Eukaryotic Chromosomes

Introductory article

Wendy A Bickmore, *MRC Human Genetics Unit, Edinburgh, Scotland, UK*

Chromosomes are the nucleoprotein structures that carry the genetic information. In eukaryotes they are located in the cell nucleus.

Article Contents

Chromosomes

The eukaryotic genome is made up of DNA/protein complexes called chromosomes. Despite the compaction of the DNA (deoxyribonucleic acid) with proteins, gene sequences embedded within chromosomes must still be available for transcription by RNA (ribonucleic acid) polymerases and all of the DNA must be capable of being copied by DNA polymerases.

Chromosomes have two main functions: to ensure that the DNA is segregated equally to daughter nuclei at cell division, and to ensure that the integrity of the genome is maintained and accurately replicated in each cell cycle. The elements responsible for these functions are centromeres, telomeres and replication origins, respectively (**Figure 1**).

Chromosome size and karyotypes

The genome size in eukaryotes varies widely, from those of the yeasts ($\sim$17 million base pairs) to those of vertebrates (3000 million base pairs or more). The genomes of some plants are huge. In *Lilium longiflorum*, for example, the genome consists of 300 000 million base pairs. Similarly, the size and number of chromosomes in any particular species – the karyotype – varies widely. For example, the closely related deer species the Chinese muntjac and the Indian muntjac have a similar genome size but the former has 23 pairs of chromosomes and the latter only three pairs of very large chromosomes. A minimum size is required for a stable eukaryotic chromosome. Small yeast artificial chromosomes (YACs), based on *Saccharomyces cerevisiae*, are stabilized by the presence of additional DNA between the centromere and telomere, and the minimal size appears to be $\sim$50 kb. A maximum limit also exists for chromosome size. It has been suggested that the longest chromosome arm must not be longer than half the length of the spindle axis at telophase.

The human karyotype consists of 22 pairs of autosomes and the sex chromosomes, XX or XY. Each human chromosome contains an average of 100 million base pairs of DNA. Aberrations of chromosome number (aneuploidy) result from errors in chromosome segregation. The presence of only one copy (monosomy) of any autosome is generally not compatible with survival in humans. On the other hand, trisomy (an extra copy of a chromosome) of chromosomes 21 (Down syndrome), 18 and 13 are found in live births in humans. The complex developmental

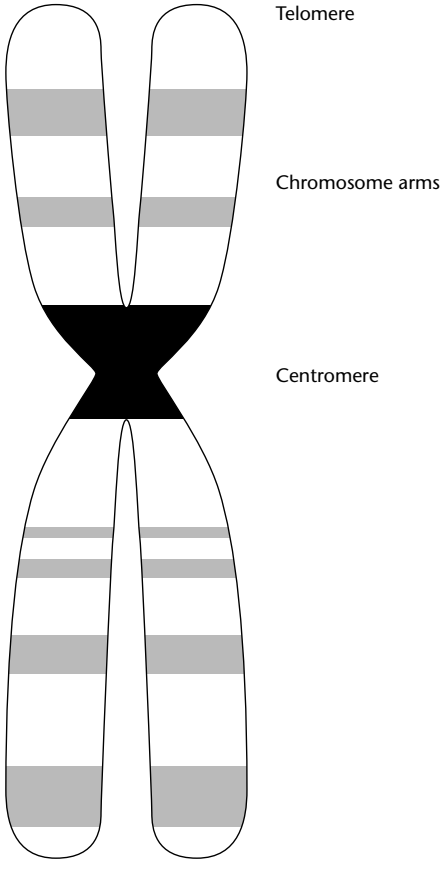

Figure 1 Elements of the chromosome. Simple representation of a metacentric eukaryotic chromosome during mitosis. The ends of the linear chromosome are capped by telomeres (from the Greek *telos*, meaning the end), chromosomal elements that are essential for maintaining the integrity of the genetic material and for preventing fusions between the ends of different chromosomes. The primary constriction marks the position of the centromere. This is the site where the sister chromatids remain attached to one another until anaphase and where the chromosome attaches to the mitotic spindle through a structure known as the kinetochore. In many organisms the centromere is also a site of heterochromatin (black). In many vertebrates the chromosome arms can be differentiated into dark or lightly-staining chromosome bands (hatched and white areas, respectively) using a variety of staining techniques. Replication origins are scattered along the length of the chromosome arms. DNA in different chromosome bands is replicated at different times.

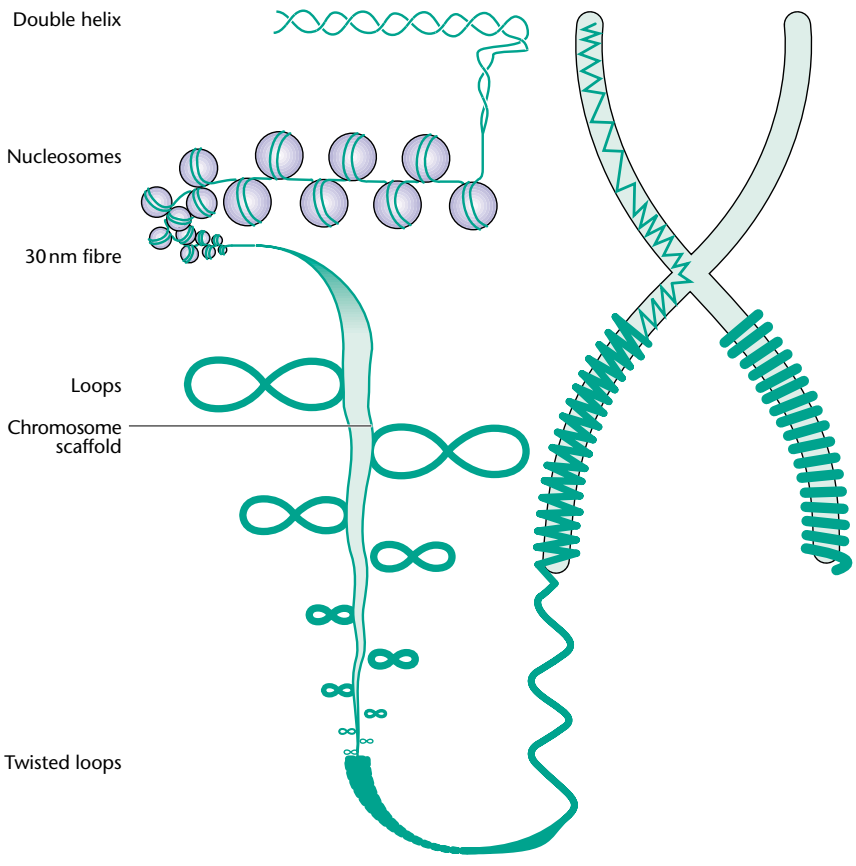

Figure 2 Chromatin packaging within the chromosome. The length of a DNA molecule is shortened by 10 000-fold during the formation of a metaphase chromosome. This is brought about by levels of chromatin packing, built one upon another. The first level of packing is brought about by the wrapping of DNA around nucleosomes to form a structure that has been likened to beads on a string. A folding or coiling of this level of chromatin together with additional chromosomal proteins, such as linker histones, produces the 30-nm diameter chromatin fibre. It is still not certain how chromatin is compacted beyond this stage. The diagram illustrates one particular model, known as the radial loop/scaffold model. In this model it is envisaged that the 30-nm fibre is arranged into loops containing ~100 kb of DNA. These loops are anchored at their bases to the chromosome scaffold. Chromatin is finally condensed both by a shortening in length of the chromosome scaffold and by a twisting of lateral loops in toward the chromosome axis/scaffold.

abnormalities of these individuals are presumed to arise from the increase in dosage of genes carried on these chromosomes. The three chromosomes involved in the viable trisomies have quite low densities of genes. Trisomies of other human chromosomes of similar or even smaller size but with high gene densities are lethal during embryonic or fetal development.

Under certain conditions and following a variety of chemical treatments, chromosomes can be seen to be banded under the light microscope. Each chromosome band contains approximately 5–10 million base pairs of DNA and the banding patterns uniquely identify each chromosome and allow abnormalities in gross chromosome structure to be identified. Chromosome bands have also been used to examine the evolutionary relationships between the karyotypes of closely related species, for example between humans and other primates.

Compaction of chromatin within chromosomes

Chromosomes are maximally condensed at metaphase of the cell cycle, when a typical mammalian chromosome will be around 10 000-fold shorter than if the same length of DNA were present as a simple, naked double helix. Packing of the DNA molecule to form the metaphase chromosome is accomplished by an ordered series of interactions with proteins (**Figure 2**). The first, and best understood, of these is the formation of nucleosomes, which compacts the DNA helix 7-fold. Nucleosomes consist of approximately 146 bp of DNA wrapped around a protein core made up of the histones H2A, H2B, H3 and H4 (core histones). The N-terminal tails of the histones, modifications of which are important in determining the type of chromatin formed, protrude from the surface of the

nucleosome. The 30-nm diameter chromatin fibre, with a packing ratio of 1 in 50, arises through the folding and/or coiling of the nucleosomes together with the linker histone H1. The 30-nm fibre is the basic chromatin fibre in eukaryotes.

The levels of packaging beyond the 30-nm fibre are poorly understood at the molecular and biochemical level because the hierarchical nature of chromatin structure makes these stages hard to study. The radial loop/scaffold model and variations on it is commonly favoured. In this model, the 30-nm fibre is formed into loops that are attached at their bases to a nonhistone protein 'scaffold' (**Figure 2**).

Chromosomes in mitosis

The final level of chromosome packaging occurs as interphase chromosomes enter mitosis or meiosis. The extent of this final level of packing differs among eukaryotes. In yeast there may be only a 2-fold further condensation of the chromatin, whilst a ratio of 9:1 has been calculated for the compaction of mammalian metaphase chromosomes over their interphase counterparts. Phosphorylation of histones H3 and H1 accompanies the condensation of chromosomes as they enter mitosis. Shortening of the chromosome may occur through a helical coiling of the central chromosome axis/scaffold (**Figure 2**). Lateral contraction may result from condensation of the radial loops toward the chromosome scaffold through folding, twisting or coiling (**Figure 2**).

The most abundant proteins of the chromosome scaffold are DNA topoisomerase II (topo II) and ScII, which is a member of the SMC (stable maintenance of chromosomes) family. These have important roles in both chromosome condensation and in holding the sister chromatids together until anaphase.

Chromosomes in the interphase nucleus

By using fluorescence *in situ* hybridization (FISH) with probes for individual chromosomes, it has been shown that each chromosome occupies a distinct globular domain, or territory, within the interphase nucleus. In addition each chromosome territory may have a preferred location within the nucleus. An important consequence of this is that individual DNA sequences within a chromosome are not free to adopt random positions within the nucleus, they are constrained to some extent by the behaviour of the whole chromosome.

Centromeres

The centromere is the site of attachment to the kinetochore, the structure where microtubules emanating from the spindle pole meet the chromosome. The centromere is also the site of sister chromatid attachment until anaphase. The cell has mechanisms to monitor the attachment of chromosomes to the spindle in metaphase. Centromeres must then release sister chromatid attachments and allow the cell to proceed into anaphase. Errors in these processes lead to aberrations in chromosome number (aneuploidy) that can be associated with tumorigenesis or with defects in development (e.g. Down syndrome).

On large chromosomes, such as those in humans, the centromere is visible as a site of constriction on metaphase chromosomes (**Figures 1** and **3**) and the kinetochore can be observed as a trilaminar structure in electron micrographs.

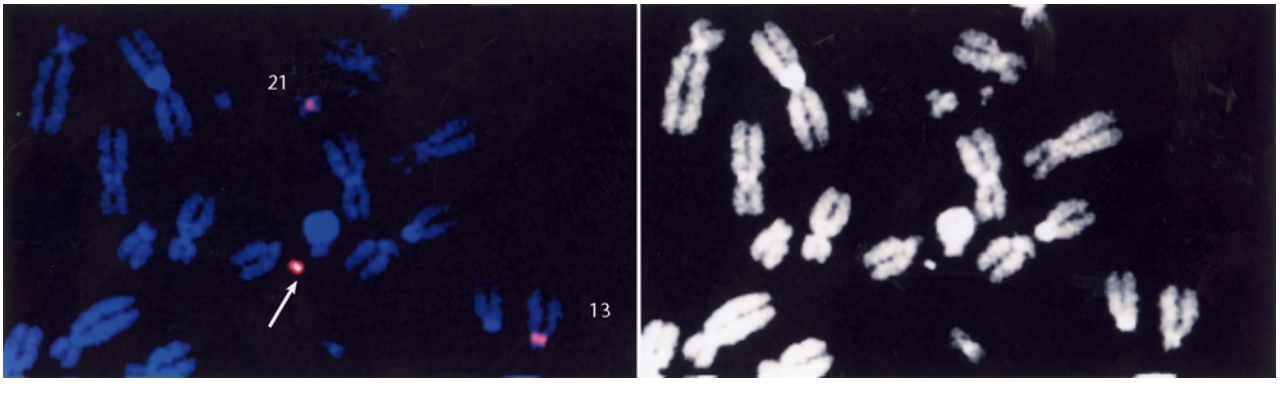

Figure 3 Natural and artificial human chromosomes: spread metaphase chromosomes from a human cell line that also harbours an artificial chromosome. In the left panel the chromosomal DNA is stained in blue and hybridized with a probe for centromeric alpha-satellite DNA (red) from the acrocentric chromosomes nos 13 and 21. This probe also lights up a small chromosome (arrowed). The green signal superimposed on the red on this small chromosome demonstrates it to be derived from a yeast artificial chromosome-based artificial chromosome. The size of this mammalian artificial chromosome (MAC) is more easily seen in the right panel where the DNA staining is shown in black and white. Note the brightly stained heterochromatin visible at the natural centromeres of the human chromosomes. Images courtesy of Dr B. Grimes, MRC Human Genetics Unit, Edinburgh, UK.

Chromosomes in which the centromere is located toward the middle of the chromosome are termed 'metacentric'; those in which the centromere is located toward one end are 'acrocentric'; and those lacking a functional centromere are 'acentric'. Chromosomes from organisms, such as some plants, in which spindle microtubules attach along the chromosome length rather than at a specific site are termed holocentric. Occasionally, chromosomes arise that appear to be dicentric, i.e. they have two centromeres. If both of these centromeres function this can lead to errors in chromosome segregation; usually, however, in such cases only one of the two centromeres is functional while the other is inactivated.

The simple centromeres of budding yeast

Centromeres of the budding yeast (*Saccharomyces cerevisiae*) appear to be quite simple and can be defined by $\sim$200 bp of specific sequence, using an assay for sequences conferring faithful segregation on acentric circular plasmids or linear molecules. The 200-bp region consists of three DNA elements CDEI, II and III. A single base pair mutation in CDEIII can abolish centromere activity. Many specific centromere-binding proteins have been identified in this yeast and they bind to the different elements of the centromere. One of them is CSE4, a protein related to the chromosomal protein histone H3.

All other eukaryotic centromeres analysed so far are complex and not defined by simple sequence motifs. They all appear to involve complex arrays of repetitive DNA and at least some aspects of their function appear to depend on the way in which the DNA sequence is packaged with proteins, rather than on the primary DNA sequence alone. The presence of a condensed form of chromatin called heterochromatin appears to be central to the functions of these centromeres.

The more complex centromeres of the fission yeast

Fission yeast (*Schizosaccharomyces pombe*) centromeres are 50–100 kb in length and consist of inverted DNA repeats surrounding a central core of unique sequence. The central core has an unusual chromatin structure and assembly of heterochromatin at the centromere in this yeast is important for function. Proteins that have motifs found in heterochromatin proteins from many organisms, are found at the centromere and mutations in these proteins compromise centromere function. The presence of underacetylated histones is also important for centromere function.

Centromeres in an invertebrate

Centromeres in the fruitfly (*Drosophila melanogaster*) are also embedded in heterochromatin. Blocks of transposable elements are found among simple sequence repeats over a 500-kb region. In this organism, centromere activity can be experimentally induced to move down the chromosome to sites that do not normally act as centromeres. This argues strongly against a specific short DNA sequence being essential for centromere function in this organism.

Mammalian centromeres

The centromere is visible as the primary constriction on mammalian chromosomes (**Figure 1**). In humans this site is marked by the presence of a particular type of tandemly repeated simple DNA sequence, termed the alpha-satellite (**Figure 3**). Alpha-satellite DNA is a 170-bp sequence repeated over many millions of base pairs of DNA. In the mouse a different, but related, simple sequence (the minor satellite) occupies the equivalent position. In humans, variant chromosomes have shown that the presence of alpha-satellite DNA is neither necessary nor sufficient for centromere activity. Heterochromatin is visible at most mammalian centromeres and the histones in these regions are underacetylated.

The study of proteins located at the centromere is essential to our understanding of centromere function and chromosome segregation. Several mammalian centromere-specific proteins have been identified:

1. CENP-B. This protein binds a short DNA sequence called the CENP-B box. This box is present both human alpha- and mouse minor satellite DNAs. CENP-B binds at both active and inactive centromeres. Disruption of the gene encoding CENP-B in the mouse showed that this protein is not essential for centromere function or cell viability. Its function, if any, remains unknown.
2. CENP-A. Like the protein CSE4 in the budding yeast, CENP-A is a variant version of a histone H3-like protein. It is present only at active centromeres.
3. CENP-C. This is a basic protein that is absent from inactive centromeres. Deletion of the *CenpC* gene in vertebrate cells has shown that this protein is required for proper centromere function.
4. CENP-E. The sequence of this protein suggests that it may act as a motor to move chromosomes along the spindle microtubules.

A variety of other proteins concentrate at centromeric regions of mammalian centromeres, including topoisomerase IIα, which is an enzyme that may be important for chromosome condensation and for separating the strands of sister chromatids. The protein kinases Mad and Bub are also found at mammalian centromeres and may have a role

in monitoring the correct attachment of chromosomes to the mitotic spindle. Mutations in the human *BUB1* gene have been found in cancers characterized by aneuploidy.

Chromosomes at Meiosis

The chromosome at meiosis has several features that make it distinct in structure and behaviour from the mitotic chromosome. First, centromere/kinetochore activity during the two cell divisions of meiosis are quite different. In the first (reduction) division, homologous chromosomes pair and a single kinetochore is formed on each sister chromatid pair, instead of on each sister chromatid as in somatic mitosis. Also the cohesion between sister chromatids is not lost at metaphase/anaphase in this division so that one pair of sister chromatids is segregated to one spindle pole and the other pair to the other pole. This results in the segregation of homologous chromosomes to opposite poles. Specialized proteins are involved in this type of cohesion. In the next division cycle sister chromatids are segregated in a normal fashion. Prior to the segregation of homologous chromosome pairs in the first meiotic division, recombination occurs between the chromosomes and in many organisms this recombination is obligatory for subsequent successful homologous chromosome segregation (disjunction). Recombination occurs through formation of a synaptonemal complex between the paired chromosomes. This complex is visible by electron microscopy and by light microscopy using antibodies against protein components of the structure.

Telomeres

Prokaryotic chromosomes are generally circular molecules, whereas eukaryotes have linear chromosomes. Telomeres are the specialized structures at the ends of these linear DNA molecules. Telomeres 'hide' the ends of the chromosome from the mechanisms within the cell that monitor DNA damage. They are also needed to overcome the problem of end replication. The end-replication problem arises because all known DNA polymerases add nucleotides to a free 3′ OH, i.e. they work only in the 5′ to 3′ direction. Replicative DNA synthesis is primed from an RNA primer that is subsequently removed. At the extreme 3′ end of a linear DNA strand being copied by lagging strand synthesis, removal of the RNA primer and ligation of Okazaki fragments leaves a gap at the end of the new strand. Because this is at the end of the DNA molecule/chromosome there is no DNA template beyond this from which to prime synthesis of DNA across this gap. Hence, without a mechanism to counteract this, the ends of linear DNA molecules would get progressively shorter through subsequent rounds of conventional DNA replication.

Telomeric DNA sequences

Most eukaryotes overcome the end-replication problem with an enzyme called telomerase (a ribonucleoprotein) that uses its own RNA template to add on simple repeats to the 3′ ends of chromosomes, elongating them. Conventional DNA polymerases can use this extended DNA strand as a template on which to synthesize the complementary strand. The telomere sequence added is very similar in a wide variety of eukaryotes. In mammals it is TTAGGG. Large stretches of telomere-like repeats are found at the ends of chromosomes. The telomeres of *Drosophila melanogaster* are unusual in that mobile repetitive sequences (non-long terminal repeat (LTR) retroposons) are found at the ends of the chromosomes rather than tandem repeats of TTAGGG.

Telomerase

Telomerase activity is highest in germline cells. Somatic cells have little telomerase and they have shorter telomeres than cells of the germline. It is estimated that 50–200 bp of DNA is lost from the end of the chromosome per cell division in the absence of telomerase activity. Telomere length also decreases with age and this has led to speculation that telomere shortening may play some role in ageing. Indeed, in the human syndrome of premature ageing telomeres are excessively short.

Mammalian telomerase is a multisubunit ribonucleoprotein complex of > 1000 kDa. Two protein subunits have been identified: p80 (TP1) binds telomerase RNA and p123 (TERT1) is similar to the budding yeast protein EST2p. Mutation of EST2 in yeast results in short telomeres. Both EST2p and TERT1 have strong homology to reverse transcriptases, especially those of non-LTR retroposons. This has led to speculation that telomeres may have evolved from a cellular parasite, analogous to the presence of retroposons at *Drosophila* chromosome ends.

The gene encoding the RNA component of telomerase has been deleted in mice. These mice have shortened telomeres, losing ~ 5 kb per generation, but they do not age prematurely, and in the first few generations they are fertile. By the fourth generation the mice begin to show chromosomal anomalies such as aneuploidy and chromosome fusions, and by the sixth generation they are unable to complete spermatogenesis successfully and the proliferative capacity of organs such as the testis, bone marrow and spleen is compromised. These experiments have demonstrated the role of telomeres and telomerase in both the maintenance of genome integrity and in the viability of organs with high rates of cell division.

Telomere capping

The ends of broken DNA molecules are 'sticky' and tend to fuse together. One job of the telomere is to 'camouflage' the

natural ends of the linear chromosomal DNA molecules from the cellular mechanisms that monitor for DNA damage and also to stop the ends of real chromosomes fusing together. These functions are achieved by DNA-binding proteins such as taz 1 (in yeast) and TRF1 and 2 (in mammals). These proteins bind to telomeres and so sequester away the linear DNA ends and they regulate telomere length through feedback to the telomerase enzyme. Overexpression of TRF1 leads to telomere shortening, whereas abrogation of TRF1 function leads to elongated telomeres. When TRF2 function is perturbed in cells end-to-end fusions of chromosomes are seen.

Senescence and transformation

Telomere shortening in somatic cells and the relationship between telomere length and the proliferative capacity of cells suggest that telomere length may play a role in cellular senescence. Telomere shortening may also be a tumour suppressor mechanism, limiting the potential of transformed cells to divide, at least in humans. Supporting this suggestion, the activity of the catalytic subunit of telomerase (TERT) is increased in human tumours and in immortalized cells, and the misexpression of telomerase is sufficient to extend cellular lifespan beyond the normal senescence point of primary cells and to immortalize them.

Replication Origins

The process of DNA replication must ensure that all of the chromosomal DNA is replicated once, but once only, in each cell cycle. Replication origins are the places where bidirectional DNA replication is initiated on chromosomes. They are spaced at intervals along eukaryotic chromosomes and so divide the genome into domains called replicons. The speed at which DNA polymerases can copy DNA means that the activity of ~ 1500 origins are needed on an average human chromosome if DNA replication is to be completed within the S (synthesis) phase of the cell cycle.

Replication origins are defined by short sequences called ARSs (autonomously replicating sequences) in *S. cerevisiae*, but in other eukaryotes they may be more complex and dependent on chromosome context. There appear to be many more potential origins on the chromosome than are normally needed, or that are actually used at any time.

DNA synthesis does not begin at all origins at the same time. Some regions of chromosomes are replicated early in S phase while the replication of others is delayed until later stages. These regions are clustered together in the human genome and correspond with chromosome bands. The type of chromosome bands that are termed R-bands replicate their DNA early in S phase. Other areas of the chromosome do not begin to replicate their DNA until much later in S phase. The replication time of an entire chromosome can change in special circumstances. For example, inactivation of an X chromosome in the somatic cells of female mammals is accompanied by late replication of this chromosome compared with its active counterpart in the same cells.

The completion of S phase is usually coupled to mitosis, ensuring that DNA replication cannot begin again until after a cell division and that mitosis cannot start until synthesis is completed. Exceptions to this rule include the overamplification of discrete regions of the genome commonly found in human tumour cells and the amplification of the entire genome in the absence of intervening mitoses to form the large polytene chromosomes of dipteran insects.

Artificial Chromosomes

The ability to form artificial chromosomes is an important prerequisite for understanding chromosome biology and for fully exploiting the potential for genetic manipulation within an organism. Since the essential elements of a chromosome (centromere, telomere and replication origin) have been defined by small DNA elements in *S. cerevisiae* it has been possible to create yeast artificial chromosomes (YACs). These have proved to be invaluable cloning vectors for large fragments of DNA. In most other eukaryotes such simple and easily manipulated artificial chromosomes have not been produced.

An important and prominent goal for many years has been to create a mammalian artificial chromosome (MAC). Recently, this has been partially achieved. Small human microchromosomes have been created by a process of chromosome fragmentation, i.e. by chipping away at an endogenous chromosome until the bare minimum for stability is left. However, such chromosomes are still too large (> 2 Mb) to be readily usable as cloning vectors. Functioning human artificial chromosomes have been put together from component parts in two different ways. In the first, megabase-long arrays of human alpha-satellite DNA were mixed in a test-tube with telomeric sequences $(TTAGGG)_n$ and with random human genomic DNA. The role of the latter is unclear; it may provide replication origins or just act as a stuffer to give the chromosomes bulk. This mixture of DNA molecules was then introduced into human cells, and in some cells functioning and stable minichromosomes were generated by ligation events within the cell. In the second approach, human alpha-satellite DNA and telomere sequences were covalently joined to one another within a YAC and this YAC was then introduced into human cells. In both of these approaches the resulting artificial chromosomes generated are visible under the light microscope (**Figure 3**). This indicates that they must be ~ 5 Mb in length and hence are considerably

longer than might have been predicted from the input DNA. It is unclear what events occurred within the human cells that rearranged the input DNAs and that led to the formation of artificial chromosomes. These uncertainties preclude the current use of such systems as readily usable cloning vectors but they do represent definitive steps toward the goal of a MAC.

As the discovery of the sequence of the human genome nears completion there is a growing focus away from considering gene sequences in isolation and towards thinking about genes and gene function within the context of whole chromosomes. In particular the chromosome and its structure may have an important role in providing the appropriate environments for correct gene expression during development. Abnormalities in the function of telomeres and centromeres are also of increasing importance in the understanding of cancer and as targets for potential anti-cancer therapies. In the future a better understanding of chromosome structure and function may enable the design of artificial chromosome based gene therapies.

Further Reading

Bickmore W and Craig J (1997) *Chromosome Bands: Patterns in the Genome*, chap. 1, 4 and 5. Heidelberg: Springer-Verlag.

Blow JJ (1996) *Eukaryotic DNA Replication. Frontiers in Molecular Biology*. Oxford: Oxford University Press.

Greider CW (1998) Telomeres and senescence: the history, the experiment, the future. *Current Biology* **8**: R178–R181.

Kipling D (1995) *The Telomere*. Oxford: Oxford University Press.

Rosenfeld MA (1997) Human artificial chromosomes get real. *Nature Genetics* **15**: 333–335.

Wolffe A (1995) *Chromatin: Structure and Function*, chap 2. London: Academic Press.

Wynford-Thomas D and Kipling D (1997) Cancer and the knockout mouse. *Nature* **389**: 551–552.