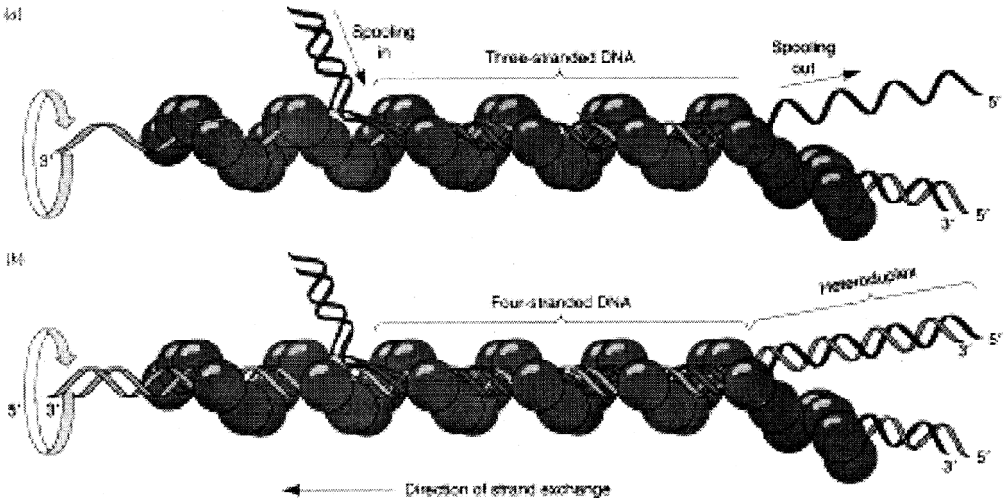


Fig. 7.15 : The strand exchange reaction probably involves the following steps: (a) RecA binds to the ssDNA partner, (b) The two molecules are aligned possible through the formation of a triple-stranded intermediate, (c) Displacement of one of the old strands. This requires concurrent migration of the RecA nucleoprotein filament along the molecule - which proceeds in one direction only (5′→3′) - and consequent winding/unwinding. ATP hydrolysis takes place during this step

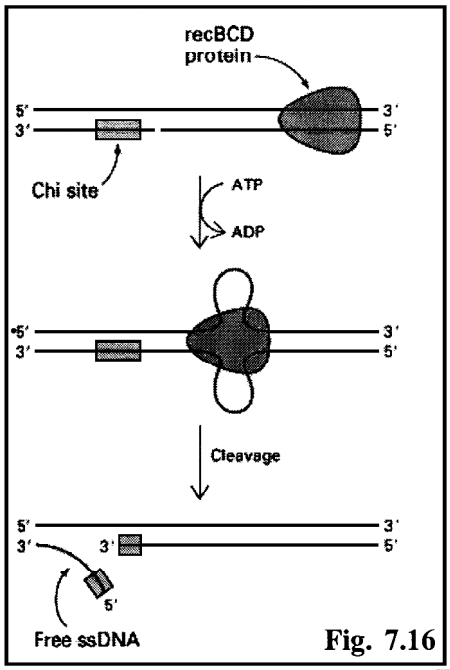


Fig. 7.16

RecBCD

The *recB, recC & recD* genes code for the three subunits of the RecBCD enzyme which has five activities: exonuclease V; a helicase activity; an endonuclease activity; an ATPase activity; and, an ssDNA exonuclease activity. The RecBCD helicase activity can unwind DNA faster than it rewinds. Thus as it travels along a DNA molecule, it can generate ssDNA loops (Fig. 7.16).

The RecBCD complex functions as a DNA exonuclease. It will bind to double-stranded breaks in DNA and degrade both strands simultaneously (Fig. 7.17). However, when **RecBCD** encounters a **Chi** sequence, its activity changes, the RecBC proteins act as a helicase to unwind the The RecD subunit is released and DNA in an ATP dependent reaction. This generates a ssDNA region that can serve (along with RecA) to initiate strand exchange and a recombination reaction

7.3.1 Proteins required for resolving holliday junctions in *E. coli*

RuvA

RuvA is a small protein whose function is to recognize a Holliday junction thereby assisting the **RuvB** helicase to promote branch migration.

The RuvA protein is 203 amino acids in length, but only 190 of them could be ssigned in the crystal structure. Most of the missing assignments represent amino

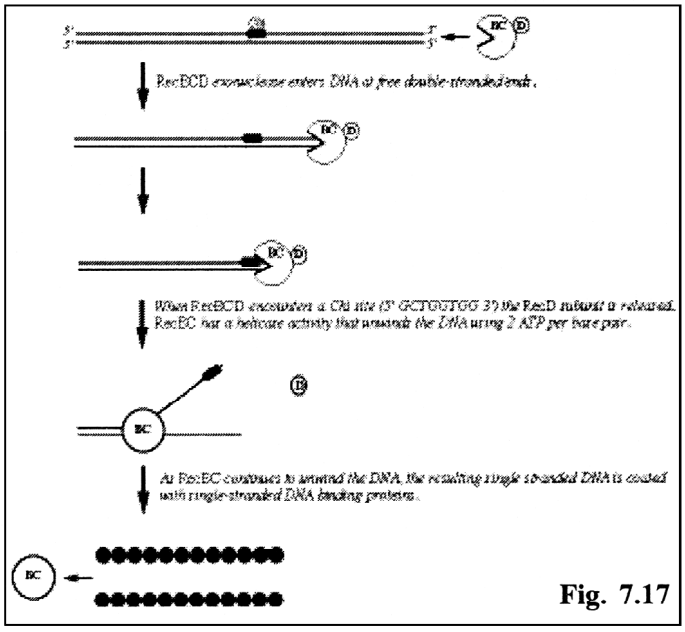
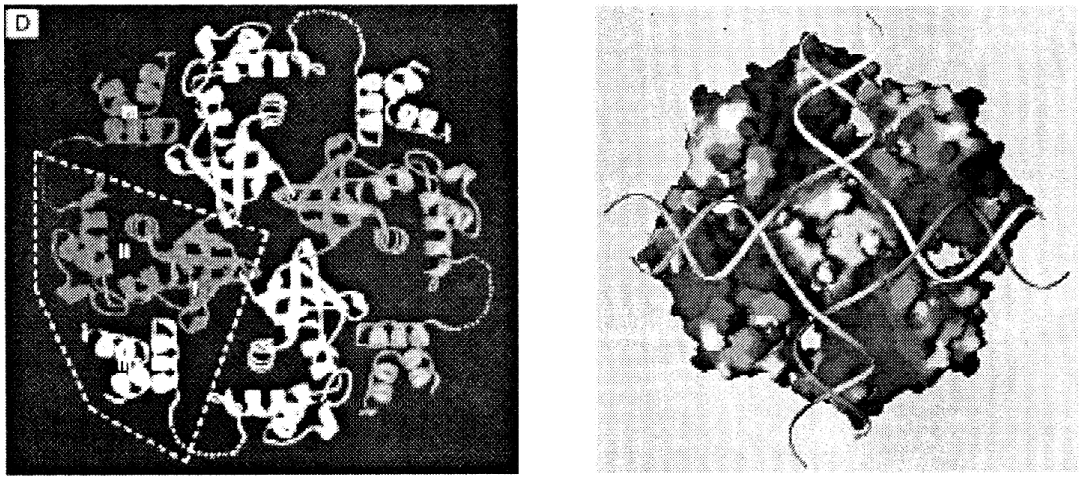


Fig. 7.17

acids in a flexible part of the protein. The crystal structure of the *E. coli* **RuvA** protein was solved at a resolution of 1.9 Å. The protein forms a tetramer in an unusual manner - though one that is ideally suited to its function.



the DNA strands at Holliday Junction

RuvB

The **RuvB** protein is a **helicase** that catalyzes branch migration of Holliday junctions. By itself it cannot bind to DNA efficiently. It functions in combination with **RuvA.** Like other helicases, RuvB functions as a hexamer; but, unlike other helicases, RuvB encloses double-stranded DNA not ssDNA.

Electron microscopy has shown that RuvB is a heptamer in solution and that it converts to a hexamer ring when it binds to DNA. Electron microscopy has also shown that the two hexamer rings of RuvB lie contacting RuvA on the two opposite sides of a RuvAB-Holliday junction complex (Fig. 7.19).

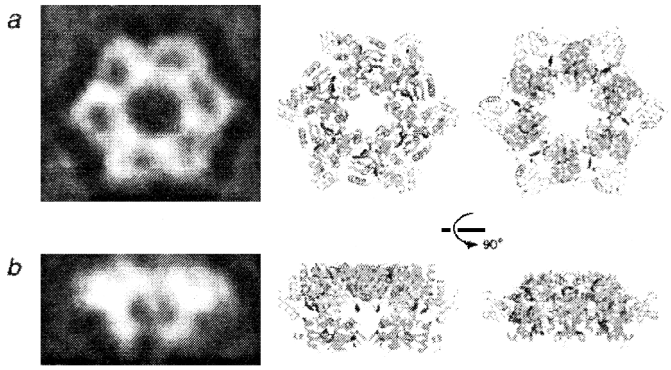


Fig. 7.19 : A hypothetical hexamer model of RuvB derived from the electron microscopy images

RuvC

The **RuvC** protein resolves the Holliday intermediate. It functions as a dimer to cleave two of the four strands that make up the central part of the intermediate. Since binding is symmetrical, **RuvC** can bind to the Holliday intermediate in two equally likely ways. Hence, Holliday intermediates can be resolved in two different, but equally likely, ways. The interaction of **RuvC** with Holliday junction is shownin Fig. 7.20.

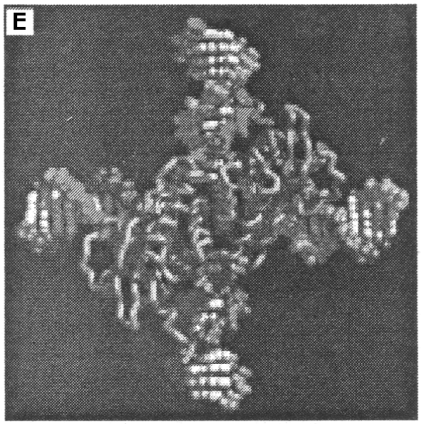


Fig. 7.20

RuvC does have some sequence specificity. It cleaves DNA at the 3′-side of thymidine, preferentially at the consensus 5′-$^{A}/_{T}$TT|$^{C}/_{G}$-3′ where '|' indicates the site of cleavage

7.4 DNA repair mechanisms

7.4.1 Introduction

Cells are always subjected to different types of stresses some of which cause alteration to the DNA molecules. For the genetic information encoded in the DNA is to remain uncorrupted, any chemical changes must be corrected. Surprisingly, both eukaryotic and prokaryotic cells have evolved an efficient DNA repair mechanism that function to maintain the integrity of the DNA molecules but allows subtle changes required to bring about variations and evolutionary changes. The recent publication of the human genome has already revealed 130 genes whose products participate in DNA repair. More will probably be identified soon.

7.4.2 Agents that damage DNA

- Certain wavelengths of **radiation**
  - ionizing radiation such as gamma rays and x-rays
  - **ultraviolet rays,** especially the UV-C rays (-260 nm) that are absorbed strongly by DNA but also the longer-wavelength UV-B that penetrates the ozone shield.
- Highly-reactive **oxygen radicals** produced during normal cellular respiration as well as by other biochemical pathways.
- Chemicals in the **environment**
  - many hydrocarbons, including some found in cigarette smoke
  - some plant and microbial products, e.g. the aflatoxins produced in moldy peanuts
- Chemicals used in **chemotherapy,** especially chemotherapy of cancers

7.4.3 Types of DNA damage

1. All four of the bases in DNA (A, T, C, G) can be covalently modified at various positions.

   o One of the most frequent is the loss of an amino group ("deamination") — resulting, for example, in a **C** being converted to a **U**.

2. During DNA replication DNA polymerase *(E. coli)* - inserts one incorrect nucleotide for every 10^5 nucleotides due to tautomeric "flickering" of the bases

3. **Mutations;**

   - Base Substitution by
     - **o point mutations**
       - *transitions*
         - pyrimidine-to-pyrimidine substitutions (T6C)
         - purine-to-purine substitutions (A6G) ?
       - *transversions*
         - pyrimidine-to-purine substitutions (T6G or A)
         - purine-to-pyrimidine substitutions (A6C or T)
   - Insertions or deletions of a small number of nucleotides
     - o **frame shift mutations**
       - recombination errors
       - transposons

Spontaneous rate of mutation at a given site on a chromosome is approximately 10^{-6} to 10^{-11} per round of replication. It is

- species and site specific
- "hot spots"
  - DNA microsatellites
    - repetitive sequences errors dues to "slippage" of DNA polymerase during replication
    - low frequency sites

4. **Mismatches** of the normal bases because of a failure of proofreading during DNA replication.
   - o example: incorporation of the pyrimidine **U** (normally found only in RNA) instead of **T.**

5. **Breaks** in the backbone : caused frequently by Ionizing radiation or by chemicals
   - o Can be limited to one of the two strands (a single-stranded break, **SSB)** or
   - o on **both strands** (a double-stranded break **(DSB).**

6. **Crosslinks** Covalent linkages can be formed between bases
   - o on the same DNA strand ("intrastrand") or
   - o on the opposite strand ("interstrand").

   Several chemotherapeutic drugs used against cancers crosslink DNA.

7.4.4 Repairing damaged bases

Damaged or inappropriate bases can be repaired by several mechanisms:

- **Direct chemical reversal** of the damage
- **Excision Repair,** in which the damaged base or bases are removed and then replaced with the correct ones in a localized burst of DNA synthesis. There are three modes of excision repair, each of which employs specialized sets of enzymes.
  1. **Base Excision Repair (BER)**
  2. **Nucleotide Excision Repair (NER)**
  3. **Mismatch Repair (MMR)**

7.4.5 Direct reversal of base damage

Perhaps the most frequent cause of point mutations in humans is the spontaneous addition of a methyl group (CH_3-) (an example of alkylation) to Cytosine followed by deamination to a Thyamine. Fortunately, most of these changes are repaired by enzymes, called glycosylases, that remove the mismatched T restoring the correct C. This is done without the need to break the DNA backbone (in contrast to the mechanisms of excision repair described below).

Some of the drugs used in cancer chemotherapy ("chemo") also damage DNA by alkylation. Some of the methyl groups can be removed by a protein encoded by our ***MGMT*** gene. However, the protein can only do it once, so the removal of each methyl group requires another molecule of protein.

This illustrates a problem with direct reversal mechanisms of DNA repair: they are quite wasteful. Each of the myriad types of chemical alterations to bases requires its own mechanism to correct. What the cell needs are more general mechanisms capable of correcting all sorts of chemical damage with a limited toolbox. This requirement is met by the mechanisms of **excision repair.**

7.4.6 Base excision repair (BER)

The steps and some key players:

1. Removal of the damaged base (estimated to occur some 20,000 times a day in each cell in our body!) by a DNA glycosylase. There exist at least 8 genes encoding different DNA glycosylases each enzyme responsible for identifying and removing a specific kind of base damage.
2. Removal of its deoxyribose phosphate in the backbone, producing a gap. We have two genes encoding enzymes with this function.
3. Replacement with the correct nucleotide. This relies on **DNA polymerase beta,** one of at least 11 DNA polymerases encoded by our genes.
4. Ligation of the break in the strand. Two enzymes are known that can do this; both require ATP to provide the needed energy.

7.4.7 Nucleotide excision repair (NER)

NER differs from BER in several ways.

- It uses different enzymes.
- Even though there may be only a single "bad" base to correct, its nucleotide is removed along with many other adjacent nucleotides; that is, NER removes a large "patch" around the damage.

The steps and some key players:

1. The damage is recognized by one or more protein factors that assemble at the location.
2. The DNA is unwound producing a "bubble". The enzyme system that does this is **Transcription Factor IIH, TFIIH,** (which also functions in normal transcription).

3. Cuts are made on both the 3' side and the 5' side of the damaged area so the tract containing the damage can be removed.
4. A fresh burst of DNA synthesis — using the intact (opposite) strand as a template — fills in the correct nucleotides. The DNA polymerases responsible are designated polymerase **delta** and **epsilon.**
5. **A DNA ligase** covalent binds the fresh piece into the backbone.

7.4.8 Transcription-coupled NER

Nucleotide-excision repair proceeds most rapidly

- in cells whose genes are being actively transcribed
- on the DNA strand that is serving as the template for transcription.

This enhancement of NER involves XPB, XPD, and several other gene products. The genes for two of them are designated ***CSA*** and ***CSB*** (mutations in them cause an inherited disorder called **Cockayne's syndrome).** The CSB product associates in the nucleus with **RNA polymerase II,** the enzyme responsible for synthesizing **messenger RNA** (mRNA), providing a molecular link between transcription and repair. One plausible scenario: If RNA polymerase II, tracking along the template (antisense) strand), encounters a damaged base, it can recruit other proteins, e.g., the CSA and CSB proteins, to make a quick fix before it moves on to complete transcription of the gene.

7.4.9 Mismatch repair (MMR)

Mismatch repair deals with correcting mismatches of the **normal bases;** that escapes proofreading mechanism of correction, and fails to maintain normal Watson-Crick base pairing (A•T, C•G)

It can enlist the aid of enzymes involved in both base-excision repair (BER) and nucleotide-excision repair (NER) as well as using enzymes specialized for this function.

- **Recognition** of a mismatch requires several different proteins including one encoded by ***MSH2.***

Cutting the mismatch out also requires several proteins, including one encoded by ***MLH1.***

Eg. In *E. coll*

- **MutS** dimer scans DNA for mismatches which distort DNA backbone
- Binds ATP at conformation changed sites where MutS bends the DNA
- ATP-MutS complex recruits **MutL & MutH.** ATP hydrolysis required for loading
- MutL activates **MutH endonuclease** activity nicks one strand near the mismatch
- DNA is unwound by **helicase UvrD** from the incision to the site of the mismatch
- **Exonuclease** digests displaced strand gap filled in by **Pol III** nick sealed by **ligase**

Mutations in either of these genes predispose the person to an inherited form of colon cancer. So these genes qualify as tumor suppressor genes. Synthesis of the repair patch is done by the same enzymes used in NER: **DNA polymerase delta** and **epsilon.** Cells also use the MMR system to enhance the fidelity of recombination; i.e., assure that only homologous regions of two DNA molecules pair up to crossover and recombine segments (e.g., in meiosis).

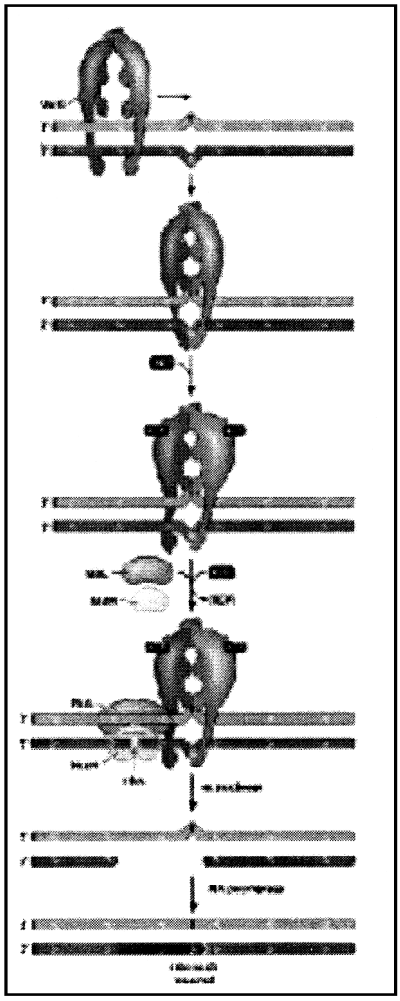


Fig. 7.21

Repairing Strand Breaks

Ionizing radiation and certain chemicals can produce both single-strand breaks **(SSBs)** and double-strand breaks **(DSBs)** in the DNA backbone.

Single-Strand Breaks (SSBs)

Breaks in a single strand of the DNA molecule are repaired using the same enzyme systems that are used in Base-Excision Repair (BER).

Double-Strand Breaks (DSBs)

There are two mechanisms by which the cell attempts to repair a complete break in a DNA molecule:

- **Direct joining** of the broken ends. This requires proteins that recognize and bind to the exposed ends and bring them together for ligating. They

would prefer to see some complementary nucleotides but can proceed without them so this type of joining is also called **Nonhomologous End-Joining (NHEJ).**

- Errors in direct joining may be a cause of the various **translocations** that are associated with cancers.
- Examples:
  - Burkitt's lymphoma
  - the Philadelphia chromosome in chronic myelogenous leukemia (CML)
  - B-cell leukemia

Meiosis I with the alignment of homologous sequences provides a mechanism for repairing damaged DNA; that is, mutations, in fact, many biologists feel that the main function of sex is to provide this mechanism for maintaining the integrity of the genome. However, most of the genes on the human Y chromosome have no counterpart on the X chromosome, and thus cannot benefit from this repair mechanism. They seem to solve this problem by having multiple copies of the same gene ––oriented in opposite directions. Looping the intervening DNA brings the duplicates together and allowing repair by homologous recombination.

7.4.10 Gene conversion

If the sequence used as a template for repairing a gene by homologous recombination differs slightly from the gene needing repair; that is, is an allele, the repaired gene will acquire the donor sequence. This **nonreciprocal transfer** of genetic information is called gene conversion. Gene conversion during meiosis alters the normal mendelian ratios. Normally, meiosis in a heterozygous **(A,a)** parent will produce gametes or spores in a 1:1 ratio; e.g., 50% **A;** 50% **a.** However, if gene conversion has occurred, other ratios will appear. If, for example, an **A** allele donates its sequence as it repairs a damaged **a** allele, the repaired gene will become **A**, and the ratio will be 75% **A;** 25% **a.**

Human diseases caused by loss of DNA repair systems: DNA repair systems play a major role in normal human health. Two examples of human pathology caused by loss of repair systems are described below.

Xeroderma pigmentosum: Xeroderma pigmentosum is a human genetic disease (or more correctly, a family of closely related genetic diseases), in which there is abnormal sensitivity to ultraviolet radiation. A number of different genes appear to be involved. Some patients exhibit defects in photoreactivation, but loss of

excision repair is more common. Mutations in at least seven different genes coding for proteins involved in excision repair can cause afflicted individuals to exhibit the symptoms of xeroderma pigmentosum.

Cockayne syndrome: This human genetic disease, whose symptoms include mental retardation, dwarfism, and premature aging, appears to be primarily due to failure of transcription-repair coupling.

Unit 8 Molecular Mapping of Genome

Structure

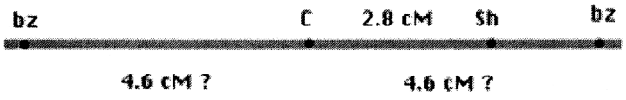
8.1 Genetic and physical maps

With the rebirth of genetics in the 20th century, it quickly became apparent that Mendel's second rule does not apply to many matings of dihybrids. In many cases, two alleles inherited from one parent show a strong tendency to stay together as do those from the other parent. This phenomenon is called **linkage.** The linkage phenomenon was utilized to construct gene maps.

8.1.1 Genetic map

A genetic map is a representation of the genes on a chromosome arrayed in linear order with distances between loci expressed as percent recombination **(map units, centimorgans).** It is also called a **linkage map.** The relative position of genes on a chromosome in a Genetic map is determined by counting the phenotypes from a cross. One **map unit =** one **centimorgan (cM)** = 1% recombination between loci. The farther apart two loci are, the more likely that a crossover will occur between them. Conversely, if two loci are close together, a crossover is less likely to occur between them. A recombination rate of 50% corresponds to independent assortment. Therefore, only distances less than 50 map units can be measured directly. Greater distances can be constructed by adding up distances between closer loci. They have been prepared for many eukaryotes, including corn, Drosophila, the mouse, and the tomato. A genetic map of chromosome 9 (the one that carries the **C, Sh,** and **bz** loci) of the corn plant (**Zea mays)** is shown on the right.

For example, test crossing a com plant that is dihybrid for the C,c (yellow/ colourless kernels) alleles and the alleles for bronze color (**Bz, bz**) produces 4.6% recombinants. So these two loci are 4.6 cM apart. However, is the **bz** locus on the same side of **c** as **sh** (smooth/shrunken) or is it on the other side? consider a cross between two different strains of corn (maize)



The answer can be found by test crossing the dihybrid **Shsh, Bzbz.** If the percentage of recombinants is less than 4.6%, then bz must be on the same side of locus c as locus **sh.** If greater than 4.6%, it must be on the other side. In fact, the recombination frequency is less than 1.8%, telling us that the actual order of loci is c — **sh** — **bz**

But there are certain difficulties with such genetic maps. Mapping by linkage analysis is best done with loci that are relatively close together; that is, within a few centimorgans of each other. Why? Because as the distance between two loci increases, the probability of a second crossover occurring between them also increases and therefore interpretation becomes difficult. There are other problems with preparing genetic maps of chromosomes.

- The probability of a crossover is not uniform along the entire length of the chromosome.
  - o Crossing over is inhibited in some regions (e.g., near the centromere).
  - o Some regions are "hot spots" for recombination (for reasons that are not clear). Approximately 80% of genetic recombination in humans is confined to just one-quarter of our genome.
- In humans, the frequency of recombination of loci on most chromosomes is higher in females than in males. Therefore, genetic maps of female chromosomes are longer than those for males.

8.1.2 Chromosome maps

The chromosome map (or cytogenetic map) is based on the karyotype of an organism. For example: All mouse chromosomes are defined at the cytogenetic

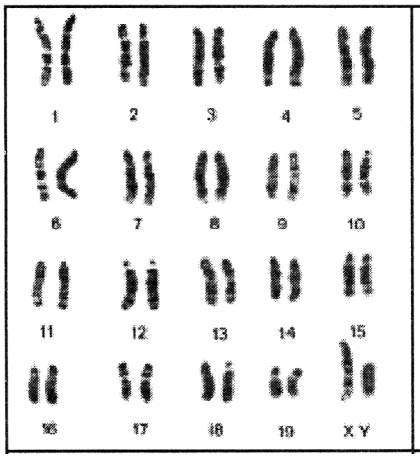


Fig. 8.1 : Karyotype of a diploid set of normal raetaphase chromosomes of *Mus musculus.*

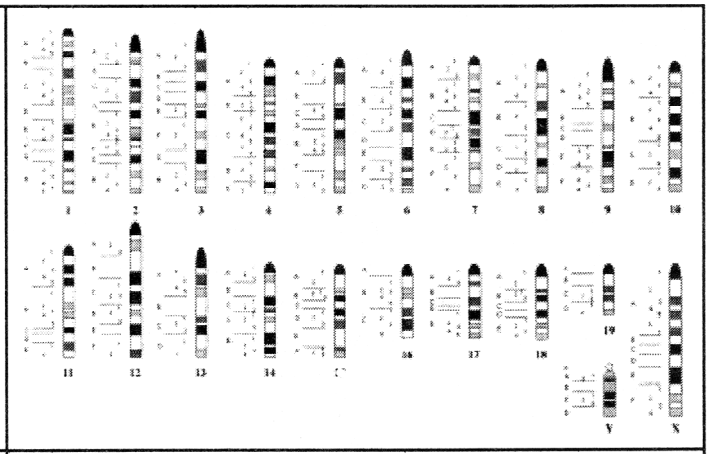


Fig 8.2 : Mouse chromosome idiograms. (Giemsa banding patterns associated with each chromosome in a normal karyotype

level according to their size and banding pattern (Fig. 8.1) and ultimately, all chromosomal assignments are made by direct cytogenetic analysis or by linkage to a locus that has previously been mapped in this way. Chromosomal map positions are indicated with the use of band names (Fig. 8.2).

Today, several different approaches, with different levels of resolution, can be used to generate chromosome maps.

1. Human/mouse cell hybrids tend to lose human chromosomes at random, leading eventually to hybrid cell lines that have one or a few human chromosomes. If a gene is always present or absent when one particular chromosome is present or absent, it can be concluded that the gene is on that chromosome.
2. Fluorescent or radioactive probes that bind to a particular gene can be observed microscopically and can be used to localize the gene on a metaphase spread.
3. Chromosomes from cells in metaphase can be sorted with high-speed electronic sorters. One can make preparations of a particular chromosome. If a particular gene can be shown to be in the preparation, it must be located on that chromosome.

8.1.3 Physical map

All physical maps are based on the direct analysis of DNA. Physical distances between and within loci are measured in basepairs (bp). Physical maps are arbitrarily divided into short range and long range. 1. Short range mapping is commonly pursued over distances ranging up to 30 kb. In very approximate terms, this is the average size of a gene and it is also the average size of cloned inserts obtained from cosmid-based genomic libraries. Cloned regions of this size can be easily mapped to high resolution with restriction enzymes or by sequencing.

2. Direct long-range physical mapping can be accomplished over megabase-sized regions with the use of rare-cutting restriction enzymes together with various methods of gel electrophoresis referred to generically as *pulsed field gel electrophoresis* or PFGE, which allow the separation and sizing of DNA fragments of 6 mb or more in length.

3. Long-range mapping can also be performed with clones obtained from large insert genomic libraries such as those based on the yeast artificial chromosome (YAC) cloning vectors, since regions within these clones can be readily isolated for further analysis.

8.1.4 Connections between maps

In theory, linkage, chromosomal, and physical maps should all provide the same information on chromosomal assignment and the order of loci (Fig. 8.3). However, the relative *distances* that are measured within each map can be quite different. Only the physical map can provide an accurate description of the actual length of DNA that separates loci from each other. This is not to say that the other two types of maps are inaccurate. Rather, each represents a version of the physical map that has been modulated according to a different parameter. Cytogenetic distances are modulated by the relative packing of the DNA molecule into different chromosomal regions. Linkage distances are modulated by the variable propensity of different DNA regions to take part in recombination events.

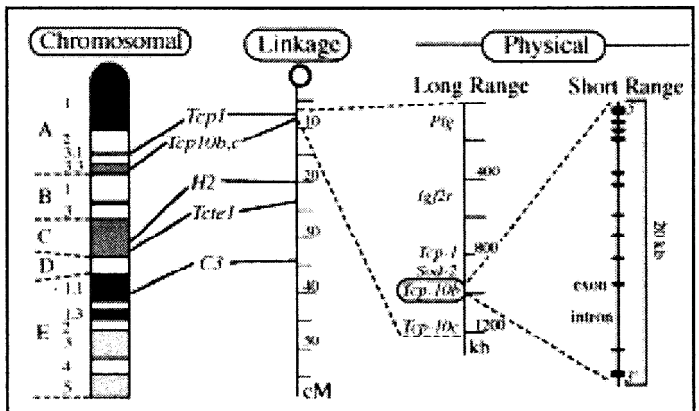


Fig. 8.3: Comparison of chromosomal, linkage, and physical maps. The three types of maps are shown for loci on Chr 17 of mouse.

In practice, genetic maps of the mouse are often an amalgamation of chromosomal, linkage, and physical maps, but at the time of this writing, it is still the case that classical recombination studies provide the great bulk of data incorporated into such integrated maps. Thus, the primary metric used to chart interlocus distances has been the centimorgan. However, it seems reasonable to predict that, within the next five years, the megabase will overtake the centimorgan as the unit for measurement along the chromosome

8.1.5 Gene mapping has important applications

A. It is useful for locating the position of genes on chromosomes, e.g. if two genes are closely linked and the position of one is known, then the other must also be nearby.

B. It is useful in estimating genetic risk, e.g. if a gene cannot be tested directly, then variation at a closely linked locus may indicate the presence or absence of a detrimental allele.

C. A major goal of the Human Genome Project is the mapping of all human

genes (as well as those of mice, Drosophila, *Caenorabditis elegans* (a nematode), *Arabidopsis thaliana* (a small plant), yeast, and the bacterium *Escherichia coli.* As of 1999, yeast, *E. coli, C. elegans,and* about a dozen other bacteria have been completely sequenced and all their genes identified, although the functions of most are unknown. Major progress has been made in mapping human genes, and a "rough draft" of the human genome is anticipated by 2000. Understanding of function of the many newly discovered human genes is being greatly aided by the studies of yeast, which has many genes similar to those of humans.

8.2 Gene cloning

8.2.1 Procedure of cloning

Gene cloning involves separating a specific gene or DNA segment from a larger chromosome, attaching it to a small molecule of carrier DNA, and then replicating this modified DNA thousands or millions of times through both an increase in cell number and the creation of multiple copies of the cloned DNA in each cell. The result is selective amplification of a particular gene or DNA segment. Cloning of DNA before the 1970 was a difficult tusk. Unlike a protein, a gene does not exist as a discrete entity in cells, but rather as a small region of a much larger DNA molecule. But with the discovery of restriction nucleases and other enzymes like ligase, poly-merases, the tusk of cloning targeted gene became rather easy. The key development that made recombinant DNA technology possible was the discovery in the late 1960s of restriction enzymes or called restriction endonucleases. The specialty of the endo-nucleases is that it recognizes and makes double-stranded cuts in the sugar-phosphate backbone of DNA molecules at specific nucleotide sequences. These enzymes are produced naturally by bacteria, where they are used in defense against viruses. In bacteria, restriction enzymes recognize particular sequences in viral DNA and then cut it up. A bacterium protects its own DNA from a restriction enzyme by modifying the recognition sequence, usually by adding methyl groups to its DNA. The endonucleases can be used to cut purified DNA at targeted sites for partially purifying the gene from a mixture.

Cloning of any DNA fragment essentially involves four steps: fragmentation, ligation, transfection, and screening/selection. Although these steps are invariable among cloning procedures a number of alternative routes can be selected, these are summarised as a 'cloning strategy'. The protocol for isolation of genes can be broken down into several steps.

Step 1.

At first, DNA need to be isolated from the desired cell, purified to ensure there is no protein contamination in the DNA sample. Presence of protein or any undesired chemicals can effect the subsequent steps of cloning.

Step 2.

DNA sample having the gene of interest requires to be segmented into suitable size. Preparation of DNA fragments for cloning is frequently achieved by means of PCR, but it may also be accomplished by restriction enzyme digestion, DNA soni-cation and fractionation by agarose gel electrophoresis. Restriction enzyme digestion, for example with EcoRl will produce small fragments of DNA with sticky ends. If there is no EcoRl site in the gene of interest, then a the DNA fragment carrying the gene will have flanking region with sticky ends.

Step 3.

The next step involves the insertion of the DNA fragment into a vector. A vector is usually a plasmid- circular DNA, which is linearised by means of restriction enzymes. The DNA fragments and the linearised vector is incubated together under appropriate condition in presence of the enzyme DNA ligase. Sticky ends or the single stranded DNA overhangs allow annealing of the DNA fragment with the vector sequence. Sticky ends may also be produced by chemical modification and attachment of adapter molecules. 'Sticky ends' allow for both higher efficiency transformations and directional insertion of the insert into the vector, thus minimising the need for subsequent screening.

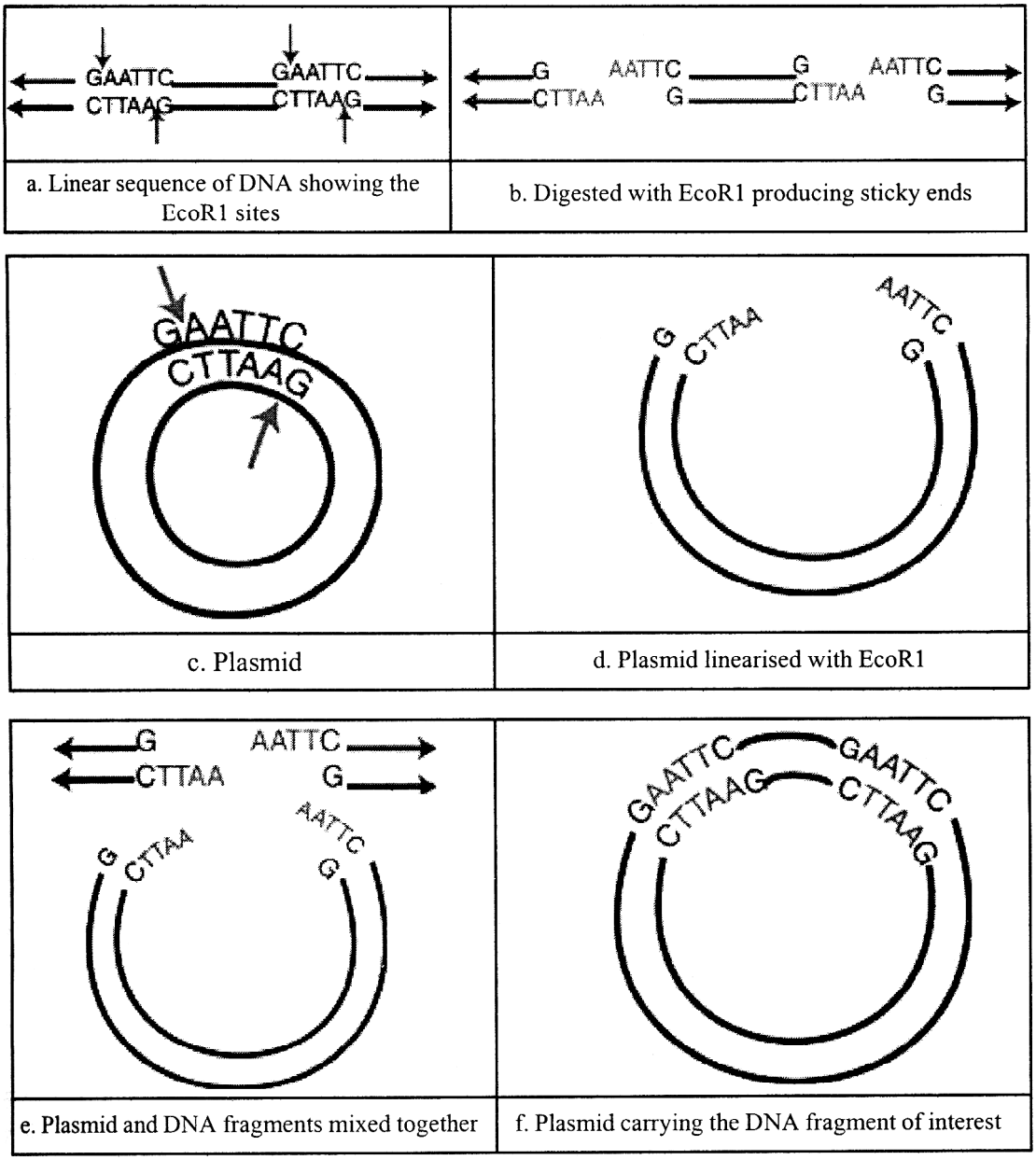
Step 4.

After the ligation precedure, the vectors with successful inserts are first identified and then is transfected into host cells. A number of alternative techniques are available, such as chemical sensitization of cells, electroporation and biolistics. Chemical sensitization of cells is frequently employed since this does not require specialised equipment and provides relatively high transformation efficiencies. Electroporation is employed when extremely high transformation efficiencies are required, as in very inefficient cloning strategies. Biolistics are mainly used in plant cell transformations, where the cell wall is a major obstacle in DNA uptake by cells.

Step 5.

Finally, the transfected cells are cultured and screened to identify the clone carrying the gene of interest. If the starting material is a PCR product, the screening step is not required. Successfully transformed carrying the gene of interest is identified primarily by hybridization technique. The required cells will be those that have been successfully transfected with the vector construct containing the

desired insertion sequence in the required orientation. Modern cloning vectors include selectable antibiotic resistance selection marker, which allow only cells in which the vector has been transfected, too grow. Additionally, the cloning vectors may contain colour selection markers which provide blue/white screening



a. Linear sequence of DNA showing the EcoR1 sites

b. Digested with EcoR1 producing sticky ends

c. Plasmid

d. Plasmid linearised with EcoR1

e. Plasmid and DNA fragments mixed together

f. Plasmid carrying the DNA fragment of interest

Fig. 8.4 : Schematic representation of the procedure of cloning

(á-factor complementation) on X-gal medium. Nevertheless, these selection steps do not absolutely guarantee that the DNA insert is present in the cells obtained. Further investigation of the resulting colonies is required to confirm that cloning was successful

8.2.2 cDNA Libraries

A collection of clones containing all the DNA fragments from one source is called a DNA library. For example, we might isolate genomic DNA from human cells, break it into fragments, and clone all of them in bacterial cells or phages. The set of bacterial colonies or phages containing these fragments is a human genomic library, containing all the DNA sequences found in the human genome. A genomic library must contain a large number of clones to ensure that all DNA sequences in the genome are represented in the library. A library of the human genome formed by using cosmids, each carrying a random DNA fragment from 35,000 to 44,000 bp long, would require about 350,000 cosmid clones to provide a 99% chance that every sequence is included in the library.

An alternative to creating a genomic library is to create a library consisting only of those DNA sequences that are transcribed into mRNA (called a cDNA library because all the DNA in this library is *complementary* to mRNA). Much of eukaryotic DNA consists of repetitive (and other DNA) sequences that are not transcribed into mRNA and such sequences are not represented in a cDNA library.

One of the most challenging tasks in molecular biology is the synthesis and cloning of cDNA. A complex series of enzymatic steps is involved in copying mRNA into double-stranded cDNA and subsequently preparing the termini for vector ligation. Many approaches have been used to generate cDNA libraries. Most cDNA molecules produced will lack a few nucleotides corresponding to the 5' end of the mRNA because second-strand replacement only proceeds from 3'-OH RNA primers. However, since all eukaryotic mRNA molecules appear to have 5' noncoding leader sequences, which commonly range from 40-80 nucleotides, it is likely that the vast majority of double-stranded cDNA will contain all of the coding sequences present in the initial cellular mRNA molecules.

However, cDNA library has two advantages. First, it is enriched with fragments from actively transcribed genes. Second, introns do not interrupt the cloned sequences and therefore easy to clone in bacterial system for expression.The disadvantage of a cDNA library is that it contains only sequences that are present in mature mRNA. Introns and any other sequences that influence transcription procedure are not present; sequences, such as promoters and enhancers, etc. It is

also important to note that the cDNA library represents only those gene sequences expressed in the tissue from which the RNA was isolated which is again dependent on the frequency of mRNA trans cribbed in the tissue. In contrast, almost all genes are present at the same frequency in a genomic DNA library.

8.2.3 Choice of primers

The classical method of cDNA synthesis uses the Oligo(dT) Primer to prime first-strand synthesis (Fig. 8.5). This method is suitable in most cases where poly(A)+ RNA of high quality can be prepared from the cell line or tissue of interest. Random Hexameric Primers (hexadeoxyribonucleotides) provide an alternative procedure (Fig. 8.6) by which first-strand cDNA synthesis is initiated from internal sites within the mRNA molecule. Random Primers can be used to prime cDNA synthesis from mRNA molecules that do not possess a poly(A)+ tail or for RNA isolated from prokaryotic sources. Random Primers also provide a scheme by which cDNA can be synthesized representing mRNA with strong 5′ secondary structure.

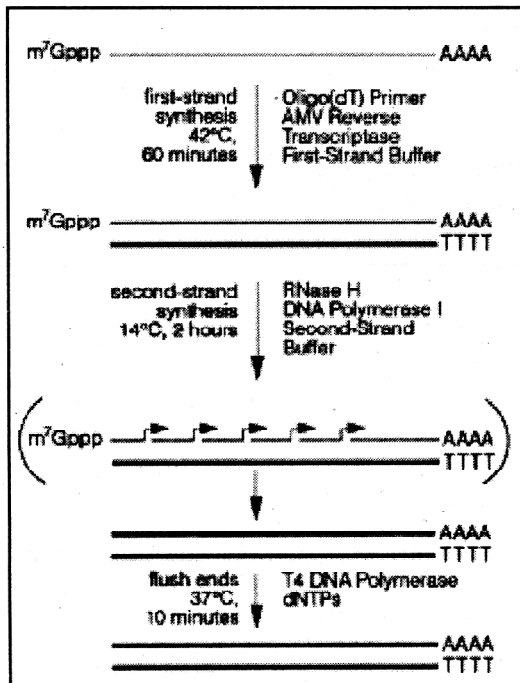


Fig.8.5: cDNA synthesis using the Oligo(dT) Primer

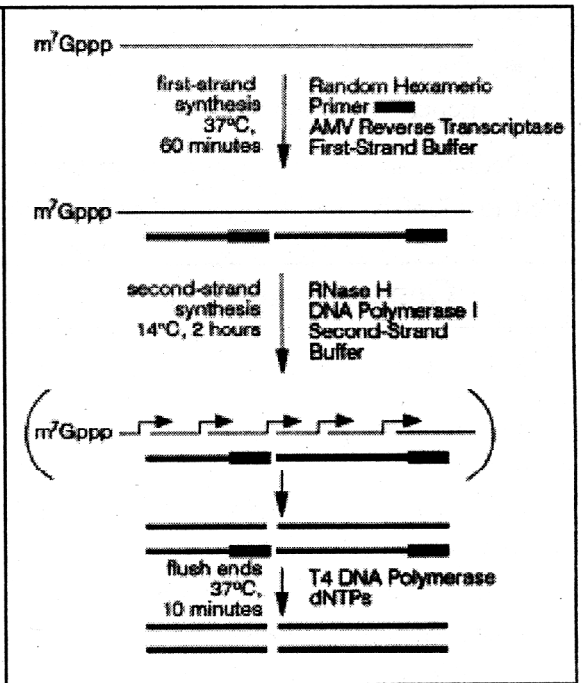


Fig. 8.6: cDNA synthesis using Random Hexameric Primers.

The mRNA molecules are then copied into cDNA by reverse transcription. Reverse transcriptase, an enzyme isolated from retroviruses, synthesizes single-stranded complementary DNA from the RNA template by adding DNA nucleotides to the 3'-OH group of the primer (Fig. 1 & 2)

The resulting RNA-DNA hybrid molecule is then converted into a double-stranded cDNA molecule by one of several methods. One common method is to treat the RNA-DNA hybrid with RNase to partly digest the RNA strand. Partial digestion leaves gaps in the RNA-DNA hybrid, allowing DNA polymerase to synthesize a second DNA strand by using the short undigested RNA pieces as primers and the first DNA strand as a template. DNA polymerase eventually displaces all the RNA fragments, replacing them with DNA nucleotides, and nicks n the sugar-phosphate backbone are sealed by DNA ligase. The cDNA thus synthesized are then subjected to end modifications and cloned into vectors for further analysis.

8.2.5 Libraries

A "library" is a convenient storage mechanism of genetic information.

- They are typically either "genomic" or "cDNA" (i.e. mRNA in DNA form) genetic information.

Deduced genetic sequences from corresponding polypeptide information can be used to identify specific genetic information within a library.

8.3 Genomic analysis

8.3.1 Introduction

The field of genomics comprises focuses on the content and organization of genomic information, and attempts to understand the function of information in genomes. Genomics is trying to look at all the genes as a dynamic system, over time, and determine how they interact and influence biological pathways and physiology, in a much more global sense. Genetics looks at single genes, one at a time, as a snapshot. Genetics is much more linear than genomics, complicated but not as complex as genomics.

The genetic information possessed by each individual is termed its **genotype** and can refer to the entirety of its genetic information or a part of it. The set of characteristics expressed by an individual is termed as **phenotype**. When two

individuals possessing the same genotype also have the same phenotype, regardless of the environmental conditions in which they exist, the character expressed is termed as **genetic trait.** Determination of the mode of inheritance of the genetic trait is called inheritance analysis. Inheritance analysis of a trait or phenotypic character is a **genetic marker.** If each phenotype can be unambiguously assigned to exactly one genotype, then the genetic marker defines a **gene marker.**

The advent of recombinant DNA technology in population genetics in the mid-1980's, gradually led to the development of **DNA markers.** However, the repertoire of genetic markers available for population genetic studies continues to increase enormously and is still relevant of genetic analysis. DNA marker analysis, though costly have some added advantage over the classical genetic markers and is rapidly replacing the old system of genetic analysis.

Technological advancement in the DNA sequencing technique contributed to development of the Genomics as a subject. Agenomic sequence is, by itself, of limited use. Functional genomics is, in essence, probing genome sequences for meaning— identifying genes, identify the unique sequences which can serve as DNA markers, recognizing their organization, and understanding their function etc. The goals of functional genomics include identifying all the RNA molecules transcribed from a genome (the **transcriptome**) and all the proteins encoded by the genome (the **proteome**).' Functional genomics exploits both bioinformatics and laboratory-based experimental approaches in its search to define the function of DNA sequences.

8.3.2 Predicting function from sequence

Several methods for identifying genes and assessing their functions have been discussed earlier. The methods include in situ hybridization, DNA footprinting, experimental mutagenesis, and the use of transgenic animals and knockouts. These methods can provide important information about the locations and functions of genetic information and can be applied to study to large numbers of genes simultaneously.

However, this biochemical approach to understanding gene function is both time consuming and expensive. A major goal of functional genomics has been to develop computational methods that allow gene function to be identified from DNA sequence alone, bypassing the laborious process of isolating and characterizing individual proteins.

8.3.3 Search for homology

One computational method for determining gene function is to conduct a homology search, which relies on comparing DNA and protein sequences from the same and different organisms. Databases containing sequences of genes and proteins for a wide array of organisms are available in gene banks which are in public domain and can be accessed for homology searches. Powerful computer programs have been developed for scanning these databases to look for particular sequences. A commonly used homology search program is BLAST used to align sequences from different or same species. If a function is known for one of these sequences, that function may provide information about the function of the newly discovered protein. Similar programs are also available that can analyze two sequences and predict the evolutionary relationship from which phylogenetic trees can be established.

8.3.4 Drug designing

Computer programs are also available that can detect single nucleotide polymorphism. One. can use the information from the analysis to predict the causes of various diseases and also can use the information to design drugs for treatment of diseases.

8.3.5 Gene expression and microarrays

The advent and developmet of the **Microarray** technique made it possible to study hundred and thousands of gene at the same time. Many important clues about gene function come from knowing when and where the genes are expressed. The microarray technique enables us to get such clues. **Microarrays** rely on nucleic acid hybridization in which a known DNA fragment is used as a probe to find complementary sequences (Fig. 8.7). In a microarray, numerous known DNA fragments are fixed to a solid support in an orderly pattern or array, usually as a series of dots. These DNA fragments (the probes) usually correspond to known genes. When the microarray has been constructed, mRNA, DNA, or cDNA isolated from experimental cells is labeled with fluorescent nucleotides and applied to the array. Any of the DNA or RNA molecules that are complementary to probes on the array will hybridize with them and emit luorescence, which can be detected by an automated scanner. An array containing tens of thousands of probes can be applied to a glass slide or silicon wafer just a few square centimeters in size.

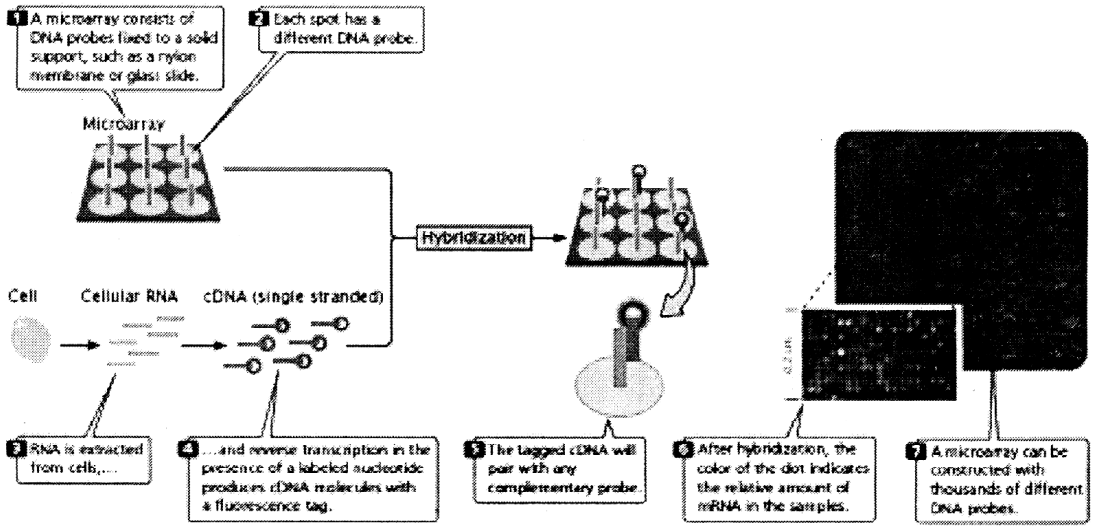


Fig 8.7 : Microarrays, used to detect the expression of many genes

For example, the experimental cells are stimulated from which mRNA is isolated and converted into cDNA labeled with red fluorescent nucleotides. mRNA from control cells is converted into cDNA and labeled with green fluorescent nucleotides. The labeled cDNAs are mixed and hybridized to the DNA chip, which contains DNA probes from different genes from the same organism. Hybridization of the red (experimental) and green (control) cDNAs is proportional to the relative amounts of mRNA in the samples. The fluorescence of each spot is assessed with microscopic scanning and appears as a single color. Red indicates the overexpression of a gene in the experimental cells relative to that in the control cells (more red-labeled cDNA hybridizes), whereas green indicates the underexpression of a gene in the experimental cells relative to that in the control cells (more green-labeled cDNA hybridizes). Yellow indicates equal expression in experimental and control cells (equal hybridization of red- and greenlabeled cDNAs), and no color indicates no expression in either experimental or control cells (Fig. 8.8).

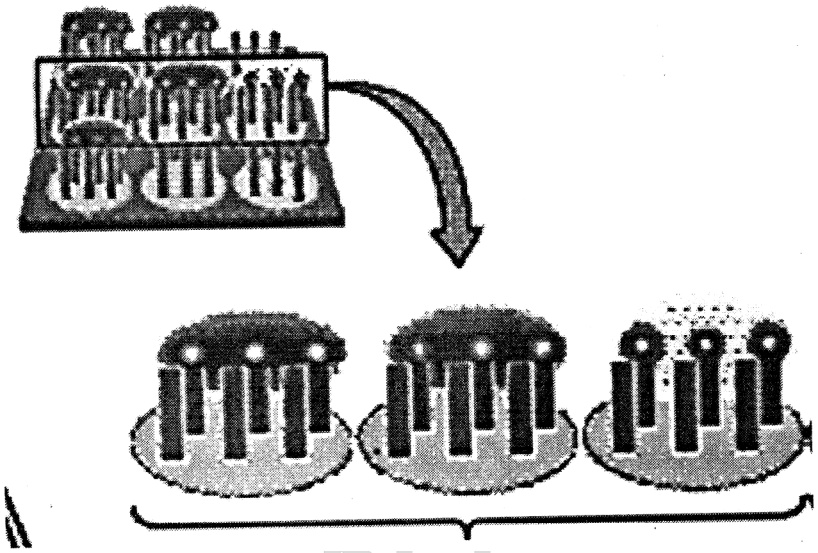


Fig: 8.8: Emission of fluorescence depends on the nature of hybridization

Microarrays allow the expression of thousands of genes to be monitored simultaneously, enabling scientists to study which genes are active in particular tissues. They can also be used to investigate how gene expression changes in the course of biological processes such as development or disease progression.

8.4 Restriction fragment length polymorphisms (RFLP)

A restriction fragment length polymorph of alternative alleles associated with restriction fragments that differ in size from each other. RFLPs are visualized by digesting DNA from different individuals with a restriction enzyme, followed by gel elec-trophoresis to separate fragments according to size, then blotting and hybridization to a labeled probe that identifies the locus under investigation. An RFLP is demonstrated whenever the Southern blot pattern obtained with one individual is different from the one obtained with another individual (Fig. 8.9). In this example, DNA samples from five individual mice were

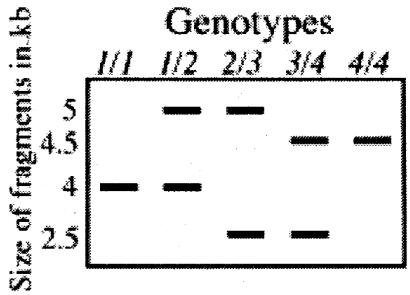


Fig. 8.9: Demonstration of RFLPs by Southern blot analysis.5 different Southern blot genotypes detected with a single probe is shown.

digested with the same enzyme, and after electrophoresis were probed with the same clone of a single-copy DNA sequence. The five patterns detected are all different from each other and are representative of five different genotypes. The first lane and the last lane are homozygous for the genes while the remaining individuals are heterozygous.

RPLPs were the predominant form of DNA variation used for linkage analysis until the advent of PCR. Even now, in the PCR age, RFLPs provide a convenient means for turning an uncharacterized DNA clone into a reagent for the detection of a genetic marker. The main advantage of RFLP analysis over PCR-based protocols is that no prior sequence information, nor oligonucleotide synthesis, is required. Furthermore, in some cases, it may not be feasible to develop a PCR protocol to detect a particular form of allelic variation. Nevertheless, if and when a PCR assay for typing a particular locus is developed, it will almost certainly be preferable over RFLP analysis. The detection of a RFLP does not provide information as to the mechanism by which it was created. Moreover, from RFLP data, it is also not possible to predict how the individuals differ from each other at the molecular level.

Attempts to identify RFLPs between different inbred strains of mice often meet with limited success even after, testing with large numbers of enzymes. In one study, RFLPs were identified at only 30% of the single copy loci tested with 22 different restriction enzymes. Furthermore, when RFLPs are identified, they are almost always 'di-allelic binary systems — the insertion, deletion, or restriction site change is either present or absent. Unfortunately, di-allelic loci can only be mapped in crosses where the two parental chromosomes carry the two alternative alleles. Thus, even if a RFLP is identified between two inbred strains of mice, there is no guarantee that another pair of strains will also happen to- carry alternative alleles. As a consequence, only a subset of the RFLP markers developed for analysis of one cross between traditional mouse strains will be of use for mapping in a cross between any other pair of inbred strains.

8.4.2 Choice of restriction enzymes to use for RFLP detection

With so many restriction enzymes available, how does one decide which ones are the best to use in the search for RFLPs? Obviously, cost is an important consideration. Another consideration is whether the enzyme is optimally active with genomic DNA obtained from animal tissues. However, a critical consideration is the rate at which RFLPs can be detected based on the enzyme that is chosen.

A systematic study of RFLP detection between B6 and *M. spretus* DNA subsequent to digestion with one often different enzymes has been reported. One

hundred and ten anonymous DNA sequences of less than 4 kb in length were used as probes. The highest rate of RFLP detection — 63% — was observed with DNA digested with *Taql.* The second highest rate — 56% — was observed with *Mspl.* In decreasing order of effectiveness were the enzymes *BainRl* (50%), *Xbal* (47%), *Pstl* (44%), *Bg*II (41%), *Hind* III (39%), *Pvull* (38%) *Rsa* I (38%), and *Eco* RI (33%). It is ironic that of the ten enzymes tested, the one most commonly used in molecular biological research — *EcoRI* — was the worst one, by a long shot, at detecting polymorphisms.

8.4.3 Minisatellites: variable number tandem repeat loci

In contrast to traditional RFLPs caused by basepair changes in restriction sites, a special class of RFLP loci present in all mammalian genomes is highly polymorphic with very large numbers of alleles. These "hypervariable" loci were first exploited in a general way by Jeffreys (1985) and his colleagues for genetic mapping in humans.

Hypervariable RFLP loci of this special class are known by a number of different names including variable number tandem repeat (VNTR) loci and *minisatellites,* which is the more commonly used term today. Minisatellites are composed of unit sequences that range from 10 to 40 bp in length and are tandemly repeated from tens to thousands of times. Although various functions have been suggested for mini satellite loci as a class, none of these has withstood the test of further analysis. Rather, it appears most likely that minisatellite loci evolve in a neutral manner through expansion and contraction caused by unequal crossing over between out-of-register repeat units. Recombination events of this type will yield reciprocal products which both represent new alleles with a change in the *number* of repeat units.

The frequency with which new alleles are created at minisatellite loci — on the order of 10^{-3} per locus per gamete — is much greater than the classical mutation rate of 10^{-5} to 10^{-6}. This leads to a much higher level of polymorphism between unrelated individuals within a population. At the same time, one change in a thousand gametes is low enough so as to not interfere with the ability to follow minisatellite alleles in classical breeding studies.

Length polymorphisms at minisatellite loci are most simply detected by digestion of genomic DNA samples with a restriction enzyme that does not cut within the minisatellite itself but does cut within closely flanking sequences. As with all other RFLP analyses, the restriction digests are fractionated by gel electrophoresis, blotted and hybridized to probes derived from the polymorphic locus. However, unlike traditional point mutation RFLPs, minisatellites are caused

by, and reflect, changes in the actual size of the locus itself.

The best restriction enzymes to use for minisatellite analysis are those with 4 bp recognition sites such as *Hae*III, *Hin*fI or *Sau*3A; it is likely that one of these enzymes will not cut within the relatively short minisatellite unit sequence, but will cut within several hundred basepairs of flanking sequence on both sides. Standard 1% agarose gels with maximal separation in the 1-4 kb range are usually best for the resolution of minisatellite bands; however, conditions can be optimized for each minisatellite system under analysis.

8.4.4 RAPD

RAPD stands for random amplification of polymorphic DNA. It is a type of PCR reaction, where random segments of genomic DNA are amplified with single primer of arbitrary nucleotide sequence and which are able to differentiate between genetically distinct individuals, although not necessarily in a reproducible way. By resolving the resulting patterns, a semi-unique profile can be gleaned from a RAPD reaction (Fig. 8.10).

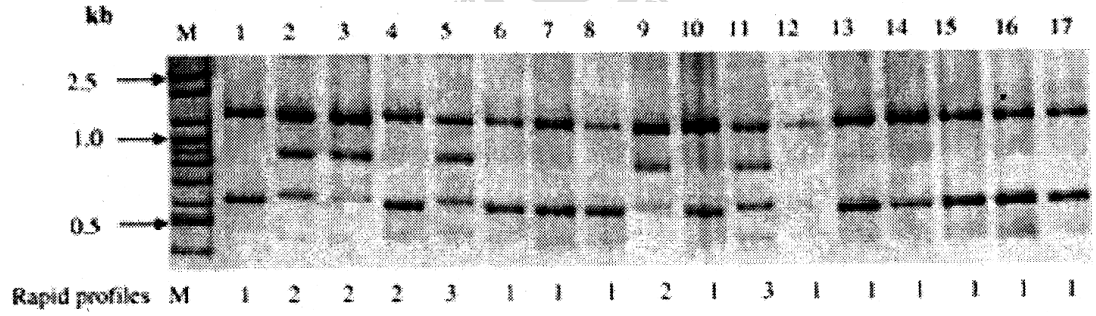


Fig. 8.10: Randomly amplified genomic DNA with short primers

No knowledge of the DNA sequence for the targeted gene is required, as the primers will bind **somewhere** in the sequence, but it is not certain exactly where. This makes the method popular for comparing the DNA of biological systems that have not had the attention of the scientific community, or in a system in which relatively few DNA sequences are compared (it is not suitable for forming a DNA databank). Due to the fact that it relies on a large, intact DNA template sequence, it has some limitations in the use of degraded DNA samples. Its resolving power is much lower than targeted, species specific DNA comparison methods, such as

short tandem repeats. In recent years, RAPD is used to characterize, and trace, the phylogeny of diverse plant and animal species.

Limitations of RAPD

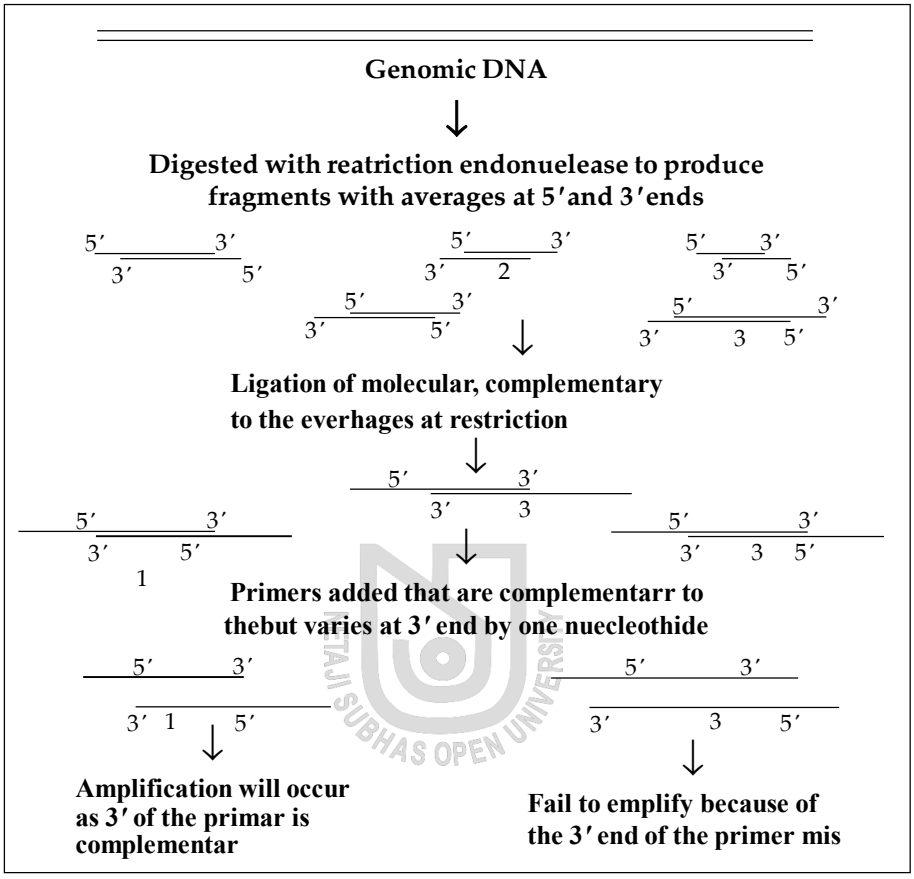
- Nearly all RAPD markers are dominant, i.e. it is not possible to distinguish whether a DNA segment is amplified from a locus that is heterozygous (1 copy) or homozygous (2 copies). Co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, are detected only rarely.

- PCR is an enzymatic reaction, therefore the quality and concentration of template DNA, concentrations of PCR components, and the PCR cycling conditions may greatly influence the outcome. Thus, the RAPD technique is notoriously laboratory dependent and needs carefully developed laboratory protocols to be reproducible.

- Mismatches, between the primer and' the template may result in the total absence of PCR product as well as in a merely decreased amount of the product. Thus, the RAPD results can be difficult to interpret.

8.4.5 AFLP

AFLP stands for Amplified Fragment Length Polymorphism which is a hybrid of RFLP and RAPD techniques. Genomic DNA is cut with restriction enzymes, as in RFLP. Typically, two different restriction enzymes are used. The idea is to produce a large number of fragments. Some of the fragments are selectively amplified with PCR using "random" primers, as in RAPD. The primers are not really random, however. Specific oligonucleotide "adapters" (these are complementary to the restriction sites) of 25-30 bp are ligated to the restricted DNA fragments. The primers are complementary to these adapters. However, the primers vary at their 3'-end, such that they will amplify only a subset of the restricted DNA fragments. Typically 50-100 restriction fragments are amplified and detected on denaturing polyacrylamide gels. AFLPs have typically been used to study variation among individuals of a species, most commonly for producing genetic maps (and in trying to find genes responsible for certain traits). They have received



limited attention as tools in systematics, perhaps because the method is relatively labor intensive when compared with other methods. The power of AFLP is based upon the molecular genetic .variations that exist between closely related species, varieties, orcultivars. These variations in DNA sequence are exploited by the AFLP technology such that "fingerprints" of particular genotypes can be routinely generated. These "fingerprints" are simply RFLPs visualized by selective PCR amplification of DNA restriction fragments.