Chapter 2

Review of Literature
2. Review of Literature

2.1 Evolution of sex chromosomes

The existence of specialized sex determining chromosomes poses a range of interesting and basic questions, such as: why there exists different sex (male and female) in an organism? What is the evolutionary significance of having male and female sexes? How is it that an entire chromosome, originally containing many other genes unrelated to sex, could become subverted to a singular developmental programme? Why has the functional specialization of the Y or W sex-determining chromosome not lead to chromosomal imbalance and consequent lethality? Attempts to address some of these questions have not received fully satisfactory answers.

Sex chromosome evolution and chromosomal sex determination are the two distinct biological phenomenon, which are linked evolutionarily. Suppression of recombination between X and Y and Z and W might have resulted in differentiation of the two chromosomes Y and W. Various studies on the evolution of sex chromosomes have indicated that sex chromosome heteromorphism must have evolved independently on many occasions. This suggests that conditions predisposing the specialization of sex chromosomes are wide spread amongst eukaryotes and the probability of their evolving is relatively high (Ohno, 1967; Charlesworth, 1978).

The Z and W chromosomes of birds and snakes can be compared with mammalian X and Y chromosomes respectively and they share many functional features. Both the chromosomes in birds and snakes are metacentric, pair during meiosis only at the ends of the short arm of the two chromosomes; therefore a small pseudoautosomal region exists (Stefos and Arrighi, 1971; Ellegren, 2000). W chromosome remains highly condensed in all somatic tissues and forms a condensed W-chromatin body. It synthesises its DNA asynchronously with the Z chromosome and autosomes (Raychaudhari et al., 1971; Raychaudhari and Singh, 1972). W chromosome decondenses extensively in developing oocytes (Singh et al., 1976). Similarly, the Y chromosome in mammals remains condensed in all somatic cells but decondenses extensively in germ cells. Such a relatively discrete condensation cycle obviously
would be consistent with the activity of one or few genes and its timing is appropriate for early embryonic development including, presumably, sex determination. Inactivated mammalian X chromosome becomes reactivated in oocytes (Gartler *et al.*, 1972; Lyon, 1974).

Why does such suppression of recombination between heteromorphic sex chromosome pair occur? Bull (1983) suggested that most likely reason is that if there are sex linked alleles with opposite effect on fitness in the two sexes, there is a selective advantage in reducing recombination between the sex determining locus and these genes. According to Rice (1987), suppression of recombination is selectively favoured by the accumulation of sexually antagonistic genes on the sex chromosome. Also recombination deficiency can result in the accumulation of mutation and other genetic changes in the heteromorphic chromosomes. Lack of crossing over in XY heterogametic males would cause accumulation of deleterious recessive mutations on Y chromosome leading gradually to its progressive heteromorphism, a phenomenon termed as Muller's Ratchet.

There is an alternative view that finds experimental reasons to conclude a specific role of *Bkm* sequences in heteromorphic degenerative evolution of Y or W sex-chromosome and where the entire chromosome seems to function like a super gene (Singh, 1995). According to Jones (1983) W-evolution is triggered by mutations which cause the control of the process of mitotic W chromosome to be modulated by the sex-determinants. According to this model, the cycle of somatic inactivation and germ cell activation of the W chromosome signifies the cycle of sex gene expression. In species where W has continued to specialize, the sex determiner assumed control over the center involved in chromosome condensation. A spreading effect or read through from the sex gene was such that when it turned off, the chromosome also condensed; the sex determiner thus fortuitously but effectively had hijacked the entire chromosome. In primitive species, the model postulates that the condensation centre was not located sufficiently close to the sex determiner for there to be a similar interaction between them. In evolutionary biology, sexual selection based on sexual heteromorphism seems to have played a very important role in the evolution of various systems for regulating the genetic interest of the two parents. Few examples
of such systems include dosage compensation, genomic imprinting and haplodiploidy system (Moore and Haig, 1991).

Meiotic pairing is normally dependent on structural and conformational homology between the two chromosomes involved. With respect to sex chromosomes, this can be illustrated in terms of behaviour of sex chromosomes in somatic and reproductive cells. In the male heterogametic XX/XY system, the XX females exhibit heterochromatinization of one of its X chromosomes in somatic cells and its euchromatinization in oocytes. In contrast, XY males exhibit euchromatinization of its X in somatic cells and heterochromatinization in spermatocytes. The problem of inadequate pairing in non-homologous region is avoided when these regions either become heterochromatinized or euchromatinized. These ideas led Jablonka and Lamb (1990) to suggest that the need to avoid effects of pairing failure may have played an important role in the evolution of sex chromosome heteromorphism. According to them, reduced crossing over between the sex chromosomes, functional and structural degeneration of the Y chromosome, sex-chromosome imprinting and dosage compensation may all be evolutionary consequences of the requirement for pairing at meiosis. There are two pathways, conformational and structural, known to be associated with initiation of reduced recombination in the non-homologous regions of the heteromorphic sex chromosomes. Conformational pathway assumes that the event initiating evolutionary changes leading to sex chromosome heteromorphism was a change in chromatin conformation in the region in which the sex determining genes were located. The structural pathway assumes that the initiating event was a structural change resulting from an inversion or a translocation in the originally homomorphic sex chromosomes (Jablonka and Lamb, 1990).

2.2 Sex chromosome evolution in human

The human X and Y chromosomes, like those of other animals, are thought to have evolved from an ordinary pair of autosomes (Charlesworth, 1996). Except for the pseudoautosomal regions at the termini of the X and Y chromosomes, most of the Y chromosome do not recombine during male meiosis, ensuring nucleotide sequence identity of the X and Y chromosomes. The non-recombining regions of the X and Y chromosomes have become highly differentiated during evolution, and only a few
sequence similarities persist within them. These similarities are the reminiscence of
the extensive sequence identity between ancestral X and Y chromosomes once
existed.

Lahn and Page (1999) described the 19 homologous genes mapping to the distal end
of the short arm of the X chromosome and throughout the euchromatic region of the
Y chromosome, which are useful for tracing human sex chromosome evolution. They
hypothesized that during evolution, differentiation of the Y from the X chromosome
was initiated in one region at a time in the order of their physical position. Bengtsson
and Goodfellow (1987) suggested that X and Y differentiation would have occurred
only after recombination ceased between them. These findings suggest that during the
evolution of X and Y chromosomes recombination was suppressed regionally. Singh
et al. (1980) and Jones and Singh (1981) showed that increase in Bkm sequences
precedes the morphological differentiation of sex chromosomes and correlates with
the evolution of chromosomal sex determination.

Jaarola et al. (1998) described that chromosomal inversions, which are known to be
capable of suppressing recombination across the broad regions in mammals, would
appear to be the most likely mechanism. The inversions on the evolving Y
chromosome must have occurred where the different regions have been scrambled,
but not on the X chromosome. Murphy et al. (1999) in their evolutionary theory
predicted that once recombination ceased within a region of the X and the Y
chromosomes, the genes on the affected portion of the Y chromosome began to decay
resulting in most of the Y-linked genes ultimately being obliterated.

In contrast, the homologous genes on the X chromosome were upregulated and
subsequently became subject to X inactivation (Jegalian and page, 1998). Kumar and
Hedges (1998) suggested that divergence between X and Y chromosomes began
shortly after the mammalian lineage arose having diverged from the lineage of birds
(with Z-W sex chromosomes) between 300 and 350Myrs ago. As the sex
chromosomes of birds appear to be completely unrelated to the mammalian sex
chromosomes, it is thought that they arose independently from a different autosomal
pair (Fridolfsson et al., 1998).
2.3 The Human Y chromosome

The Y chromosome is one of the smallest chromosomes in the human genome whose size is about 59Mb and represents around 2–3% of haploid genome. The pseudoautosomal regions (PAR) are where the Y chromosome pairs and exchanges genetic material with the pseudoautosomal region of the X chromosome during male meiosis. PAR1 is located at the terminal region of the short arm (Yp), and the PAR2 at the tip of the long arm (Yq). PAR1 and PAR2 cover approximately 2.4Mb and 0.4Mb of DNA respectively. Due to pairing in these regions, genes located within the PAR are inherited in the same manner as autosomal genes. The euchromatic region is distal to the PAR1 and consists of the short arm paracentromeric region, the centromere and the long arm paracentromeric region of the Y chromosome. The heterochromatic region comprises distal Yq corresponding to Yq12. This region is assumed to be genetically inert and polymorphic in length in different male populations. Heterochromatin is composed mainly of two highly repetitive sequence families, DYZ1 and DYZ2, containing about 5000 and 2000 copies of each respectively (Nakahori et al., 1986; Manz et al., 1992). About 95% of the Y is made up of the “Non-Recombining Y” (NRY). This includes the euchromatic and heterochromatic regions of the Y chromosome.

2.4 Physical and molecular mapping of human Y chromosome

The physical mapping of the Y chromosome has mainly depended on naturally occurring deletions on this chromosome. The creation of a deletion map and the resultant ordering of DNA loci along the chromosome, are very useful not only in locating genes but also in studying the structural diversity of the Y within and among human populations. The first attempts at mapping the Y were based on cytogenetically detectable deletions on this chromosome and then from the limited accuracy and resolution of chromosome banding patterns. Tiepolo and Zuffardi (1976) hypothesized that gene or genes located on Yq were related to spermatogenic failure. Vollrath et al. (1992) constructed a deletion map of the Y chromosome based on the detections of about 200 sequence-tagged sites (STS). The presence or absence of these STS’s on a large set of patients with a wide range of Y anomalies subdivided the euchromatin into 43 ordered intervals, all defined by naturally occurring
chromosomal breakpoints. These 43 deletion intervals further refined the seven­interval map of Vergnaud et al. (1986). This collection of ordered STS’s along the Y chromosome have been extensively used in order to define shortest deleted regions associated with particular phenotypes, in identifying Y chromosomal genes, and in exploring the origin of Y chromosome disorders. A library of yeast artificial clones (YAC) from a human XYYYY male was screened with the Y-specific STS’s in order to identify those containing the corresponding sequences. With this, a complete physical map of the Y chromosome was generated with 196 overlapping DNA clones, which covered 98 percent of the euchromatic region (Foote et al., 1992). Based on the location and apparent function, known human Y genes can be classified under pseudoautosomal loci and three basic classes of non-recombining, male-specific loci.

2.5 Pseudoautosomal region (PAR) of the Y chromosome

About a dozen pseudoautosomal genes, most of them on the short arm have been identified. Most of these genes elude X inactivation, as would be expected of genes with sex-uniform dosage. Curiously, two genes on the Y long arm of human PAR, SYBL1 (synaptobrevinlike1) and HSPRY3 (sprouty Drosophila homologue 3), reportedly undergo X and Y inactivation in females and males respectively, which indicates that this region might have a complex evolutionary history that involves recent X-to-Y translocation (Ciccodicola et al., 2000).

2.6 Non-recombining region of the Y (NRY)

The non-recombining region of the Y (NRY) consists largely of highly repetitive sequences that are rich in transposons and other repetitive elements. Of the ~56Mb of NRY, ~24Mb are euchromatic. Most of the remainder is a block of heterochromatin on the long arm. So far, 21 distinct genes or gene families have been identified in the human NRY. The genes present in this region and their expression pattern are listed in the Table I.
2.7  The Y chromosome genes

Compared to the other human chromosomes, the Y chromosome has a limited number of genes. This might have been the result of the degeneration of the genes of the Y chromosome during evolution (Graves, 1995). The most ancestral gene functions were retained on the nascent X chromosome but deteriorated on NRY portion of the emerging Y giving females with two copies but males with only one copy of many genes. The gene dosage problem has been solved through inactivation of one X chromosome in females. Lahn and Page (1997) identified 12 novel genes or gene families and assessed their expression in diverse human tissues. All NRY genes can be divided into two different categories; (i) those genes which are ubiquitously expressed, have X homologues, appear in a single copy on the NRY, and exhibit housekeeping cell functions, (ii) the genes expressed specifically in the testes, exist in multiple copies (with the exception of SRY) on the NRY, and encode proteins with more specialized functions.

2.8  SRY: a master regulatory gene

Sinclair et al. (1990) identified the gene responsible for testicular determination named SRY (Sex-determining Region on the Y chromosome). SRY gene was cloned by isolation of small fragments of translocated Y on XX sex reversed patients. This gene is located on the short arm of the Y chromosome close to the pseudoautosomal boundary. It comprises of a single exon encoding a protein of 204 amino acids which presents conserved DNA-binding domain (the HMG-box: High Mobility Group), suggesting that this protein regulates gene expression. The SRY has been proposed to be the master gene regulating the cascade of testis determination. Although many genes and loci have been proposed to interact with SRY protein such as WT-1 (Wilm's tumour gene), SF-1 (Steroidogenic Factor 1) and SOX-9, the question of how these genes are regulated by SRY is still unanswered.

Differentiation of bipotential gonadal primordium to testis or ovary is a developmental process common to animal kingdom. However, since the discovery of SRY, the Y-linked testis-determining gene in mammals, extensive efforts have failed to find its orthologue in other organisms. The fact that there are vertebrates such as
alligators and crocodiles in which incubation temperature determines whether the bipotential gonadal primordium is going to develop as testis or as ovary, testifies that the gene(s) involved in gonadogenesis cannot be confined to only one sex. These have to be present in both sexes, otherwise development of all males or all females simply by incubating at higher or lower temperature, respectively, is not possible. This may strongly suggest that the genes involved in gonadogenesis have perhaps not yet been discovered. In our laboratory, Wdr13 gene knock out male mice are sterile and seminiferous tubules are completely devoid of germ cells (Singh et al., in press and our unpublished observations). The Wdr13 gene knockout mice may help in dissecting out the early events of gonadogenesis and identifying the genes involved in this process. Several other genes which have been implicated in sex determination in mammals are: SOX9, AMH, WT1, SF1, DAX1 and DMRT1. Analysis of these genes have revealed that sex determination results from a complex interplay between the genes in this network which may differ in other networks (Morrish and Sinclair, 2002)

2.9 Genes involved in spermatogenesis

Tiepolo and Zuffardi (1976) reported the occurrence of cytogenetically detectable de novo deletions in six azoospermic individuals, suggesting the role of the Y chromosome in spermatogenesis. These observations led the authors to postulate the existence of a locus called Azoospermia Factor (AZF) on Yq11 required for complete spermatogenesis. The location of AZF in Yq11 was further confirmed by numerous studies at cytogenetic and molecular level. AZF was localised to the deletion intervals 5 and 6. Vollrath et al. (1992) reported about 200 Y-specific STS markers that allowed a much simpler Y chromosome screening for microdeletions. The original AZF region was further subdivided into three different non-overlapping sub regions in Yq11 associated with male infertility named AZFa, AZFb, and AZFc (Vogt et al., 1996). Each one of these regions contains several genes proposed as candidate genes involved in male infertility.

2.10 Azoospermia Factor a (AZFa)

The AZFa region is located in proximal Yq within the deletion interval 5. Genes such as DBY, UTY, TB4Y, and DFFRY/USP9Y have been mapped to this region. The first
three genes have no apparent specialized function and they seem to be involved in cellular "housekeeping." By contrast, the DFFRY gene has been proposed to play a role in gametogenesis (Sun et al., 1999; Brown et al., 1998).

2.11 Azoospermia Factor b (AZFb)

The AZFb region is located between deletion interval 5 and proximal deletion interval 6. Genes such as RBM, CDY, XKRY and SMCY are mapped to this region. The RBM gene encodes germ cell specific nuclear proteins containing RNA-binding motif and it is present in multiple copies along the Y. However, not all of these copies are functional and most may be pseudogenes. It has been strongly proposed as a candidate infertility gene since its expression is testis-specific; it is recurrently deleted in azoospermic men, and seems to be specifically expressed in spermatogonia and primary spermatocytes (Ma et al., 1993).

2.12 Azoospermia Factor c (AZFc)

The AZFc region is located in the proximity of the heterochromatic region distal to Yq11 (Reijo et al., 1995). This region contains the DAZ (Deleted in Azoospermia) gene cluster, two copies of the PTP-BL Related Y (PRY), Basic Protein Y2 (BPY2), as well as copies of CDY and RBM. DAZ encodes a testis-specific RNA-binding protein (Reijo et al., 1995). It has been hypothesized that DAZ originated from a translocation and subsequent amplification of this ancestral autosomal gene (Reijo et al., 1996; Saxena et al., 1996). Cooke et al. (1996) described the homologue of the human Y-linked DAZ gene named Dazla (DAZ like autosomal) in the mouse which is located on chromosome 17. Knockout mice for this gene have been shown to be infertile in both the sexes (Ruggiu et al., 1997). Although DAZ has been proposed as the cause of the AZFc phenotype, other genes must be involved since deletions within AZFc region, without including DAZ, have been reported (Stuppia et al., 1996).

The other genes identified within this region, PRY, BPY2, and TTY2, also exhibit a testis-specific expression and are present in multiple copies on the Y. Many of the AZF genes have been proposed as candidate genes involved in human male fertility.
on the basis of their expression profiles (testis-specific or highly expressed in testis) and sterile phenotypes from targeted disruption of their homologues in mice.

2.13 Degeneration of the Y chromosome

In mammals, sex chromosomes probably arose with the differentiation of \textit{SRY} from its homologue \textit{SOX3}, which persists on the mammalian X (Stevanovic \textit{et al.}, 1993). Sequence and expression comparisons indicate that \textit{SRY} and \textit{SOX3} descend from a specific progenitor gene, \textit{SRY} having gained and kept the male-determining function (Foster and Graves, 1994). The mammalian X and Y diverged, the gross structure of the X changing remarkably little, while the Y rapidly degenerated (Bull, 1983; Lahn and Page, 1999). The uncontrolled erosion of gene activity from evolving Y has long been noted. Potential causes and mechanisms of Y-specific degeneration have drawn exciting theories and speculations. Why and how have large X and Y regions stopped recombining with each other? And why Y genes might tend to decay once they stop recombining with their X counterparts?

Studies have indicated that on the evolutionary lineage leading to humans, the mutually non-recombining portions of the human X and Y greatly expanded several times, each time converting a block of previously freely recombining sequence into X and Y-specific regions (Lahn and page, 1999). The striking similarity in gene order seen among mammalian X, compared to the relative scrambling of genes seen among mammalian Y, indicates that such coarse block-wise consolidation of Y-haplotype linkage was probably caused by serial, large scale inversion of much of the Y itself. Such inversions would have disrupted alignment, and thus recombination between progressively larger regions of the X and Y chromosomes. At least four multigene inversions seem to mark the human Y lineage; the first ~300 million years ago and the last ~30 million years ago (Lahn and page, 1999).
Schematic representation of deletion intervals on human Y chromosome.
Table I: Genes on the human Y chromosome.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Gene</th>
<th>Deletion interval</th>
<th>Copy number</th>
<th>Function</th>
<th>Expression</th>
<th>X-homology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SRY</td>
<td>1</td>
<td>single</td>
<td>Transcription, Sex determination</td>
<td>Testis</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>ZFY</td>
<td>1</td>
<td>single</td>
<td>Zinc finger transcription factor</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>RPS4Y</td>
<td>1</td>
<td>single</td>
<td>Protein of small ribosomal subunit</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>TSPY</td>
<td>2, 4A</td>
<td>multiple</td>
<td>Cyclin B binding protein</td>
<td>Testis</td>
<td>no</td>
</tr>
<tr>
<td>5</td>
<td>PCDHY</td>
<td>2</td>
<td>single</td>
<td>Protocadherin cell adhesion</td>
<td>Brain</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>PRK</td>
<td>3, 6</td>
<td>single</td>
<td>Ser/Thr protein kinase</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>7</td>
<td>AMELY</td>
<td>3</td>
<td>single</td>
<td>Tooth enamel formation</td>
<td>Tooth bud</td>
<td>yes</td>
</tr>
<tr>
<td>8</td>
<td>PRY</td>
<td>3, 6</td>
<td>multiple</td>
<td>Protein tyrosine phosphatase</td>
<td>Testis</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>TTY1&amp;2,</td>
<td>3, 4A, 6</td>
<td>multiple</td>
<td>Testis transcript</td>
<td>Testis</td>
<td>no</td>
</tr>
<tr>
<td>10</td>
<td>USP9Y</td>
<td>5</td>
<td>single</td>
<td>Deubiquinating enzyme</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>11</td>
<td>DBY</td>
<td>5</td>
<td>single</td>
<td>DEAD box RNA helicase</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>12</td>
<td>UTY</td>
<td>5</td>
<td>single</td>
<td>TPR-motif</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>13</td>
<td>TB4Y</td>
<td>5</td>
<td>single</td>
<td>Actin sequestration</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>14</td>
<td>VCY</td>
<td>5, 6</td>
<td>multiple</td>
<td>Variable charged protein</td>
<td>Testis</td>
<td>yes</td>
</tr>
<tr>
<td>15</td>
<td>CDY</td>
<td>5, 6</td>
<td>multiple</td>
<td>Chromodomain protein</td>
<td>Testis</td>
<td>no</td>
</tr>
<tr>
<td>16</td>
<td>XKRY</td>
<td>5</td>
<td>multiple</td>
<td>Membrane transport protein</td>
<td>Testis</td>
<td>No</td>
</tr>
<tr>
<td>17</td>
<td>SMCY</td>
<td>5</td>
<td>single</td>
<td>Transcription factor</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>18</td>
<td>EIF1AY</td>
<td>5</td>
<td>single</td>
<td>Transcription initiation factor</td>
<td>Ubiquitous</td>
<td>yes</td>
</tr>
<tr>
<td>19</td>
<td>DAZ</td>
<td>6</td>
<td>multiple (−4)</td>
<td>RNA binding protein</td>
<td>Testis</td>
<td>no</td>
</tr>
<tr>
<td>20</td>
<td>RBMY</td>
<td>4A, 6</td>
<td>Multiple (−30)</td>
<td>RNA binding protein</td>
<td>Testis</td>
<td>yes</td>
</tr>
<tr>
<td>21</td>
<td>BPY2</td>
<td>6</td>
<td>multiple</td>
<td>Basic protein</td>
<td>Testis</td>
<td>no</td>
</tr>
</tbody>
</table>
Makrinou et al. (2001) in an effort to discover Y-linked, testis-expressed genes, identified nineteen cDNAs that showed similarity with the \textit{TTY2} (Testis specific transcript Y-linked) gene and indicated that \textit{TTY2} is part of a large gene family. Further, screening of a panel of Y-linked cosmids clones revealed that the \textit{TTY2} gene family includes at least 26 members organized in 14 sub families. Their studies showed that \textit{TTY2} genes are arranged in tandemly arrayed clusters on both arms of the Y chromosome and each gene is comprised of a series of tandemly arranged repeats. Interestingly, none of the genes contained any open reading frame. The authors suggested that the \textit{TTY2} gene family is composed of multiple copies, some of which may function as non-coding RNA transcripts and some may be pseudogenes.

2.14 Non-coding RNA genes

Non-coding RNA (ncRNA) genes produce functional RNA molecules rather than proteins. Recently, several different systematic screens have identified a surprisingly large number of new ncRNA genes. Non-coding RNAs seem to be particularly abundant in roles that require highly specific nucleic acid recognition without complex catalysis, such as in directing post-transcriptional regulation of gene expression or in guiding RNA modifications. Knowledge of ncRNAs has been limited to biochemically abundant species and anecdotal discoveries. Even after the completion of many sequencing of genomes, both the number and diversity of ncRNA genes remain largely unknown. How many ncRNA genes are there? How important are they? What functions does a cell delegate to RNA instead of protein, and why? These are the questions, which are not answered yet. Recently, several groups have carried out systematic ncRNA gene identification screens along three main lines: (i) cDNA cloning and sequencing, customized to find new small non-mRNAs (Hüttenhofer et al., 2001); (ii) specially designed cDNA cloning screens for a new regulatory RNA gene family of tiny RNAs called microRNAs (miRNAs) (Lee and Ambros, 2001; Lau et al., 2001) and (iii) general ncRNA gene-finding exercises using computational comparative genomics (Argaman et al., 2001).

New non-coding RNAs continue to appear; among these more interesting are that RNAs having roles in chromatin structure (Kelley and Kuroda, 2000). For an example, in human the \textit{XIST} (X-inactive-specific transcript) RNA, a 17Kb ncRNA,
plays a key role in dosage compensation and X chromosome inactivation (Avner et al., 2001). Drosophila also seems to control dosage compensation using small chromatin-associated roX (RNA on the X) RNAs (Franke and Baker, 2000). Several large ncRNAs have been found to be expressed from imprinted regions of vertebrate chromosomes, including the IPW (imprinted in Prader–Willi syndrome) and H19 (imprinted maternally expressed untranslated mRNA) transcripts (Tilghman, 1999). Many of these RNAs are cis-antisense RNAs that overlap coding genes on the other genomic strand. Various cis-antisense RNAs have been observed in prokaryotes (Cavaille et al., 2000), plants (Terryn and Rouze, 2000) and animals (Erdmann et al., 2001), and their roles are unlikely to be limited to those in imprinting and chromatin structure. The nucleolus is rich in snoRNAs, most of which are 70–250 nucleotides in length (Fournier and Maxwell, 1993). Some snoRNAs have roles in ribosomal RNA processing, but most function in rRNA modification (Eliceiri, 1999).

Although untranslated RNA genes, such as transfer RNAs and ribosomal RNAs, perform essential housekeeping roles in all living organisms, growing numbers of other RNAs, some widely conserved across phyla and others limited to certain species, are being uncovered and shown to fulfill specific functions (Pasquinelli, 2002). Animal genomes contain an abundance of small genes that produce regulatory RNAs of about 22 nucleotides in length. These microRNAs are diverse in sequence and expression patterns, and are evolutionarily widespread, suggesting that they may participate in a wide range of genetic regulatory pathways (Ambros, 2001). Lee and Ambros (2001) using bioinformatics and cDNA cloning, found 15 new miRNA genes in C. elegans. Several of those genes express small transcripts that vary in abundance during C. elegans larval development, and three of them have apparent homologs in mammals and/or insects.

In animals, the double-stranded RNA-specific endonuclease Dicer produces two classes of functionally distinct, tiny RNAs: microRNAs (miRNAs) and small interfering RNAs (siRNAs). miRNAs regulate mRNA translation, whereas siRNAs direct RNA destruction via the RNA interference (RNAi) pathway (Grishok and Mello, 2002). RNA interference (RNAi) is a form of post-transcriptional gene silencing that has been described in a number of plant, nematode, protozoan, and invertebrate species (Hannon, 2002). Genes for tiny RNAs have been found to be
plentiful in the genomes of worms, flies, humans and probably all animals. Some of these microRNAs have been conserved through evolution, and many are expressed only at specific times or places. How they act is just beginning to be understood (Moss, 2002). RNAi represents a conserved regulatory motif, which is present in a wide range of eukaryotic organisms. Paddison et al. (2002) showed that endogenously encoded triggers of gene silencing act through elements of the RNAi machinery to regulate the expression of protein-coding genes. These small temporal RNAs (stRNAs) are transcribed as short hairpin precursors (approximately 70 nt), processed into active 21-nt RNAs by Dicer, and recognize target mRNAs via base-pairing interactions.

A 20Kb silent domain at the mating-type region of fission yeast was shown as a model for heterochromatin formation (Hall et al., 2002). It has been shown that histone H3 methylated at lysine 9 (H3 Lys9) directly recruits heterochromatin protein Swi6/HP1, the critical determinant for H3 Lys9 methylation, to spread in cis and to be inherited through mitosis and meiosis is Swi6 itself. Further they have shown that centromere-homologous repeat (cenH) present at the silent mating-type region is sufficient for heterochromatin formation at an ectopic site, and that its repressive capacity is mediated by components of the RNA interference (RNAi) machinery.

Hutvagner and Zamore (2002) showed that in human cell extracts, the miRNA let-7 naturally enters the RNAi pathway, which suggests that only the degree of complementarity between a miRNA and its RNA target determines its function. Human let-7 is a component of a previously identified miRNA-containing ribonucleoprotein particle, which they show as an RNAi enzyme complex. Llave et al. (2002) showed that in Arabidopsis thaliana miRNA 39 (also known as miR171), a 21-ribonucleotide species that accumulates predominantly in inflorescence tissues, is produced from an intergenic region in chromosome III and functionally interacts with mRNA targets encoding several members of the Scarecrow-like (SCL) family of putative transcription factors. miRNA 39 is complementary to an internal region of three SCL mRNAs. These interactions result in specific cleavage of target mRNA within the region of complementarity, indicating that this class of miRNA functions like small interfering RNA associated with RNA silencing.
2.15 Segmental duplications

Segmental duplications are large, nearly identical copies of genomic DNA, which range in size from 1 to >200Kb and are present in at least two locations in the human genome (The International Human Genome Sequencing Consortium 2001; Bailey et al., 2001). It is now estimated that ~5% of our genetic material is composed of segmental duplications that have emerged during the past 35 million years of evolution of our species (Venter et al., 2001). Many of these recently duplicated segments are located in regions that are hot spots of chromosomal and/or evolutionary instability, which indicates that there is likely to be an important link between the processes of chromosomal rearrangement and duplication.

Their high sequence identity (90–100%) provides ample substrate for paralogous recombination events to occur and they have been identified on every human chromosome. The distribution of these segments among human chromosomes seems non-uniform, some chromosomes such as the Y, showing peculiar enrichments for these types of duplication (Tilford et al., 2001). The identification and characterization of segmental duplications have been on the basis of both computational and fluorescence in situ hybridization (FISH) analysis of the human genome (Cheung et al., 2001). Both the types of analyses consider only large blocks with 90–100% sequence similarity. In silico analysis has revealed a 3–5 fold enrichment of duplication within 100Kb and 1Mb of telomeres and centromeres respectively (Bailey et al., 2001). Large intrachromosomal duplications (>200Kb) have been documented for many human chromosomes (Hattori et al., 2000; Adams et al., 2000; Horvath et al., 2000).

2.16 Segmental duplication on human Y chromosome

Kuroda-Kawaguchi et al. (2001) determined the complete nucleotide sequence of AZFc by identifying and distinguishing between near-identical amplicons (massive repeat units) using an interactive mapping-sequencing process. They identified a complex of three palindromes with 99.97% identity with the AZFc region. These palindromes are constructed from six distinct families of amplicons with unit lengths of 115-678Kb, and may have resulted from tandem duplication and inversion during
primate evolution. Also, the authors have reported that the palindromic complex contains 11 families of transcription units, all expressed in testis. Blanco et al. (2001) showed that the human Xq21.3/Yp11.2 blocks exhibit a homology of 99% nucleotide identity with the exception of an internal X-specific region containing the marker DXS214.

2.17 Junk or non-coding DNA sequences

Early work on organization of eukaryotic genome by using renaturation kinetics (Britten and Kohn, 1968) revealed that a large fraction of the genome consists of non-coding repetitive DNA sequences, also called as Junk DNA. It has been estimated that 97% of the human DNA is junk or non-genic and the remaining 3% is genic (Nowak, 1994). However, recent studies on this aspect are clearly providing evidence for indispensable role of Junk DNA in the regulation and organization of centromere, telomere, chromosomal architecture and gene activity. The following gives a brief account of the known families of the junk DNA (repeated DNA sequences) characteristic of eukaryotic systems with their possible known functions.

2.18 Introns

Most eukaryotic genes contain introns that do not code for proteins. Some of the introns are now known to code for small nucleolar RNA (snRNA) believed to be essential for the assembly of ribosomes (Bachellerie et al., 1995). Based on an analysis of the cDNA to genomic alignments in all of the complete or nearly complete genomes from the multicellular organisms, Wong et al. (2000) proposed that in animals but not in plants, most of the "junk" is intron DNA. Muller et al. (2002) showed that the intron/exon borders are highly conserved in genes from sponges to human.

2.19 Satellites

They are short DNA sequences 140-180bp long, repeated 100-1000 times at a stretch. They mainly occur at the ends and the centromers of the chromosome. They are essential for the structural and functional integrity of chromosomes, as their absence
leads to chromosomal disintegration. Monod et al. (2002) showed that the suppression of white-mottled gene (wm4) position-effect variegation (PEV) by MATH20 protein is due to a displacement of D1, an artificial AT-hook proteins, from its preferred binding site and provides additional support for a direct role of D1 in the assembly of AT-rich satellite heterochromatin.

2.20 Minisatellites

Minisatellites are similar to satellites but are shorter up to 40bp long and occur throughout the genome. Defective minisatellites are found to be associated with cancer. In human and yeast model systems, it has been shown that some minisatellite loci are relatively stable in somatic cells but not in the germline. Unlike microsatellite sequences, mini satellites are not destabilized by mismatch repair mutations (Lopes et al., 2002).

2.21 Microsatellites

They are shorter than minisatellites, usually 1-6bp long (see below). GATA sequences of banded krait minor satellite are examples of microsatellite and are known to be essential for the organization and activity of heteromorphic sex chromosome (Singh, 1995).

2.22 SINES (small interspersed elements) and LINES (long interspersed elements)

They occur in numerous copies. One example of SINE is 300bp Alu sequences that occur about 500,000 times in human genome. LINES are similar to SINES but longer up to 700bp repeat unit. Both LINES and SINES hop about the genome and cause mutation if they land in a gene (Nowak, 1994). Ogiwara et al. (2002) have identified a new superfamily of vertebrate short interspersed repetitive elements (SINES) designated as V-SINES, that are widespread in fishes and frogs. V-SINE includes a central conserved domain preceded by a 5'-end tRNA-related region and followed by a potentially recombinogenic (TG), tract with a 3' tail derived from the 3' untranslated
region (UTR) of the corresponding partner long interspersed repetitive element (LINE) that encodes a functional reverse transcriptase. This also suggests that V-SINEs may have some functions that have been maintained by the co-evolution of SINEs and LINEs during the evolution of vertebrates.

2.23 Banded krait minor satellite DNA sequences

Banded krait minor satellite DNA (Bkm) was identified and isolated from a poisonous female Indian snake banded krait (Bungarus fasciatus) by isopycnic density gradient centrifugation (Singh et al., 1980). Bkm sequences are distributed all along the length of W chromosome of most of the snakes and are present in nearly all eukaryotes studied, predominantly concentrated on the heteromorphic sex chromosome (the W chromosome in snakes and birds and the Y chromosome in mouse and human). Sequencing of Bkm positive clones from snake, mouse, human and Drosophila revealed tetrancleotide repeat of GATA as a major conserved component of Bkm. Southern hybridization pattern with GATA rich clone Bkm 2(8) was similar to the pattern with uncloned Bkm in mouse and man, though in human, the binding intensity was higher with uncloned Bkm (Singh et al., 1981; 1984; Singh and Jones, 1982; 1986). This confirmed GATA repeats as the major conserved component of Bkm.

Bkm sequences exhibit a tendency to be concentrated on sex chromosomes thereby implicating their definite role in the sex determining process and/or in the allocyclic behaviour and evolution of sex chromosomes (Jones, 1984; Singh et al., 1984). They are predominantly located on the W chromosome of snakes (Singh et al., 1980) and on the X chromosome of D. melanogaster (Singh et al., 1981). Similar studies in mice showed the presence of a cluster of Bkm sequences in the proximal regions of the Y that is essential for the sex determination. Close linkage of Bkm to sex determination can be seen in Sxr mice in which sex reversal is regularly associated with an exchange of Bkm cluster to the X chromosome (Singh and Jones, 1982). However, Bkm sequences are known to occur on autosomes as in some species of Lepidoptera (Traut and Marec, 1997) and mouse (Kiel-Netzger et al., 1984).
2.24 Evolutionary conservation and functional significance of Bkm sequences

*Bkm* sequences are highly conserved among eukaryotes and have been reported from yeast and from invertebrates like dinoflagelates, echinoderms and arthropods. These sequences are present in most of the vertebrates, amphibians, reptiles including snake, crocodiles, alligators, gharials, turtles and birds (Singh *et al.*, 1994). However, Miklos *et al.*, (1989) reported its absence in bovine and ovine sp. under the hybridization condition that detects single copy DNA sequences. This is however, not correct. *Bkm* sequences have been found in the bovines and ovines also. *In situ* hybridization studies have shown that *Bkm* sequences are preferentially located on the X chromosome of *Drosophila* (Singh *et al.*, 1981), W chromosome of snakes and birds (Singh *et al.*, 1976), Y chromosome of mice (Jones and Singh, 1981), thus suggesting the possibility of sex specific role in the organization and functioning of sex chromosomes (Singh *et al.*, 1976).

2.25 *Bkm* in snakes and birds

Both snakes and birds having female heterogamety (ZZ males/ZW females) show hybridization on the W chromosome with *Bkm* probe (Singh *et al.*, 1976). Quantitative filter hybridization shows the presence of *Bkm* related DNA even in males on autosomes though much lesser in amount compared to the W chromosome (Jones, 1985).

Highly evolved species of snakes belonging to family viperidae and elapidae, have morphologically distinct W chromosome having a very high concentration of *Bkm* sequences. The primitive snakes of the family boidae, however, neither shows morphologically distinct sex chromosomes nor high amount of *Bkm* in their genomes. In some of the snakes of the family colubridae, the sex chromosomes cannot be distinguished morphologically but the W chromosome can be identified at DNA level by *in situ* hybridization with *Bkm* and by its allocyclic Giemsa staining behaviour (Singh, 1972). Thus the increase in *Bkm* sequences precedes the morphological differentiation of sex chromosomes and correlates with the evolution of chromosomal sex determination (CSD) (Singh *et al.*, 1980; Jones and Singh, 1981).
2.26  \textit{Bkm} in mice

\textit{Bkm} sequences are predominantly concentrated in the paracentric region of the mouse Y chromosome (Singh \textit{et al.}, 1981; Jones and Singh 1981; Jones and Singh, 1982; Epplen \textit{et al.}, 1982). This region was later confirmed to be the short arm of the Y chromosome, (Bishop \textit{et al.}, 1988; McLaren \textit{et al.}, 1988; Roberts \textit{et al.}, 1988). A dominant mutation known as Sxr was shown to be the cause of the sex reversal of chromosomally female (XX) individuals into sterile males (Cattanach \textit{et al.}, 1971). Singh and Jones (1982) by their elegant experiments showed that \textit{Bkm} positive region of the mouse Y chromosome is necessary and sufficient to convert a female into a male. When this region is transferred to the distal end of the X chromosome, it causes sex reversal in XX mice. They also showed that in carrier mice (XYSxr), the sex determining \textit{Bkm} positive region of the Y chromosome is duplicated and transferred to the distal end of the long arm of the Y chromosome. The XYSxr males, during the spermatogenesis, pass on this \textit{Bkm}-rich region from the tip of one of the chromatids of the long arm of the Y to one of the chromatids of the X through meiotic cross over between the X and the Y chromosomes. This results in 25% of the progeny as sex reversed XXSxr males, when crossed to normal females. These observations, not only explained the inheritance pattern of Sxr in these XYSxr males and narrowed down the male determining region to the proximal region of the Y chromosome, but also associated \textit{Bkm} intimately, if not specifically, with sex determination. A region in mouse on chromosome 17, rich in \textit{Bkm}, was implicated in autosomal sex reversal (Washburn and Eicher, 1983; Kiel-Metzer and Ericson, 1984) and was shown to contain genes expressed in the testis. These observations further strengthen the role of \textit{Bkm}, as proposed by Singh and Jones (1982), in controlling the expression of genes involved in sex determination and/or spermatogenesis.

A Y chromosome-specific clone M34 was isolated by double screening of male mouse genomic library, initially with uncloned \textit{Bkm} and subsequently with cloned \textit{Bkm} 2(8) containing long stretches of tetranucleotide repeats of GATA. Southern and \textit{in situ} hybridization showed that M34 is distributed all along the Y chromosome, except for the sex determining and pseudoautosomal region (Singh \textit{et al.}, 1988, 1994). Singh (1994) implicated these sequences in bringing about coordinated decondensation of the entire Y chromosome.
2.27  *Bkm* in human

*Bkm* sequences are distributed all over the genome and show very high rate of polymorphism in human DNA. However, Southern analysis of human males and females using *Bkm*, a clone rich in GATA repeats, and uncloned *Bkm* probe showed no sex-specific hybridization pattern. This is in contrast to the observation in mice (Singh *et al.*, 1981; Jones and Singh, 1982). Intriguingly, *in situ* hybridization with uncloned *Bkm* on human male chromosomes showed hybridization with most of the small acrocentric chromosomes, including the Y. The grains were concentrated more on the proximal region of the Y chromosome. This region associates with sex determination (Singh and Jones, 1986; Arnemann *et al.*, 1986; Rasheed *et al.*, 1991).

Screening of the human Y chromosomal DNA library with uncloned *Bkm* yielded C102, among many positive clones. Genomic DNA hybridization with p102d (2), a subclone of C102, showed male-specific pattern, suggesting, its origin from the Y chromosome. *In situ* hybridization showed its distribution along the length of the Y chromosome, except for the short arm. This distribution of p102d(2) is astonishingly similar to the distribution pattern of M34 on the mouse Y chromosome (Singh and Majumdar, 1993).

2.28  *Bkm* in *Drosophila*

*In situ* hybridization with the chromosomes of the larval salivary gland of *Drosophila melanogaster* localized *Bkm* to a small well-defined euchromatic region, 19F-20AB, near the base of the X chromosome. Southern blot analysis showed no difference in the hybridization pattern of the *Bkm* with adult male and female DNA. *Bkm*-positive clones recovered from *Drosophila* genomic library showed hybridization, as expected, to the base of the X chromosome in the region 19F-20 AB (Singh *et al.*, 1981).

2.29  Sequence analysis and transcription pattern of *Bkm*

Sequencing of *Bkm* positive fragments isolated from *Drosophila*, snake, mouse, and human clones showed simple repeats of tetranucleotide GATA as a major repeat type.
(Epplen et al., 1982; Singh et al., 1984, 1994; Schafer et al., 1986). Bkm associated sex chromosome-specific repeats recovered from mouse and human however, were found to be species-specific. The only conserved component of repeats between snake, mouse and human clones was GATA repeats.

2.30  **Bkm is the structural and functional component of the sex chromosomal chromatin**

Bkm sequences are distributed along the length of the W chromosome in snakes. Similarly Bkm related sex and species-specific sequences of mouse and man i.e. M34 and p102d(2) respectively are distributed all along the length of the Y chromosome except the sex determining region. Thus the structural organization of W chromosome of the snakes and the Y chromosome of mammals seems to be similar with respect to Bkm sequences.

The W chromosome of snakes and the Y chromosome of mammals are heterochromatic, transcriptionally inactive, showing allocyly in their DNA replication, and remain highly condensed in all somatic tissues (Raychaudri et al., 1971; Singh and Raychaudri, 1975; Eicher and Washburn, 1986). In situ hybridization with Bkm revealed predominant localization of grains in a single region of every interphase nucleus of somatic cell of snakes, thus confirming the condensed state of W chromosome in the form of a W-chromatin body (Singh, 1979). The W chromosome of snake and Y chromosome of mouse remain highly condensed in various somatic tissues (Singh et al., 1979) but decondense extensively in the germ cells, developing oocytes in the ovary of snakes and testis of mouse. Similarly in *Drosophila*, extensive decondensation of Bkm positive region is seen in the germ cells, while this region remains highly condensed in somatic cells of the ovary.

2.31  **Interphase chromatin is highly dynamic**

The position of genes within the nucleus has been correlated with their transcriptional activity. The interchromosome domain model of nuclear organization suggests that genes preferentially locate at the surface of chromosome territories. Conversely, high-resolution analysis of chromatin fibers suggests that chromosome territories do not
present accessibility barriers to transcription machinery. Mahy et al. (2002) demonstrated that transcription of both ubiquitous and tissue-restricted genes is not confined to the periphery of chromosome territories, suggesting that the basal transcription machinery and transcription factors can readily gain access to the chromosome interior. In *Saccharomyces cerevisiae*, Harata et al. (2002) showed the correlation between chromatin association and transcriptional regulation for the Act3p/Arp4 nuclear actin-related protein.

Chromatin exhibits extremely dynamic behaviour in both yeast and *Drosophila* nuclei (Marshall et al., 1997; Heun et al., 2001). By high-resolution tracking, a tagged site inserted near the telomere of the X chromosome was examined in *Drosophila* spermatocytes. Vazquez et al. (2001) observed large rapid movements (>0.5 μm in a 10-s interval) of the internal chromosomal loci in G1 and S-phase nuclei in yeast. However, interphase chromatin movements were not completely random. Individual movement lengths have a roughly Gaussian distribution in the studies performed on *Drosophila* spermatocytes (Vazquez et al., 2001).

Periodic attachment of chromatin to the nuclear margin seems to be a conserved mechanism for positioning chromosomes, although such attachment does not necessarily occur through telomeric sequences. Marshall et al. (1996) by a three-dimensional (3D) deconvolution microscopy on fixed cells has provided a clear evidence for anchorage sites along *Drosophila* chromosomal arms. Rapid time-lapse microscopy of such sites tagged with green fluorescence protein (GFP) has not yet been performed; thus it can only be inferred from their reproducible positions that these sites will show constrained mobility. Whether this involves short inserts of repetitive DNA or a clustering of unlinked or distant sites, as observed for yeast telomeres (Heun et al., 2001), is not yet known. The mobility of a large cluster of a 359bp satellite repeat on the *Drosophila* X chromosome was visualized through its affinity for a fluorescently modified topoisomerase II and was tracked by 3D time-lapse microscopy of stage-12 embryos (Marshall et al., 1996).
2.32 Repetitive DNA in chromatin organization

The analysis using fluorescent probes for the large satellite repeats found in higher eukaryotic centromeres has shown that these domains, like the clusters of simple repeat DNA (10 to 1000 copies) that are found dispersed along the chromosomal arms, move little in interphase nuclei (Shelby et al., 1996). Li et al. (1998) showed that this region assumes the characteristics of heterochromatin, being late replicating and tightly condensed. Although heterochromatin-like inserts in mammalian nuclei are less dynamic than tagged sites in yeast or flies, further studies of protein diffusion coefficients based on photobleaching assays suggest that DNA bound proteins in mammalian cells can be highly mobile (Dundr and Misteli, 2001). It remains to be seen whether large blocks of simple repetitive DNA, like centromeric satellite repeats, form zones that are relatively immobile (Shelby et al., 1996).

At metaphase, DNA in a human chromosome is estimated to be compacted at least 10,000 fold in length; though, the higher order mechanisms by which the chromosomes are organized in interphase and consequently further condensed in mitosis have largely remained unexplained. One generally overlooked that participant in chromosome condensation is DNA replication (Pflumm, 2002). The movement of chromatin in nuclei appears to be generally constrained in comparison with many nuclear proteins. These restrictions on motion are proposed to reflect the attachment of chromatin to immobile nuclear substructures. Chubb et al. (2002) suggested a role for the nucleolus and nuclear periphery in maintaining the three-dimensional organization of chromatin in the human nucleus. They also reported that loci at nucleoli or the nuclear periphery are significantly less mobile than other more nucleoplasmic loci. Disruption of nucleoli increases the mobility of nucleolar-associated loci, testifying nuclear substructures constraining the movements of chromatin.

2.33 Matrix attachment region (MAR)

The compact nuclear structure and the bigger genome size of eukaryotes raise questions such as: (a) How gene expression is controlled in a spatial and temporal
manner during various stages of development? (b) How regulatory proteins recognize the DNA information?

Thompson (1992) proposed that some organizational properties of the eukaryotic genome reside in the ability of chromatin to establish autonomous functional units, which specify levels and patterns of gene expression. Chromosome domains are units of gene expression in which transcription from a specific promoter results only from activation by enhancer located within that domain. However, these enhancers cannot act on the promoter that are located in different domains. MAR/SAR are AT rich DNA sequences containing topoisomerase II cleavage sites that mediate the anchoring of chromatin to the nuclear matrix and delineate the boundaries of discrete and topological independent higher order domains (Laemmli et al., 1992). Attachment of DNA to the nuclear matrix via DNA sequence of matrix attachment region (MAR) sequences, and interaction with MAR-binding proteins have been shown to alter chromatin conformation, promote histone acetylation, and influence gene transcription.

Proteins that are bound to the MARs tether the chromatin fiber into structural loops that are held together by the components of the chromosome scaffold. It is possible that such structural domains that organize the chromosome could play a functional role by organizing genes into specific functional units for regulated transcription. The topological chromosomal domains established by MARs reveal that they are not static structures but rather dynamic ones that can change during developmental stages to adopt new expression profiles (Kalos and Fournier, 1995). This insulating effect of MAR sequences may not be constitutive and might depend on cell type and developmental stage of the organism (Thompson et al., 1992). Kieffer et al. (2002) identified elements regulating the human CD8 gene complex. They mapped nuclear matrix attachment regions (MARs) and DNase I hypersensitive (HS) sites over a 100-Kb region that included the CD8B gene, the intergenic region, and the CD8A gene. This clustering of DNase HS sites and MARs capable of binding SATB1 and GATA-3 at the 3' end of the CD8B gene suggests that this region is an epigenetic regulator of CD8 expression. This also suggests that MARs can facilitate long-range chromatin remodeling required for enhancer activity. Liebich et al., (2002) have created the database S/MARt DB (S/MAR transaction database), covering scaffold/matrix
attached regions (S/MARs) and nuclear matrix proteins that are involved in chromosomal attachment to the nuclear scaffold.

2.34 Chromatin organization and gene expression

The findings on the structure of chromatin, its organization in the nucleus, and its involvement in regulating gene expression have given new insight about the understandings of gene expression and regulation (Berger and Felsenfeld, 2001). Studies on chromatin boundaries and insulators revealed that they act as transcriptional regulatory elements and modulate interactions between enhancers and promoters, protecting genes from silencing effects by the adjacent chromatin. The phenomenon was first found in *Drosophila*, and now insulators have been found in a variety of organisms, ranging from yeast to humans. They have been found interspersed with regulatory sequences in complex genes and at the boundaries between active and inactive chromatin. Insulators might modulate transcription by organizing the chromatin fiber within the nucleus through the establishment of higher-order domains of chromatin structure (Gerasimova and Corces, 2001).

Recent evidence indicates that these chicken alpha- and beta-globin genomic domains differ significantly in genomic organization as well as in their mode of packaging into chromatin. Recillas-Targa and Razin (2001) reported a comparative analysis of chicken alpha- and beta-globin gene clusters. Based on their findings, they suggested the existence of three distinct mechanisms that corroborate the activation/repression of genomic domains, namely: (i) the relocation of genomic domains into nuclear compartments, the long-term modifications of the mode of chromatin packaging within the domains, (iii) the domain activation by continuous action of multiple transcription factors and remodeling complexes.

Chen et al. (2001) studied the nonchromatin structure or nuclear matrix in developing spermatogenic cells of the rat and demonstrated that the nuclear matrix of spermatogenic cells consisted of a three-dimensional network of filaments of variable thicknesses. In spermatogonia and spermatocytes, the nuclear matrix consisted of relatively thin filaments while that of round spermatids consisted of a thicker
interconnecting network of filament. Their results showed that the proteins of nuclear matrix changed in a cell stage-specific manner. These stage-specific changes corresponded to the major transitions of chromatin structure and function during spermatogenesis. These results suggest that the nuclear matrix in spermatogenic cells may be involved in mediating DNA modifications and maintaining nuclear organization during spermatogenesis.

Cremer and Cremer (2001) in their review on chromosome territories, nuclear architecture and gene regulation in mammalian cells described that the expression of genes is regulated at many levels: (1) chromosomes occupy discrete territories in the cell nucleus and contain distinct chromosome-arm and chromosome-band domains, (2) chromosome territories (CTs) with different gene densities occupy distinct nuclear positions, (3) gene-poor, mid-to-late-replicating chromatin is enriched in nuclear compartments that are located at the nuclear periphery and at the perinucleolar region, (4) a compartment for gene-dense, early-replicating chromatin is separated from the compartments for mid-to-late-replicating chromatin, (5) chromatin domains with a DNA content of 1Mb can be detected in nuclei during interphase and in non-cycling cells, (6) the interchromatin compartment (IC) contains various types of non-chromatin domains with factors for transcription, splicing, DNA replication and repair, (7) the CT–IC model predicts that a specific topological relationship between the IC and chromatin domains is essential for gene regulation, (8) the transcriptional status of genes correlates with gene positioning in CTs and (9) a dynamic repositioning of genes with respect to centromeric heterochromatin has a role in gene silencing and activation.

2.35 Simple sequence repeats or Microsatellites

Microsatellites or simple sequence repeats (SSRs) are tandemly repeated tracts of DNA composed of 1-6 base pair (bp) long units. They are present both in prokaryotes and eukaryotic genomes (Bell, 1996; Gur-Arie et al., 2000; Toth et al., 2000; Li et al., 2002). Tandem repeats are also reported to be present in the protein coding regions of genes (Borstnik and Pumpernik, 2002). A subset of SSRs, namely trinucleotide
repeats, are of great interest because of the role they play in many human neuro
degenerative disorders (fragile X syndrome, Huntington’s disease, myotonic
dystrophy, spinal-bulbar muscular atrophy, spinocerebellar ataxia, etc). (Warren and
Nelson, 1993; Bates and Lehrach, 1994; Ashley and Warren, 1995; Reddy and
Housman, 1997) and in some human cancers, e.g. hereditary non polyposiscolorectal
carcinoma (Wooster et al., 1994; Arzimanoglou et al., 1998). The alteration
responsible for these genetic diseases is the expansion of triplet repeats, where the rate
of mutation depends on the number of tandem units within the repeat. Hence the term
‘dynamic mutation’ was coined by Richards and Sutherland (1992).

Microsatellites can be found anywhere in the genome, both in protein-coding and
non-coding regions. Because of their high mutability, microsatellites are thought to
play a significant role in genome evolution by creating and maintaining quantitative
genetic variation (Tautz et al., 1986; Kashi et al., 1997). In promoter regions, the
length of SSRs may influence transcriptional activity (Kashi et al., 1997). Length of
poly glutamine or poly proline tracts encoded by SSRs may affect protein-protein
interactions involving transcription factors (Gerber et al., 1994; Perutz et al., 1994). It
has been shown that SSRs in exons are less abundant than in non-coding regions
(Hancock, 1995), and that different taxa exhibit different preferences for SSR types
(Beckman and Weber, 1992; Lagercrantz et al., 1993; Tautz and Schlotterer, 1994).
Moreover, the over all microsatellite content in the genome correlates with the
genome size of the organisms (Hancock, 1996).

SSRs are inherently unstable. Two models have been proposed to explain
microsatellite generation and instability: (i) DNA polymerase slippage and (ii)
equal recombination. The first model involves transient dissociation of the
replicating DNA strands, followed by misaligned reassociation (Richards and
Sutherland, 1994). The slipped structure may be stabilized by hairpin, triplex, or
quadruplex arrangement of DNA strands (Pearson and Sinden, 1998; Sinden, 1999).
Thus, it is expected that those repeats that are able to form such alternative DNA
conformations would be generated more frequently than others. The possible
structures of triplet repeats involved in human diseases have been studied extensively (Baldi et al., 1999).

SSRs that show a considerable potential to form alternative structures include (CAG)n, (CCG)n, (GAA)n, (CCT)n, and (CCA)n (Gacy et al., 1995; Bidichandani et al., 1998; Usdin, 1998). However, some sequences with theoretically high hairpin-forming potential e.g. (CCG)n show the slowest in vitro slippage rate (Schlotterer and Tautz, 1992). Moreover, the rate of alterations is likely to be controlled at multiple steps in vivo. In Drosophila, recent studies have revealed differences both in density and length of microsatellites among species (Pascual et al., 2000). Distant organisms, S. cerevisiae, C. elegans, Arabidopsis, Drosophila, and human show a congruent increase in overall microsatellite presence with genome size when both length and density of microsatellites are considered (Kruglyak et al., 1998, 2000). This relationship is not universal with the puffer fish (Fugu rubripes) having a higher density and longer microsatellite repeats than human but having a genome eight times smaller (Elgar et al., 1999). An active role of the DNA mismatch repair system to stabilize simple sequence repeats has been revealed in E.coli, yeast, and human (Sia et al., 1997). Although a number of experimental results argue in favor of the above model, homologous recombination may also result in genetic instability of certain SSRs (Jakupciak and Wells, 2000; Richard and Paques, 2000). It is expected that the fixation of de novo generated SSRs is determined by the interplay of several factors, of which the repeat type, the genomic position of the SSR, and the genetic-biochemical background of the cell are the most important.