CHROMOSOMES

5.19.1 Introduction

Key Terms

The **nucleoid** is the region in a prokaryotic cell that contains the genome. The DNA is bound to proteins and is not enclosed by a membrane.

Chromatin describes the state of nuclear DNA and its associated proteins during the interphase (between mitoses) of the eukaryotic cell cycle.

A **chromosome** is a discrete unit of the genome carrying many genes. Each chromosome consists of a very long molecule of duplex DNA and an approximately equal mass of proteins. It is visible as a morphological entity only during cell division.

The **packing ratio** is the ratio of the length of DNA to the unit length of the fiber containing it.

A general principle is evident in the organization of all cellular genetic material. It exists as a compact mass, confined to a limited volume; and its various activities, such as replication and transcription, must be accomplished within this space. The organization of this material must accommodate transitions between inactive and active states.

The condensed state of nucleic acid results from its binding to basic proteins. The positive charges of these proteins neutralize the negative charges of the nucleic acid. The structure of the nucleoprotein complex is determined by the interactions of the proteins with the DNA (or RNA).

A common problem is presented by the packaging of DNA into phages and viruses, into bacterial cells and eukaryotic nuclei. The length of the DNA as an extended molecule would vastly exceed the dimensions of the compartment that contains it. The DNA (or in the case of some viruses, the RNA) must be compressed exceedingly tightly to fit into the space available. *So in contrast with the customary picture of DNA as an extended double helix, structural deformation of DNA to bend or fold it into a more compact form is the rule rather than exception.*

The magnitude of the discrepancy between the length of the nucleic acid and the size of its compartment is evident from the examples summarized in **Figure 19.1**. For bacteriophages and for eukaryotic viruses, the nucleic acid genome, whether single-stranded or double-stranded DNA or RNA, effectively fills the container (which can be rod-like or spherical).

DNA is highly compressed in all types of genomes				
Compartment	Shape	Dimensions	Type of Nucleic Acid	Length
TMV	filament	0.008 x 0.3 mm	1 single-stranded RNA	2 mm = 6.4 kb
Phage fd	filament	0.006 x 0.85 mm	1 single-stranded DNA	2 mm = 6.0 kb
Adenovirus	icosahedron	0.07 mm diameter	1 double-stranded DNA	11 mm = 35.0 kb
Phage T4	icosahedron	0.065 x 0.10 mm	1 double-stranded DNA	55 mm = 170.0 kb
E. coli	cylinder	1.7 x 0.65 mm	1 double-stranded DNA	1.3 mm = 4.2×10^3 kb
Mitochondrion (human)	oblate spheroid	3.0 x 0.5 mm	~10 identical double-stranded DNAs	50 mm = 16.0 kb
Nucleus (human)	spheroid	6 mm diameter	46 chromosomes of double-stranded DNA	1.8 m = 6×10^6 kb

Figure 19.1 The length of nucleic acid is much greater than the dimensions of the surrounding compartment.

For bacteria or for eukaryotic cell compartments, the discrepancy is hard to calculate exactly, because the DNA is contained in a compact area that occupies only part of the compartment. The genetic material is seen in the form of the **nucleoid** in bacteria and as the mass of **chromatin** in eukaryotic nuclei at interphase (between divisions).

The density of DNA in these compartments is high. In a bacterium it is ~10 mg/ml, in a eukaryotic nucleus it is ~100 mg/ml., and in the phage T4 head it is >500 mg/ml. Such a concentration in solution would be equivalent to a gel of great viscosity. We do not entirely understand the physiological implications, for example, what effect this has upon the ability of proteins to find their binding sites on DNA.

The packaging of chromatin is flexible; it changes during the eukaryotic cell cycle. At the time of division (mitosis or meiosis), the genetic material becomes even more tightly packaged, and individual **chromosomes** become recognizable.

The overall compression of the DNA can be described by the **packing ratio**, the length of the DNA divided by the length of the unit that contains it. For example, the smallest human chromosome contains ~4.6×10^7 bp of DNA (~10 times the genome size of the bacterium *E. coli)*. This is equivalent to 14,000 µm (= 1.4 cm) of extended DNA. At the most condensed moment of mitosis, the chromosome is ~2 µm long. So the packing ratio of DNA in the chromosome can be as great as 7000.

Packing ratios cannot be established with such certainty for the more amorphous overall structures of the bacterial nucleoid or eukaryotic chromatin. However, the usual reckoning is that mitotic chromosomes are likely to be 5-10× more tightly packaged than interphase chromatin, which therefore has a typical packing ratio of 1000-2000.

A major unanswered question concerns the *specificity* of packaging. Is the DNA folded into a *particular* pattern, or is it different in each individual copy of the genome? How does the pattern of packaging change when a segment of DNA is replicated or transcribed?

CHROMOSOMES

5.19.2 Viral genomes are packaged into their coats

Key Terms

A **capsid** is the external protein coat of a virus particle.

Icosahedral symmetry is typical of viruses that have capsids that are polyhedrons.

The **nucleation center** of TMV (tobacco mosaic virus) is a duplex hairpin where assembly of coat protein with RNA is initiated.

A **terminase** enzyme cleaves multimers of a viral genome and then uses hydrolysis of ATP to provide the energy to translocate the DNA into an empty viral capsid starting with the cleaved end.

Monster particles of bacteriophages form as the result of an assembly defect in which the capsid proteins form a head that is much longer than usual.

An **assembly factor** is a protein that is required for formation of a macromolecular structure but is not itself part of that structure.

Key Concepts

- The length of DNA that can be incorporated into a virus is limited by the structure of the head shell.
- Nucleic acid within the head shell is extremely condensed.
- Filamentous RNA viruses condense the RNA genome as they assemble the head shell around it.
- Spherical DNA viruses insert the DNA into a preassembled protein shell.

From the perspective of packaging the *individual* sequence, there is an important difference between a cellular genome and a virus. The cellular genome is essentially indefinite in size; the number and location of individual sequences can be changed by duplication, deletion, and rearrangement. So it requires a *generalized* method for packaging its DNA, insensitive to the total content or distribution of sequences. By contrast, two restrictions define the needs of a virus. The amount of nucleic acid to be packaged is *predetermined* by the size of the genome. And it must all fit within a coat assembled from a protein or proteins coded by the viral genes.

A virus particle is deceptively simple in its superficial appearance. The nucleic acid genome is contained within a **capsid**, a symmetrical or quasi-symmetrical structure assembled from one or only a few proteins. Attached to the capsid, or incorporated into it, are other structures, assembled from distinct proteins, and necessary for infection of the host cell.

The virus particle is tightly constructed. The internal volume of the capsid is rarely much greater than the volume of the nucleic acid it must hold. The difference is usually less than twofold, and often the internal volume is barely larger than the

nucleic acid.

In its most extreme form, the restriction that the capsid must be assembled from proteins coded by the virus means that the entire shell is constructed from a single type of subunit. The rules for assembly of identical subunits into closed structures restrict the capsid to one of two types. The protein subunits stack sequentially in a helical array to form a *filamentous* or rod-like shape. Or they form a pseudospherical shell, a type of structure that conforms to a polyhedron with **icosahedral symmetry**. Some viral capsids are assembled from more than a single type of protein subunit, but although this extends the exact types of structures that can be formed, viral capsids still all conform to the general classes of quasi-crystalline filaments or icosahedrons (590).

There are two types of solution to the problem of how to construct a capsid that contains nucleic acid:

- The protein shell can be assembled around the nucleic acid, condensing the DNA or RNA by protein-nucleic acid interactions during the process of assembly.
- Or the capsid can be constructed from its component(s) in the form of an empty shell, into which the nucleic acid must be inserted, being condensed as it enters.

The capsid is assembled around the genome for single-stranded RNA viruses. The principle of assembly is that *the position of the RNA within the capsid is determined directly by its binding to the proteins of the shell.* The best characterized example is TMV (tobacco mosaic virus). Assembly starts at a duplex hairpin that lies within the RNA sequence (3157; 3158). From this **nucleation center**, it proceeds bidirectionally along the RNA, until reaching the ends. The unit of the capsid is a two-layer disk, each layer containing 17 identical protein subunits. The disk is a circular structure, which forms a helix as it interacts with the RNA. At the nucleation center, the RNA hairpin inserts into the central hole in the disk, and the disk changes conformation into a helical structure that surrounds the RNA. Then further disks are added, each disk pulling a new stretch of RNA into its central hole. The RNA becomes coiled in a helical array on the inside of the protein shell, as illustrated in **Figure 19.2** (589; for review see 3159; 3160).

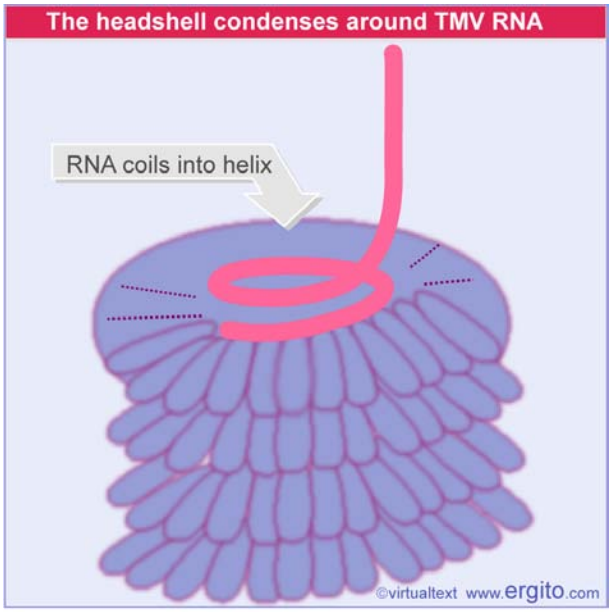

Figure 19.2 A helical path for TMV RNA is created by the stacking of protein subunits in the virion.

The spherical capsids of DNA viruses are assembled in a different way, as best characterized for the phages lambda and T4. In each case, an empty headshell is assembled from a small set of proteins. *Then the duplex genome is inserted into the head,* accompanied by a structural change in the capsid.

Figure 19.3 summarizes the assembly of lambda. It starts with a small headshell that contains a protein "core". This is converted to an empty headshell of more distinct shape. Then DNA packaging begins, the headshell expands in size though remaining the same shape, and finally the full head is sealed by addition of the tail.

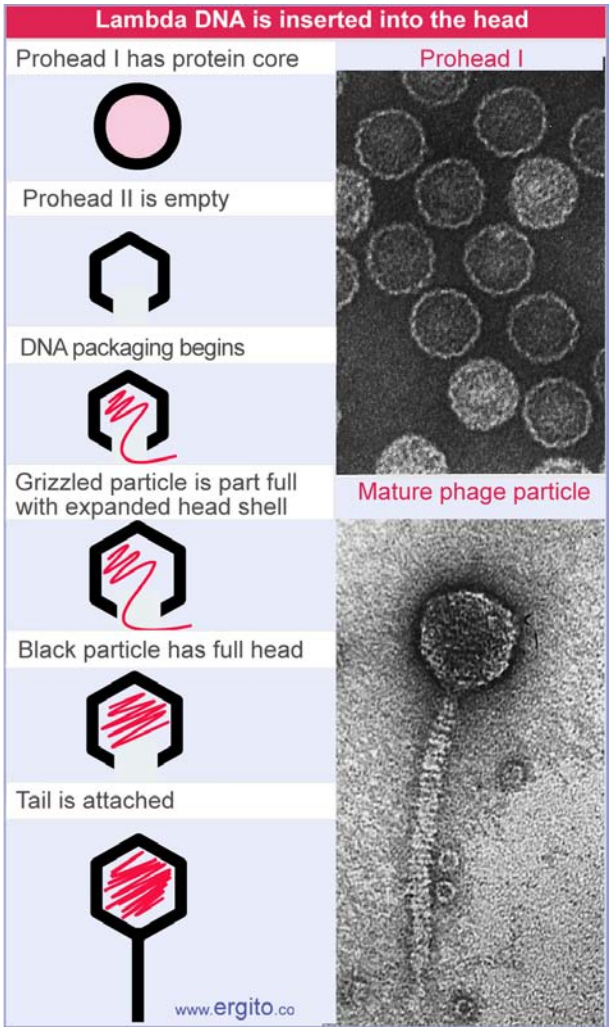

Figure 19.3 Maturation of phage lambda passes through several stages. The empty head changes shape and expands when it becomes filled with DNA. The electron micrographs show the particles at the start and end of the maturation pathway. Photographs kindly provided by A. F. Howatson.

Now a double-stranded DNA considered over short distances is a fairly rigid rod. Yet it must be compressed into a compact structure to fit within the capsid. We should like to know whether packaging involves a smooth coiling of the DNA into the head or requires abrupt bends.

Inserting DNA into a phage head involves two types of reaction: translocation and condensation. Both are energetically unfavorable.

Translocation is an active process in which the DNA is driven into the head by an ATP-dependent mechanism. A common mechanism is used for many viruses that replicate by a rolling circle mechanism to generate long tails that contain multimers of the viral genome. The best characterized example is phage lambda. The genome is packaged into the empty capsid by the **terminase** enzyme. **Figure 19.4** summarizes the process.

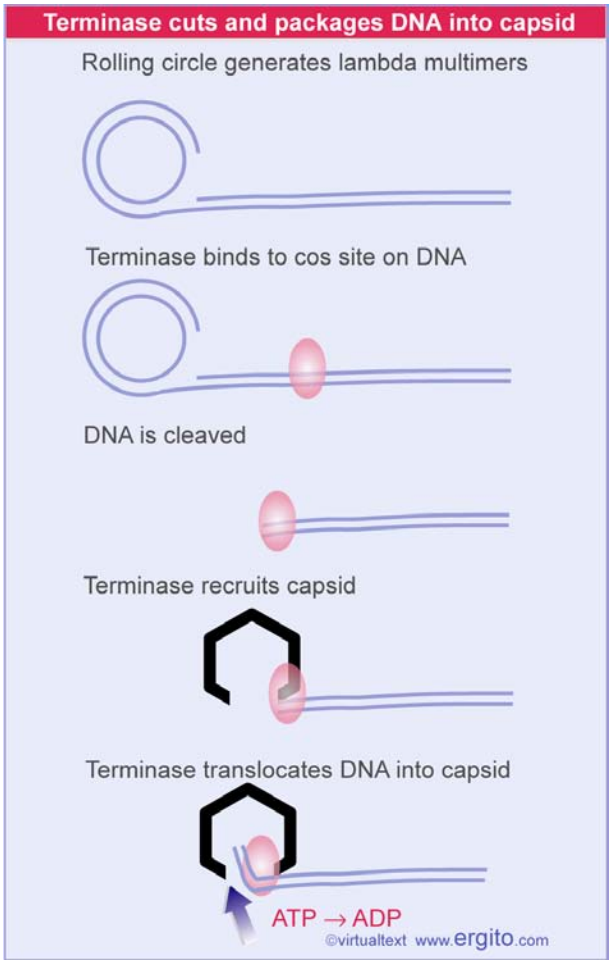

Figure 19.4 Terminase protein binds to specific sites on a multimer of virus genomes generated by rolling circle replication. Terminase cuts the DNA and binds to an empty virus capsid. It uses energy from hydrolysis of ATP to insert the DNA into the capsid.

The terminase was first recognized for its role in generating the ends of the linear phage DNA by cleaving at *cos* sites. (The name *cos* reflects the fact that it generates cohesive ends that have complementary single-stranded tails.) The phage genome codes two subunits that make up the terminase. One subunit binds to a *cos* site; then it is joined by the other subunit, which cuts the DNA. The terminase assembles into a hetero-oligomer in a complex that also includes IHF (the integration host factor coded by the bacterial genome) (3161). It then binds to an empty capsid and uses ATP hydrolysis to power translocation along the DNA. The translocation drives the DNA into the empty capsid.

Another method of packaging uses a structural component of the phage. In the *B. subtilis* phage ϕ29, the motor that inserts the DNA into the phage head is the structure that connects the head to the tail. It functions as a rotary motor, where the motor action effects the linear translocation of the DAN into the phage head (3163; 3162). The same motor is used to eject the DNA from the phage head when it infects a bacterium.

Little is known about the mechanism of condensation into an empty capsid, except that the capsid contains "internal proteins" as well as DNA. One possibility is that they provide some sort of "scaffolding" onto which the DNA condenses. (This would be a counterpart to the use of the proteins of the shell in the plant RNA viruses.)

How specific is the packaging? It cannot depend on particular sequences, because deletions, insertions, and substitutions all fail to interfere with the assembly process. The relationship between DNA and the headshell has been investigated directly by determining which regions of the DNA can be chemically crosslinked to the proteins of the capsid. The surprising answer is that all regions of the DNA are more or less equally susceptible. This probably means that when DNA is inserted into the head, it follows a general rule for condensing, but the pattern is not determined by particular sequences (for review see 192).

These varying mechanisms of virus assembly all accomplish the same end: packaging a single DNA or RNA molecule into the capsid. However, some viruses have genomes that consist of multiple nucleic acid molecules. Reovirus contains ten double-stranded RNA segments, all of which must be packaged into the capsid. Specific sorting sequences in the segments may be required to ensure that the assembly process selects one copy of each different molecule in order to collect a complete set of genetic information. In the simpler case of phage ϕ6, which packages three different segments of double-stranded RNA into one capsid, the RNA segments must bind in a specific order; as each is incorporated into the capsid, it triggers a change in the conformation of the capsid that creates binding sites for the next segment (for review see 3164).

Some plant viruses are multipartite: their genomes consist of segments, each of which is packaged into a *different* capsid. An example is alfalfa mosaic virus (AMV), which has four different single-stranded RNAs, each packaged independently into a coat comprising the same protein subunit. A successful infection depends on the entry of one of each type into the cell.

The four components of AMV exist as particles of different sizes. This means that the same capsid protein can package each RNA into its own characteristic particle. This is a departure from the packaging of a unique length of nucleic acid into a capsid of fixed shape.

The assembly pathway of viruses whose capsids have only one authentic form may be diverted by mutations that cause the formation of aberrant **monster** particles in which the head is longer than usual. These mutations show that a capsid protein(s) has an intrinsic ability to assemble into a particular type of structure, but the exact size and shape may vary. Some of the mutations occur in genes that code for **assembly factors**, which are needed for head formation, but are not themselves part of the headshell. Such ancillary proteins limit the options of the capsid protein so that it assembles only along the desired pathway. Comparable proteins are employed in the assembly of cellular chromatin (see *Molecular Biology 5.20 Nucleosomes*).

Last updated on 12-10-2002

Reviews

192. Black, L. W. (1989). *DNA packaging in dsDNA bacteriophages*. Annu. Rev. Immunol. 43, 267-292.

3159. Klug, A. (1999). *The tobacco mosaic virus particle: structure and assembly.* Philos. Trans. R. Soc. Lond. B Biol. Sci. 354, 531-535.

3160. Butler, P. J. (1999). *Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed.* Philos. Trans. R. Soc. Lond. B Biol. Sci. 354, 537-550.

3164. Mindich, L. (2000). *Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage phi6.* Microbiol. Mol. Biol. Rev. 63, 149-160.

References

589. Fraenkel-Conrat, H. and Williams, R. C. (1955). *Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components*. Proc. Natl. Acad. Sci. USA 41, 690-698.

590. Caspar, D. L. D. and Klug, A. (1962). *Physical principles in the construction of regular viruses*. Cold Spring Harbor Symp. Quant. Biol. 27, 1-24.

3157. Zimmern, D. and Butler, P. J. (1977). *The isolation of tobacco mosaic virus RNA fragments containing the origin for viral assembly.* Cell 11, 455-462.

3158. Zimmern, D. (1977). *The nucleotide sequence at the origin for assembly on tobacco mosaic virus RNA.* Cell 11, 463-482.

3161. de Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M., and Catalano, C. E. (2002). *Insights into specific DNA recognition during the assembly of a viral genome packaging machine.* Mol. Cell 9, 981-991.

3162. Jiang, Y. J., Aerne, B. L., Smithers, L., Haddon, C., Ish-Horowicz, D., and Lewis, J. (2000). *Notch signalling and the synchronization of the somite segmentation clock.* Nature 408, 475-479.

3163. Dube, P., Tavares, P., Lurz, R., and van Heel, M. (1993). *The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry.* EMBO J. 12, 1303-1309.

CHROMOSOMES

5.19.3 The bacterial genome is a nucleoid

Key Terms

The **nucleoid** is the region in a prokaryotic cell that contains the genome. The DNA is bound to proteins and is not enclosed by a membrane.

Key Concepts

- The bacterial nucleoid is ~80% DNA by mass and can be unfolded by agents that act on RNA or protein.
- The proteins that are responsible for condensing the DNA have not been identified.

Although bacteria do not display structures with the distinct morphological features of eukaryotic chromosomes, their genomes nonetheless are organized into definite bodies. The genetic material can be seen as a fairly compact clump or series of clumps that occupies about a third of the volume of the cell. **Figure 19.5** displays a thin section through a bacterium in which this **nucleoid** is evident (for review see 191).

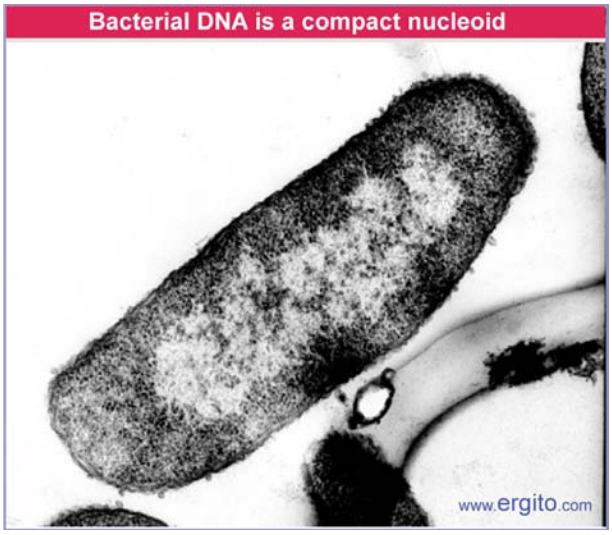

Figure 19.5 A thin section shows the bacterial nucleoid as a compact mass in the center of the cell. Photograph kindly provided by Jack Griffith.

When *E. coli* cells are lysed, fibers are released in the form of loops attached to the broken envelope of the cell. As can be seen from **Figure 19.6**, the DNA of these loops is not found in the extended form of a free duplex, but is compacted by association with proteins.

Figure 19.6 The nucleoid spills out of a lysed *E. coli* cell in the form of loops of a fiber. Photograph kindly provided by Jack Griffith.

Several DNA-binding proteins with a superficial resemblance to eukaryotic chromosomal proteins have been isolated in *E. coli*. What criteria should we apply for deciding whether a DNA-binding protein plays a structural role in the nucleoid? It should be present in sufficient quantities to bind throughout the genome. And mutations in its gene should cause some disruption of structure or of functions associated with genome survival (for example, segregation to daughter cells). None of the candidate proteins yet satisfies the genetic conditions (for review see 190).

Protein HU is a dimer that condenses DNA, possibly wrapping it into a bead-like structure. It is related to IHF (integration host factor), another dimer, which has a structural role in building a protein complex in specialized recombination reactions. Null mutations in either of the genes coding for the subunits of HU (*hupA,B*) have little effect, but loss of both functions causes a cold-sensitive phenotype and some loss of superhelicity in DNA. These results raise the possibility that HU plays some general role in nucleoid condensation.

Protein H1 (also known as H-NS) binds DNA, interacting preferentially with sequences that are bent. Mutations in its gene have turned up in a variety of guises (*osmZ, bglY, pilG*), each identified as an apparent regulator of a different system.

These results probably reflect the effect that H1 has on the local topology of DNA, with effects upon gene expression that depend upon the particular promoter.

We might expect that the absence of a protein required for nucleoid structure would have serious effects upon viability. Why then are the effects of deletions in the genes for proteins HU and H1 relatively restricted? One explanation is that these proteins are *redundant,* that any one can substitute for the others, so that deletions of *all* of them would be necessary to interfere seriously with nucleoid structure. Another possibility is that we have yet to identify the proteins responsible for the major features of nucleoid integrity.

The nucleoid can be isolated directly in the form of a very rapidly sedimenting complex, consisting of ~80% DNA by mass. (The analogous complexes in eukaryotes have ~50% DNA by mass; see *Molecular Biology 5.19.4 The bacterial genome is supercoiled*.) It can be unfolded by treatment with reagents that act on RNA or protein. The possible role of proteins in stabilizing its structure is evident. The role of RNA has been quite refractory to analysis.

Reviews

190. Drlica, K. and Rouviere-Yaniv, J. (1987). *Histone-like proteins of bacteria*. Microbiol. Rev. 51, 301-319.

191. Brock, T. D. (1988). *The bacterial nucleus: a history*. Microbiol. Rev. 52, 397-411.

CHROMOSOMES

5.19.4 The bacterial genome is supercoiled

Key Terms

A **domain** of a chromosome may refer *either* to a discrete structural entity defined as a region within which supercoiling is independent of other domains; *or* to an extensive region including an expressed gene that has heightened sensitivity to degradation by the enzyme DNAase I.

Key Concepts

- The nucleoid has ~100 independently negatively supercoiled domains.
- The average density of supercoiling is ~1 turn/100 bp.

The DNA of the bacterial nucleoid isolated *in vitro* behaves as a closed duplex structure, as judged by its response to ethidium bromide. This small molecule intercalates between base pairs to generate *positive* superhelical turns in "closed" circular DNA molecules, that is, molecules in which both strands have covalent integrity. (In "open" circular molecules, which contain a nick in one strand, or with linear molecules, the DNA can rotate freely in response to the intercalation, thus relieving the tension.)

In a natural closed DNA that is *negatively* supercoiled, the intercalation of ethidium bromide first removes the negative supercoils and then introduces positive supercoils. The amount of ethidium bromide needed to achieve zero supercoiling is a measure of the original density of negative supercoils.

Some nicks occur in the compact nucleoid during its isolation; they can also be generated by limited treatment with DNAase. But this does not abolish the ability of ethidium bromide to introduce positive supercoils. This capacity of the genome to retain its response to ethidium bromide in the face of nicking means that it must have many independent chromosomal **domains**; *the supercoiling in each domain is not affected by events in the other domains.*

This autonomy suggests that the structure of the bacterial chromosome has the general organization depicted diagrammatically in **Figure 19.7**. Each domain consists of a loop of DNA, the ends of which are secured in some (unknown) way that does not allow rotational events to propagate from one domain to another. There are ~100 such domains per genome; each consists of ~40 kb (13 µm) of DNA, organized into some more compact fiber whose structure has yet to be characterized.

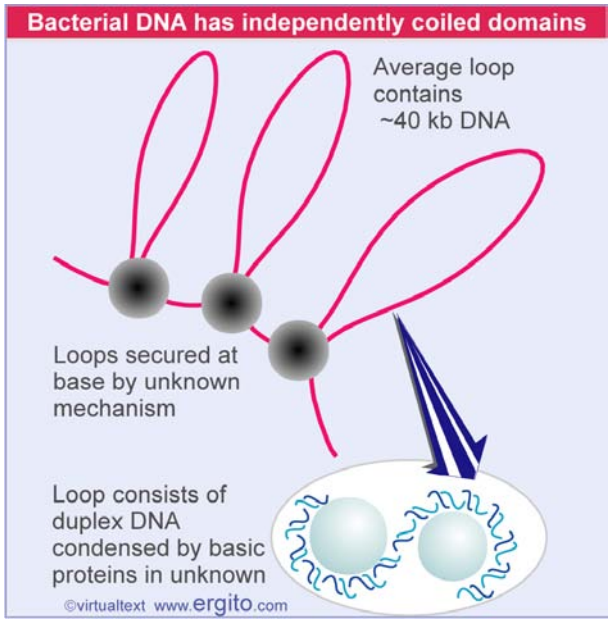

Figure 19.7 The bacterial genome consists of a large number of loops of duplex DNA (in the form of a fiber), each secured at the base to form an independent structural domain.

The existence of separate domains could permit different degrees of supercoiling to be maintained in different regions of the genome. This could be relevant in considering the different susceptibilities of particular bacterial promoters to supercoiling (see *Molecular Biology 3.9.15 Supercoiling is an important feature of transcription*).

Supercoiling in the genome can in principle take either of two forms, as summarized in in **Figure 19.8**:

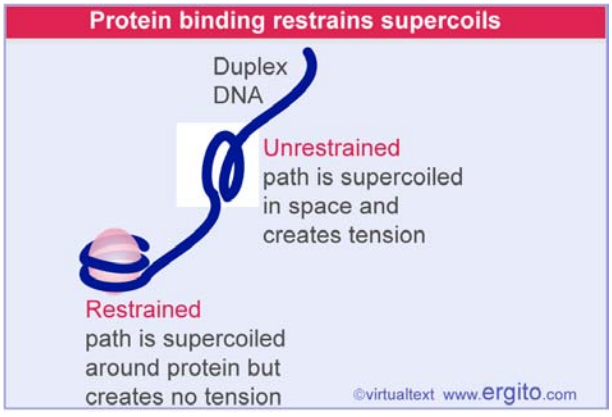

Figure 19.8 An unrestrained supercoil in the DNA path creates tension, but no tension is transmitted along DNA when a supercoil is restrained by protein binding.

- If a supercoiled DNA is free, its path is *unrestrained*, and negative supercoils generate a state of torsional tension that is transmitted freely along the DNA

within a domain. It can be relieved by unwinding the double helix, as described in *Molecular Biology 4.15.12 Supercoiling affects the structure of DNA*. The DNA is in a dynamic equilibrium between the states of tension and unwinding.

- Supercoiling can be *restrained* if proteins are bound to the DNA to hold it in a particular three-dimensional configuration. In this case, the supercoils are represented by the path the DNA follows in its fixed association with the proteins. The energy of interaction between the proteins and the supercoiled DNA stabilizes the nucleic acid, so that no tension is transmitted along the molecule.

Are the supercoils in *E. coli* DNA restrained *in vivo* or is the double helix subject to the torsional tension characteristic of free DNA? Measurements of supercoiling *in vitro* encounter the difficulty that restraining proteins may have been lost during isolation. Various approaches suggest that DNA is under torsional stress *in vivo*.

One approach is to measure the effect of nicking the DNA. Unrestrained supercoils are released by nicking, but restrained supercoils are unaffected. Nicking releases ~50% of the overall supercoiling, suggesting that about half of the supercoiling is transmitted as tension along DNA, but the other half is absorbed by protein binding (3245).

Another approach uses the crosslinking reagent psoralen, which binds more readily to DNA when it is under torsional tension. The reaction of psoralen with *E. coli* DNA *in vivo* corresponds to an average density of one negative superhelical turn / 200 bp ($\sigma = -0.05$).

We can also examine the ability of cells to form alternative DNA structures; for example, to generate cruciforms at palindromic sequences. From the change in linking number that is required to drive such reactions, it is possible to calculate the original supercoiling density. This approach suggests an average density of $\sigma = -0.025$, or one negative superhelical turn / 100 base pairs.

So supercoils *do* create torsional tension *in vivo*. There may be variation about an average level, and although the precise range of densities is difficult to measure, it is clear that the level is sufficient to exert significant effects on DNA structure, for example, in assisting melting in particular regions such as origins or promoters (for review see 3235).

Many of the important features of the structure of the compact nucleoid remain to be established. What is the specificity with which domains are constructed – do the same sequences always lie at the same relative locations, or can the contents of individual domains shift? How is the integrity of the domain maintained? Biochemical analysis by itself is unable to answer these questions fully, but if it is possible to devise suitable selective techniques, the properties of structural mutants should lead to a molecular analysis of nucleoid construction.

Reviews

3235. Hatfield, G. W. and Benham, C. J. (2002). *DNA topology-mediated control of global gene expression in Escherichia coli.* Annu. Rev. Genet. 36, 175-203.

References

3245. Pettijohn, D. E. and Pfenninger, O. (1980). *Supercoils in prokaryotic DNA restrained in vitro.* Proc. Natl. Acad. Sci. USA 77, 1331-1335.

CHROMOSOMES

5.19.5 Eukaryotic DNA has loops and domains attached to a scaffold

Key Terms

A chromosome **scaffold** is a proteinaceous structure in the shape of a sister chromatid pair, generated when chromosomes are depleted of histones.

Key Concepts

- DNA of interphase chromatin is negatively supercoiled into independent domains of ~85 kb.
- Metaphase chromosomes have a protein scaffold to which the loops of supercoiled DNA are attached.

Interphase chromatin is a tangled mass occupying a large part of the nuclear volume, in contrast with the highly organized and reproducible ultrastructure of mitotic chromosomes. What controls the distribution of interphase chromatin within the nucleus?

Some indirect evidence on its nature is provided by the isolation of the genome as a single, compact body. Using the same technique that was developed for isolating the bacterial nucleoid (see *Molecular Biology 5.19.4 The bacterial genome is supercoiled*), nuclei can be lysed on top of a sucrose gradient. This releases the genome in a form that can be collected by centrifugation. As isolated from *D. melanogaster*, it can be visualized as a compactly folded fiber (10 nm in diameter), consisting of DNA bound to proteins.

Supercoiling measured by the response to ethidium bromide corresponds to about one negative supercoil / 200 bp. These supercoils can be removed by nicking with DNAase, although the DNA remains in the form of the 10 nm fiber. This suggests that the supercoiling is caused by the arrangement of the fiber in space, and represents the existing torsion.

Full relaxation of the supercoils requires one nick / 85 kb, identifying the average length of "closed" DNA. This region could comprise a loop or domain similar in nature to those identified in the bacterial genome. Loops can be seen directly when the majority of proteins are extracted from mitotic chromosomes. The resulting complex consists of the DNA associated with ~8% of the original protein content. As seen in **Figure 19.9**, the protein-depleted chromosomes take the form of a central **scaffold** surrounded by a halo of DNA.

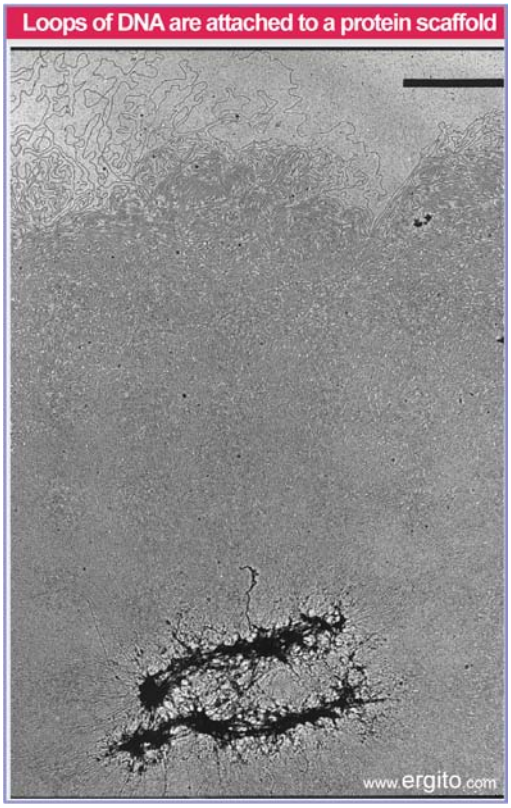

Figure 19.9 Histone-depleted chromosomes consist of a protein scaffold to which loops of DNA are anchored. Photograph kindly provided by Ulrich K. Laemmli.

The metaphase scaffold consists of a dense network of fibers. Threads of DNA emanate from the scaffold, apparently as loops of average length 10-30 µm (30-90 kb). The DNA can be digested without affecting the integrity of the scaffold, which consists of a set of specific proteins. This suggests a form of organization in which loops of DNA of ~60 kb are anchored in a central proteinaceous scaffold.

The appearance of the scaffold resembles a mitotic pair of sister chromatids. The sister scaffolds usually are tightly connected, but sometimes are separate, joined only by a few fibers. Could this be the structure responsible for maintaining the shape of the mitotic chromosomes? Could it be generated by bringing together the protein components that usually secure the bases of loops in interphase chromatin?

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.19.5

CHROMOSOMES

5.19.6 Specific sequences attach DNA to an interphase matrix

Key Terms

A **matrix attachment site (MAR)** is a region of DNA that attaches to the nuclear matrix. It is also known as a scaffold attachment site (SAR).

Key Concepts

- DNA is attached to the nuclear matrix at specific sequences called MARs or SARs.
- The MARs are A·T-rich but do not have any specific consensus sequence.

Is DNA attached to the scaffold via specific sequences? DNA sites attached to proteinaceous structures in interphase nuclei are called **MAR** (matrix attachment regions); they are sometimes also called *SAR* (scaffold attachment regions). The nature of the structure in interphase cells to which they are connected is not clear. Chromatin often appears to be attached to a matrix, and there have been many suggestions that this attachment is necessary for transcription or replication. When nuclei are depleted of proteins, the DNA extrudes as loops from a residual proteinaceous structure. However, attempts to relate the proteins found in this preparation to structural elements of intact cells have not been successful.

Are particular DNA regions associated with this matrix? *In vivo* and *in vitro* approaches are summarized in **Figure 19.10**. Both start by isolating the matrix as a crude nuclear preparation containing chromatin and nuclear proteins. Different treatments can then be used to characterize DNA in the matrix or to identify DNA able to attach to it.

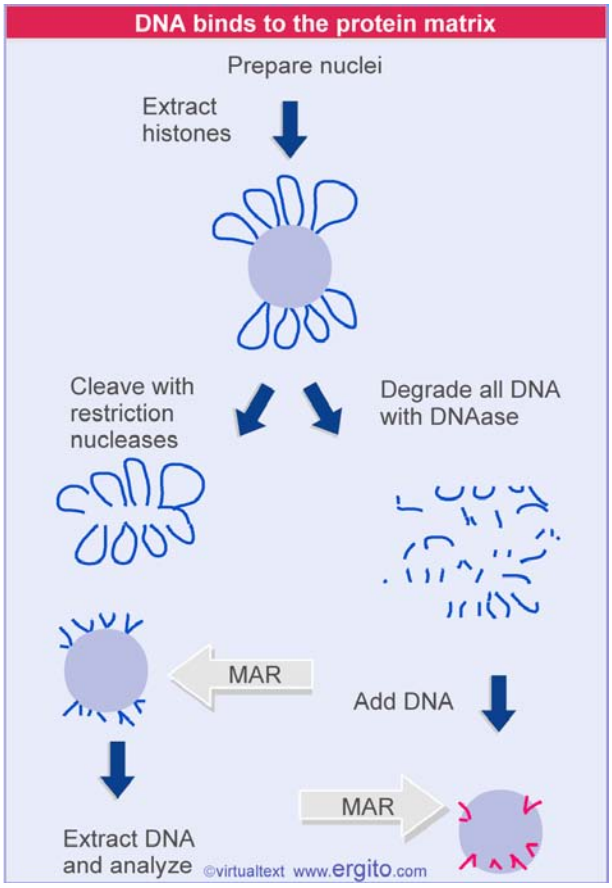

Figure 19.10 Matrix-associated regions may be identified by characterizing the DNA retained by the matrix isolated *in vivo* or by identifying the fragments that can bind to the matrix from which all DNA has been removed

To analyze the existing MAR, the chromosomal loops can be decondensed by extracting the proteins. Removal of the DNA loops by treatment with restriction nucleases leaves only the (presumptive) *in vivo* MAR sequences attached to the matrix.

The complementary approach is to remove *all* the DNA from the matrix by treatment with DNAase; then isolated fragments of DNA can be tested for their ability to bind to the matrix *in vitro.*

The same sequences should be associated with the matrix *in vivo* or *in vitro.* Once a potential MAR has been identified, the size of the minimal region needed for association *in vitro* can be determined by deletions. We can also then identify proteins that bind to the MAR sequences.

A surprising feature is the lack of conservation of sequence in MAR fragments. They are usually ~70% A·T-rich, but otherwise lack any consensus sequences. However, other interesting sequences often are in the DNA stretch containing the MAR. *Cis*-acting sites that regulate transcription are common. And a recognition site for topoisomerase II is usually present in the MAR. It is therefore possible that an MAR

serves more than one function, providing a site for attachment to the matrix, but also containing other sites at which topological changes in DNA are effected.

What is the relationship between the chromosome scaffold of dividing cells and the matrix of interphase cells? Are the same DNA sequences attached to both structures? In several cases, the same DNA fragments that are found with the nuclear matrix *in vivo* can be retrieved from the metaphase scaffold. And fragments that contain MAR sequences can bind to a metaphase scaffold. It therefore seems likely that DNA contains a single type of attachment site, which in interphase cells is connected to the nuclear matrix, and in mitotic cells is connected to the chromosome scaffold.

The nuclear matrix and chromosome scaffold consist of different proteins, although there are some common components. Topoisomerase II is a prominent component of the chromosome scaffold, and is a constituent of the nuclear matrix, suggesting that the control of topology is important in both cases.

CHROMOSOMES

5.19.7 Chromatin is divided into euchromatin and heterochromatin

Key Terms

Euchromatin comprises all of the genome in the interphase nucleus except for the heterochromatin. The euchromatin is less tightly coiled than heterochromatin, and contains the active or potentially active genes.

Heterochromatin describes regions of the genome that are highly condensed, are not transcribed, and are late-replicating. Heterochromatin is divided into two types, which are called constitutive and facultative.

The **chromocenter** is an aggregate of heterochromatin from different chromosomes.

Key Concepts

- Individual chromosomes can be seen only during mitosis.
- During interphase, the general mass of chromatin is in the form of euchromatin, which is less tightly packed than mitotic chromosomes.
- Regions of heterochromatin remain densely packed throughout interphase.

Each chromosome contains a single, very long duplex of DNA. This explains why chromosome replication is semiconservative like the individual DNA molecule. (This would not necessarily be the case if a chromosome carried many independent molecules of DNA.) The single duplex of DNA is folded into a fiber that runs continuously throughout the chromosome. *So in accounting for interphase chromatin and mitotic chromosome structure, we have to explain the packaging of a single, exceedingly long molecule of DNA into a form in which it can be transcribed and replicated, and can become cyclically more and less compressed.*

Individual eukaryotic chromosomes come into the limelight for a brief period, during the act of cell division. Only then can each be seen as a compact unit. **Figure 19.11** is an electron micrograph of a sister chromatid pair, captured at metaphase. (The sister chromatids are daughter chromosomes produced by the previous replication event, still joined together at this stage of mitosis.) Each consists of a fiber with a diameter of ~30 nm and a nubbly appearance. The DNA is 5–10× more condensed in chromosomes than in interphase chromatin.

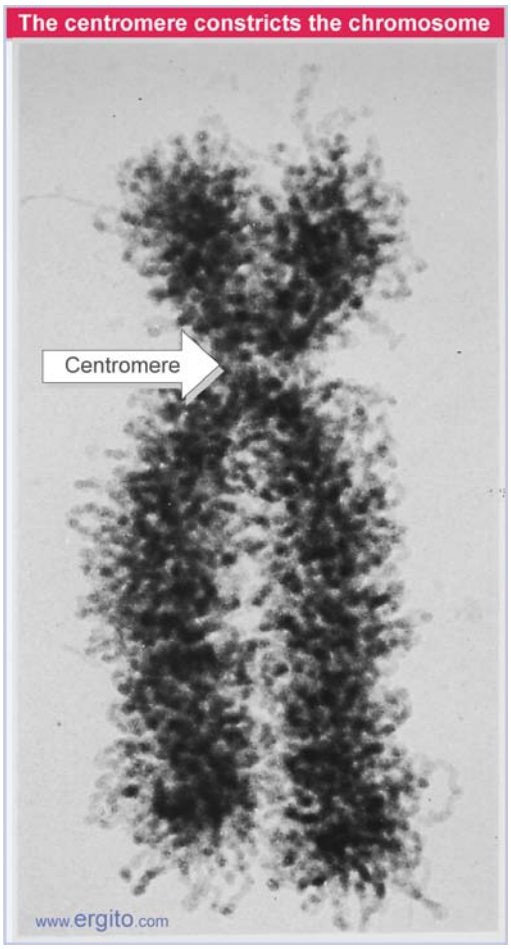

Figure 19.11 The sister chromatids of a mitotic pair each consist of a fiber (~30 nm in diameter) compactly folded into the chromosome. Photograph kindly provided by E. J. DuPraw.

During most of the life cycle of the eukaryotic cell, however, its genetic material occupies an area of the nucleus in which individual chromosomes cannot be distinguished. The structure of the interphase chromatin does not change visibly between divisions. No disruption is evident during the period of replication, when the amount of chromatin doubles. Chromatin is fibrillar, although the overall configuration of the fiber in space is hard to discern in detail. The fiber itself, however, is similar or identical to that of the mitotic chromosomes.

Chromatin can be divided into two types of material, which can be seen in the nuclear section of **Figure 19.12**:

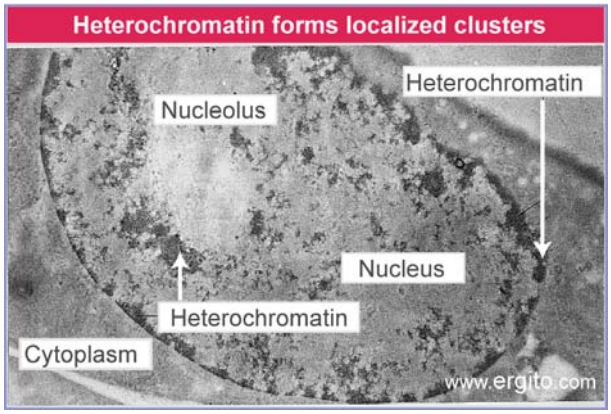

Figure 19.12 A thin section through a nucleus stained with Feulgen shows heterochromatin as compact regions clustered near the nucleolus and nuclear membrane. Photograph kindly provided by Edmund Puvion.

- In most regions, the fibers are much less densely packed than in the mitotic chromosome. This material is called **euchromatin**. It has a relatively dispersed appearance in the nucleus, and occupies most of the nuclear region in **Figure 19.12**.

- Some regions of chromatin are very densely packed with fibers, displaying a condition comparable to that of the chromosome at mitosis. This material is called **heterochromatin**. It is typically found at centromeres, but occurs at other locations also. It passes through the cell cycle with relatively little change in its degree of condensation. It forms a series of discrete clumps in **Figure 19.12**, but often the various heterochromatic regions aggregate into a densely staining **chromocenter**. (This description applies to regions that are always heterochromatic, called constitutive heterochromatin; in addition, there is another sort of heterochromatin, called facultative heterochromatin, in which regions of euchromatin are converted to a heterochromatic state).

The same fibers run continuously between euchromatin and heterochromatin, which implies that these states represent different degrees of condensation of the genetic material. In the same way, euchromatic regions exist in different states of condensation during interphase and during mitosis. So the genetic material is organized in a manner that permits alternative states to be maintained side by side in chromatin, and allows cyclical changes to occur in the packaging of euchromatin between interphase and division. We discuss the molecular basis for these states in *Molecular Biology 5.23 Controlling chromatin structure*.

The structural condition of the genetic material is correlated with its activity. The common features of constitutive heterochromatin are:

- It is permanently condensed.

- It often consists of multiple repeats of a few sequences of DNA that are not transcribed.

- The density of genes in this region is very much reduced compared with heterochromatin; and genes that are translocated into or near it are often inactivated.
- Probably resulting from the condensed state, it replicates late in S phase and has a reduced frequency of genetic recombination.

We have some molecular markers for changes in the properties of the DNA and protein components (see *Molecular Biology 5.23.15 Heterochromatin depends on interactions with histones*). They include reduced acetylation of histone proteins, increased methylation of one histone protein, and hypermethylation of cytidine bases in DNA. These molecular changes cause the condensation of the material, which is responsible for its inactivity.

Although active genes are contained within euchromatin, only a small minority of the sequences in euchromatin are transcribed at any time. So location in euchromatin is *necessary* for gene expression, but is not *sufficient* for it.

Last updated on 4-23-2002

CHROMOSOMES

5.19.8 Chromosomes have banding patterns

Key Terms

G-bands are generated on eukaryotic chromosomes by staining techniques and appear as a series of lateral striations. They are used for karyotyping (identifying chromosomal regions by the banding pattern).

Key Concepts

- Certain staining techniques cause the chromosomes to have the appearance of a series of striations called G-bands.
- The bands are lower in G·C content than the interbands.
- Genes are concentrated in the G·C-rich interbands.

Because of the diffuse state of chromatin, we cannot directly determine the specificity of its organization. But we can ask whether the structure of the (mitotic) chromosome is ordered. Do particular sequences always lie at particular sites, or is the folding of the fiber into the overall structure a more random event?

At the level of the chromosome, each member of the complement has a different and reproducible ultrastructure. When subjected to certain treatments and then stained with the chemical dye Giemsa, chromosomes generate a series of **G-bands**. **Figure 19.13** presents an example of the human set.

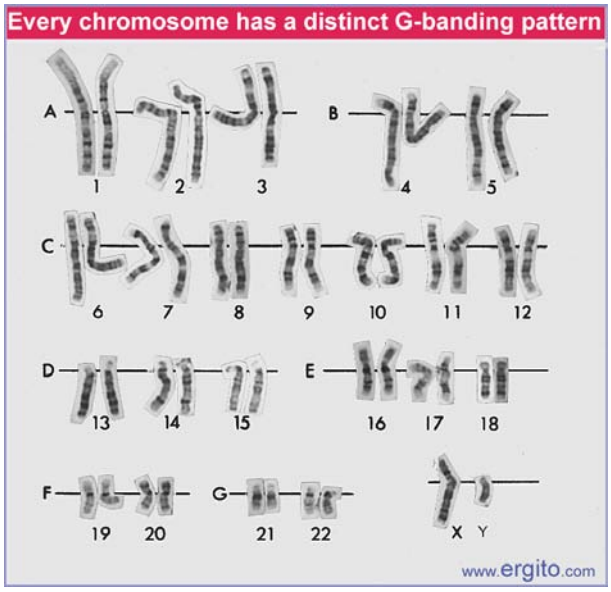

Figure 19.13 G-banding generates a characteristic lateral series of bands in each member of the chromosome set. Photograph kindly provided by Lisa Shaffer.

Until the development of this technique, chromosomes could be distinguished only by their overall size and the relative location of the centromere. G-banding allows each chromosome to be be identified by its characteristic banding pattern. This pattern allows translocations from one chromosome to another to be identified by comparison with the original diploid set. **Figure 19.14** shows a diagram of the bands of the human X chromosome. The bands are large structures, each ~10^7 bp of DNA, which could include many hundreds of genes.

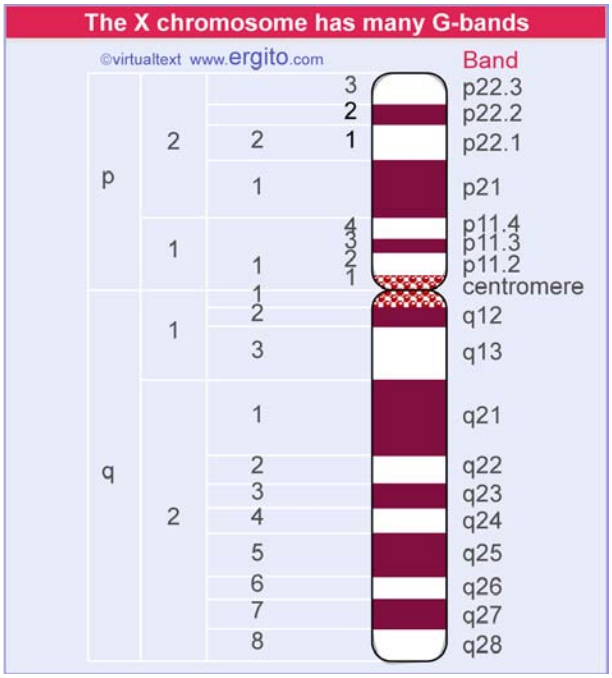

Figure 19.14 The human X chromosome can be divided into distinct regions by its banding pattern. The short arm is *p* and the long arm is *q*; each arm is divided into larger regions that are further subdivided. This map shows a low resolution structure; at higher resolution, some bands are further subdivided into smaller bands and interbands, e.g. *p*21 is divided into *p*21.1, *p*21.2, and *p*21.3.

The banding technique is of enormous practical use, but the mechanism of banding remains a mystery. All that is certain is that the dye stains untreated chromosomes more or less uniformly. So the generation of bands depends on a variety of treatments that change the response of the chromosome (presumably by extracting the component that binds the stain from the nonbanded regions). But similar bands can be generated by a variety of treatments.

The only known feature that distinguishes bands from interbands is that the bands have a lower G·C content than the interbands. This is a peculiar result. If there are ~10 bands on a large chromosome with a total content of ~100 Mb, this means that the chromosome is divided into regions of ~5 Mb in length that alternate between low G·C (band) and high G·C (interband) content (1514). There is a tendency for genes (as identified by hybridization with mRNAs) to be located in the interband regions. All of this argues for some long-range sequence-dependent organization.

The human genome sequence confirms the basic observation (1439; 1440). **Figure 19.15** shows that there are distinct fluctuations in G-C content when the genome is divided into small tranches. The average of 41% G·C is common to mammalian genomes. There are regions as low as 30% or as high as 65%. When longer tranches are examined, there is less variation. The average length of regions with >43% G·C is 200-250 kb. This makes it clear that the band/interband structure does not represent homogeneous segments that alternate in G-C content, although the bands do contain a higher content of low G·C segments. Genes are concentrated in regions of higher G-C content. We have yet to understand how the G·C content affects chromosome structure.

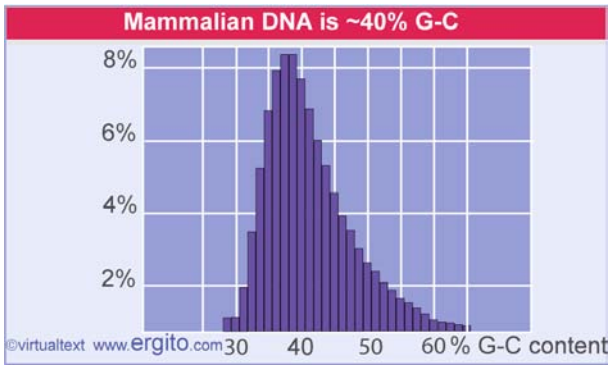

Figure 19.15 There are large fluctuations in G-C content over short distances. Each bar shows the per cent of 20 kb fragments with the given G-C content.

Last updated on 2-16-2001

References

1439. Venter, J. C. et al. (2001). *The sequence of the human genome*. Science 291, 1304-1350.

1440. International Human Genome Sequencing Consortium. (2001). *Initial sequencing and analysis of the human genome.* Nature 409, 860-921.

1514. Saccone, S., De Sario, A., Wiegant, J., Raap, A. K., Della Valle, G., and Bernardi, G. (1993). *Correlations between isochores and chromosomal bands in the human genome.* Proc. Natl. Acad. Sci. USA 90, 11929-11933.

CHROMOSOMES

5.19.9 Lampbrush chromosomes are extended

Key Terms

Chromomeres are densely staining granules visible in chromosomes under certain conditions, especially early in meiosis, when a chromosome may appear to consist of a series of chromomeres.

Lampbrush chromosomes are the extremely extended meiotic bivalents of certain amphibian oocytes.

Key Concepts

- Sites of gene expression on lampbrush chromosomes show loops that are extended from the chromosomal axis.

It would be extremely useful to visualize gene expression in its natural state, to see what structural changes are associated with transcription. The compression of DNA in chromatin, coupled with the difficulty of identifying particular genes within it, makes it impossible to visualize the transcription of individual active genes.

Gene expression can be visualized directly in certain unusual situations, in which the chromosomes are found in a highly extended form that allows individual loci (or groups of loci) to be distinguished. Lateral differentiation of structure is evident in many chromosomes when they first appear for meiosis. At this stage, the chromosomes resemble a series of beads on a string. The beads are densely staining granules, properly known as **chromomeres**. However, usually there is little gene expression at meiosis, and it is not practical to use this material to identify the activities of individual genes. But an exceptional situation that allows the material to be examined is presented by **lampbrush chromosomes**, which have been best characterized in certain amphibians.

Lampbrush chromosomes are formed during an unusually extended meiosis, which can last up to several months! During this period, the chromosomes are held in a stretched-out form in which they can be visualized in the light microscope. Later during meiosis, the chromosomes revert to their usual compact size. So the extended state essentially proffers an unfolded version of the normal condition of the chromosome.

The lampbrush chromosomes are meiotic bivalents, each consisting of two pairs of sister chromatids. **Figure 19.16** shows an example in which the sister chromatid pairs have mostly separated so that they are held together only by chiasmata. Each sister chromatid pair forms a series of ellipsoidal chromomeres, ~1-2 µm in diameter, which are connected by a very fine thread. This thread contains the two sister duplexes of DNA, and runs continuously along the chromosome, through the chromomeres.

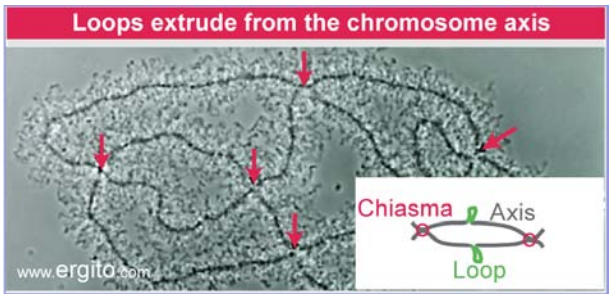

Figure 19.16 A lampbrush chromosome is a meiotic bivalent in which the two pairs of sister chromatids are held together at chiasmata (indicated by arrows). Photograph kindly provided by Joe Gall.

The lengths of the individual lampbrush chromosomes in the newt *Notophthalmus viridescens* range from 400-800 µm, compared with the range of 15-20 µm seen later in meiosis. So the lampbrush chromosomes are ~30 times less tightly packed. The total length of the entire lampbrush chromosome set is 5-6 mm, organized into ~5000 chromomeres.

The lampbrush chromosomes take their name from the lateral loops that extrude from the chromomeres at certain positions. (These resemble a lampbrush, an extinct object.) The loops extend in pairs, one from each sister chromatid. The loops are continuous with the axial thread, which suggests that they represent chromosomal material extruded from its more compact organization in the chromomere.

The loops are surrounded by a matrix of ribonucleoproteins. These contain nascent RNA chains. Often a transcription unit can be defined by the increase in the length of the RNP moving around the loop. An example is shown in **Figure 19.17**.

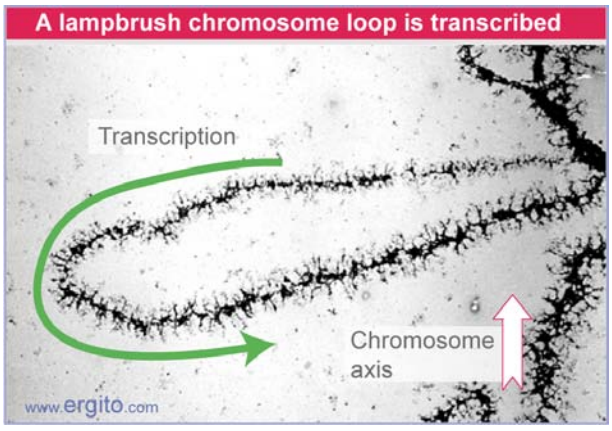

Figure 19.17 A lampbrush chromosome loop is surrounded by a matrix of ribonucleoprotein. Photograph kindly provided by Oscar Miller.

So the loop is an extruded segment of DNA that is being actively transcribed. In some cases, loops corresponding to particular genes have been identified. Then the structure of the transcribed gene, and the nature of the product, can be scrutinized *in situ*.

CHROMOSOMES

5.19.10 Polytene chromosomes form bands

Key Terms

Polytene chromosomes are generated by successive replications of a chromosome set without separation of the replicas.

Bands of polytene chromosomes are visible as dense regions that contain the majority of DNA. They include active genes.

Interbands are the relatively dispersed regions of polytene chromosomes that lie between the bands.

***In situ* hybridization (Cytological hybridization)** is performed by denaturing the DNA of cells squashed on a microscope slide so that reaction is possible with an added single-stranded RNA or DNA; the added preparation is radioactively labeled and its hybridization is followed by autoradiography.

Key Concepts

- Polytene chromosomes of Dipterans have a series of bands that can be used as a cytological map.

The interphase nuclei of some tissues of the larvae of Dipteran flies contain chromosomes that are greatly enlarged relative to their usual condition. They possess both increased diameter and greater length. **Figure 19.18** shows an example of a chromosome set from the salivary gland of *D. melanogaster*. They are called **polytene** chromosomes.

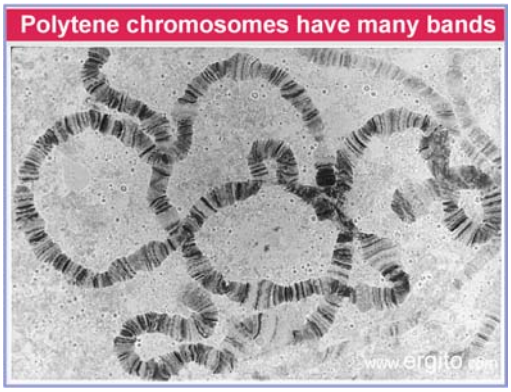

Figure 19.18 The polytene chromosomes of *D. melanogaster* form an alternating series of bands and interbands. Photograph kindly provided by Jose Bonner.

Each member of the polytene set consists of a visible series of **bands** (more properly, but rarely, described as chromomeres). The bands range in size from the largest with a breadth of ~0.5 µm to the smallest of ~0.05 µm. (The smallest can be distinguished

only under the electron microscope.) The bands contain most of the mass of DNA and stain intensely with appropriate reagents. The regions between them stain more lightly and are called **interbands**. There are ~5000 bands in the *D. melanogaster* set.

The centromeres of all four chromosomes of *D. melanogaster* aggregate to form a chromocenter that consists largely of heterochromatin (in the male it includes the entire Y chromosome). Allowing for this, ~75% of the haploid DNA set is organized into alternating bands and interbands. The length of the chromosome set is ~2000 µm. The DNA in extended form would stretch for ~40,000 µm, so the packing ratio is ~20. This demonstrates vividly the extension of the genetic material relative to the usual states of interphase chromatin or mitotic chromosomes.

What is the structure of these giant chromosomes? Each is produced by the successive replications of a synapsed diploid pair. The replicas do not separate, but remain attached to each other in their extended state. At the start of the process, each synapsed pair has a DNA content of 2C (where C represents the DNA content of the individual chromosome). Then this doubles up to 9 times, at its maximum giving a content of 1024C. The number of doublings is different in the various tissues of the *D. melanogaster* larva.

Each chromosome can be visualized as a large number of parallel fibers running longitudinally, tightly condensed in the bands, less condensed in the interbands. Probably each fiber represents a single (C) haploid chromosome. This gives rise to the name polytene. The degree of polyteny is the number of haploid chromosomes contained in the giant chromosome.

The banding pattern is characteristic for each strain of *Drosophila*. The constant number and linear arrangement of the bands was first noted in the 1930s, when it was realized that they form a *cytological map* of the chromosomes. Rearrangements – such as deletions, inversions, or duplications – result in alterations of the order of bands.

The linear array of bands can be equated with the linear array of genes. So genetic rearrangements, as seen in a linkage map, can be correlated with structural rearrangements of the cytological map. Ultimately, a particular mutation can be located in a particular band. Since the total number of genes in *D. melanogaster* exceeds the number of bands, there are probably multiple genes in most or all bands.

The positions of particular genes on the cytological map can be determined directly by the technique of ***in situ* hybridization**. The protocol is summarized in **Figure 19.19**. A radioactive probe representing a gene (most often a labeled cDNA clone derived from the mRNA) is hybridized with the denatured DNA of the polytene chromosomes *in situ*. Autoradiography identifies the position or positions of the corresponding genes by the superimposition of grains at a particular band or bands. An example is shown in **Figure 19.20**. With this type of technique at hand, it is possible to determine directly the band within which a particular sequence lies.

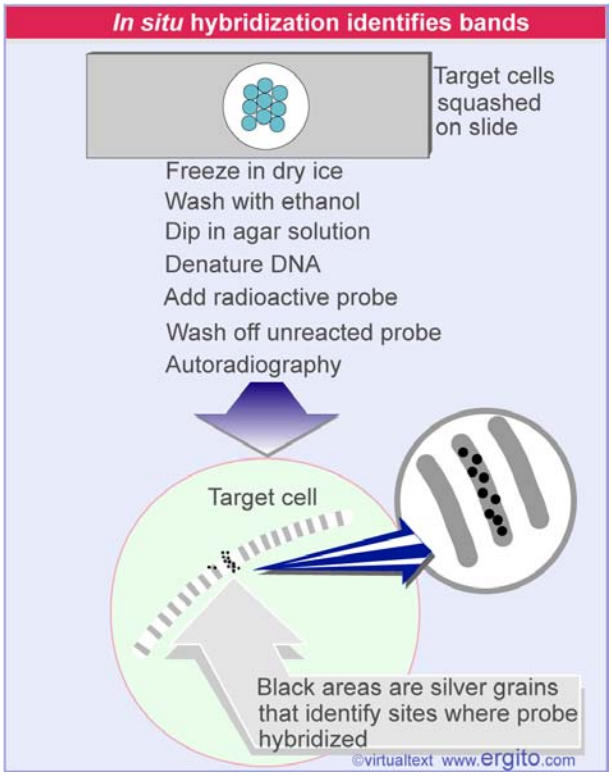

Figure 19.19 Individual bands containing particular genes can be identified by *in situ* hybridization.

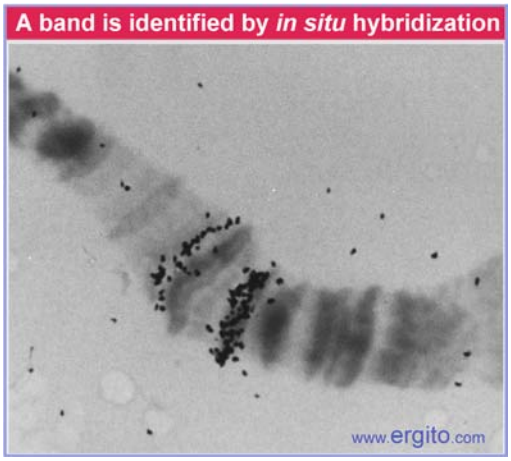

Figure 19.20 A magnified view of bands 87A and 87C shows their hybridization *in situ* with labeled RNA extracted from heat-shocked cells. Photograph kindly provided by Jose Bonner.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.5.19.10

CHROMOSOMES

5.19.11 Polytene chromosomes expand at sites of gene expression

Key Terms

A **puff** is an expansion of a band of a polytene chromosome associated with the synthesis of RNA at some locus in the band.

Key Concepts

- Bands that are sites of gene expression on polytene chromosomes expand to give "puffs".

One of the intriguing features of the polytene chromosomes is that active sites can be visualized. Some of the bands pass transiently through an expanded state in which they appear like a **puff** on the chromosome, when chromosomal material is extruded from the axis. An example of some very large puffs (called Balbiani rings) is shown in **Figure 19.21**.

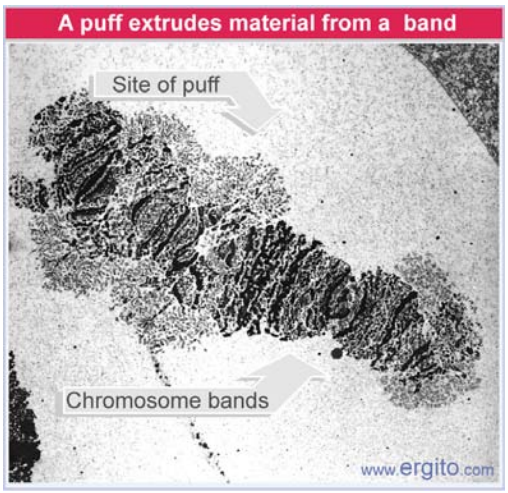

Figure 19.21 Chromosome IV of the insect *C. tentans* has three Balbiani rings in the salivary gland. Photograph kindly provided by Bertil Daneholt.

What is the nature of the puff? It consists of a region in which the chromosome fibers unwind from their usual state of packing in the band. The fibers remain continuous with those in the chromosome axis. Puffs usually emanate from single bands, although when they are very large, as typified by the Balbiani rings, the swelling may be so extensive as to obscure the underlying array of bands.

The pattern of puffs is related to gene expression. During larval development, puffs appear and regress in a definite, tissue-specific pattern. A characteristic pattern of puffs is found in each tissue at any given time. Puffs are induced by the hormone

ecdysone that controls *Drosophila* development. Some puffs are induced directly by the hormone; others are induced indirectly by the products of earlier puffs.

The puffs are *sites where RNA is being synthesized.* The accepted view of puffing has been that expansion of the band is a consequence of the need to relax its structure in order to synthesize RNA. Puffing has therefore been viewed as a consequence of transcription. A puff can be generated by a single active gene. The sites of puffing differ from ordinary bands in accumulating additional proteins, which include RNA polymerase II and other proteins associated with transcription.

The features displayed by lampbrush and polytene chromosomes suggest a general conclusion. In order to be transcribed, the genetic material is dispersed from its usual more tightly packed state. The question to keep in mind is whether this dispersion at the gross level of the chromosome mimics the events that occur at the molecular level within the mass of ordinary interphase euchromatin.

Do the bands of a polytene chromosome have a functional significance, that is, does each band correspond to some type of genetic unit? You might think that the answer would be immediately evident from the sequence of the fly genome, since by mapping interbands to the sequence it should be possible to determine whether a band has any fixed type of identity. However, so far, no pattern has been found that identifies a functional significance for the bands.

Last updated on 9-1-2001

CHROMOSOMES

5.19.12 The eukaryotic chromosome is a segregation device

Key Terms

A **microtubule organizing center (MTOC)** is a region from which microtubules emanate. The major MTOCs in a mitotic cell are the centrosomes.

The **centromere** is a constricted region of a chromosome that includes the site of attachment (the kinetochore) to the mitotic or meiotic spindle.

An **acentric fragment** of a chromosome (generated by breakage) lacks a centromere and is lost at cell division.

C-bands are generated by staining techniques that react with centromeres. The centromere appears as a darkly-staining dot.

The **kinetochore** is the structural feature of the chromosome to which microtubules of the mitotic spindle attach. Its location determines the centromeric region.

Key Concepts

- A eukaryotic chromosome is held on the mitotic spindle by the attachment of microtubules to the kinetochore that forms in its centromeric region.
- Centromeres often have heterochromatin that is rich in satellite DNA sequences.

During mitosis, the sister chromatids move to opposite poles of the cell. Their movement depends on the attachment of the chromosome to microtubules, which are connected at their other end to the poles. (The microtubules comprise a cellular filamentous system, reorganized at mitosis so that they connect the chromosomes to the poles of the cell.) The sites in the two regions where microtubule ends are organized – in the vicinity of the centrioles at the poles and at the chromosomes – are called **MTOCs** (microtubule organizing centers).

Figure 19.22 illustrates the separation of sister chromatids as mitosis proceeds from metaphase to telophase. The region of the chromosome that is responsible for its segregation at mitosis and meiosis is called the **centromere**. The centromeric region on each sister chromatid is pulled by microtubules to the opposite pole. Opposing this motive force, "glue" proteins called cohesins hold the sister chromatids together. Initially the sister chromatids separate at their centromeres, and then they are released completely from one another during anaphase when the cohesins are degraded (the cohesins are discussed in more detail in *Molecular Biology 6.29.19 Cohesins hold sister chromatids together*).

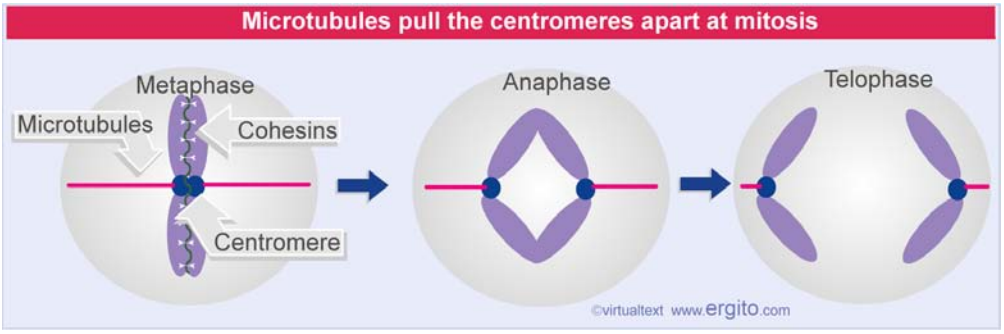

Figure 19.22 Chromosomes are pulled to the poles via microtubules that attach at the centromeres. The sister chromatids are held together until anaphase by glue proteins (cohesins). The centromere is shown here in the middle of the chromosome (metacentric), but can be located anywhere along its length, including close to the end (acrocentric) and at the end (telocentric).

The centromere is pulled toward the pole during mitosis, and the attached chromosome is dragged along behind, as it were. The chromosome therefore provides a device for attaching a large number of genes to the apparatus for division. It contains the site at which the sister chromatids are held together prior to the separation of the individual chromosomes. This shows as a constricted region connecting all four chromosome arms, as in the photograph of **Figure 19.11**, which shows the sister chromatids at the metaphase stage of mitosis.

The centromere is essential for segregation, as shown by the behavior of chromosomes that have been broken. A single break generates one piece that retains the centromere, and another, an **acentric fragment**, that lacks it. The acentric fragment does not become attached to the mitotic spindle; and as a result it fails to be included in either of the daughter nuclei.

(When chromosome movement relies on discrete centromeres, there can be *only* one centromere per chromosome. When translocations generate chromosomes with more than one centromere, aberrant structures form at mitosis, since the two centromeres on the *same* sister chromatid can be pulled toward different poles, breaking the chromosome. However, in some species the centromeres are "diffuse," which creates a different situation. Only discrete centromeres have been analyzed at the molecular level.)

The regions flanking the centromere often are rich in satellite DNA sequences and display a considerable amount of heterochromatin. Because the entire chromosome is condensed, centromeric heterochromatin is not immediately evident in mitotic chromosomes. However, it can be visualized by a technique that generates **C-bands**. In the example of **Figure 19.23**, all the centromeres show as darkly staining regions. Although it is common, heterochromatin cannot be identified around *every* known centromere, which suggests that it is unlikely to be essential for the division mechanism.

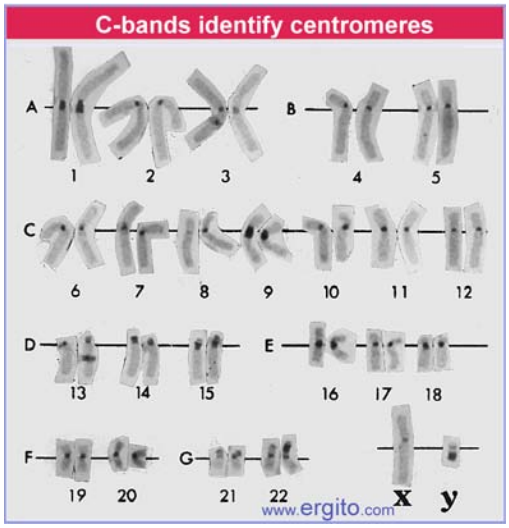

Figure 19.23 C-banding generates intense staining at the centromeres of all chromosomes. Photograph kindly provided by Lisa Shaffer.

The region of the chromosome at which the centromere forms is defined by DNA sequences (although the sequences have been defined in only a very small number of cases). The centromeric DNA binds specific proteins that are responsible for establishing the structure that attaches the chromosome to the microtubules. This structure is called the **kinetochore**. It is a darkly staining fibrous object of diameter or length ~400 nm. The kinetochore provides the MTOC on a chromosome (for review see 197). **Figure 19.24** shows the hierarchy of organization that connects centromeric DNA to the microtubules. Proteins bound to the centromeric DNA bind other proteins that bind to microtubules (see *Molecular Biology 5.19.14 The centromere binds a protein complex*).

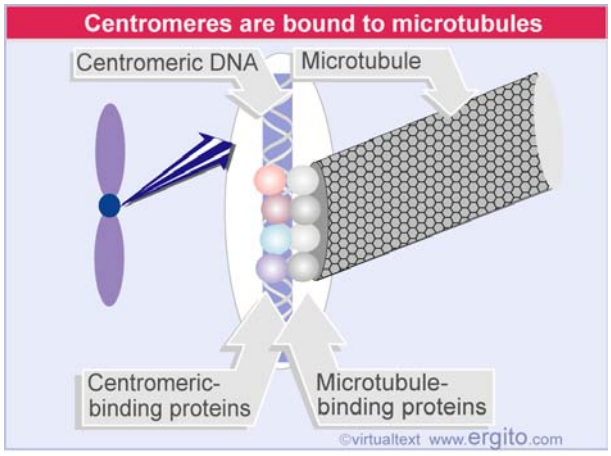

Figure 19.24 The centromere is identified by a DNA sequence that binds specific proteins. These proteins do not themselves bind to microtubules, but establish the site at which the microtubule-binding proteins in turn bind.

Reviews

197. Hyman, A. A. and Sorger, P. K. (1995). *Structure and function of kinetochores in budding yeast*. Annu. Rev. Cell Dev. Biol. 11, 471-495.

CHROMOSOMES

5.19.13 Centromeres have short DNA sequences in *S. cerevisiae*

Key Concepts

- CEN elements are identified in *S. cerevisiae* by the ability to allow a plasmid to segregate accurately at mitosis.
- CEN elements consists of short conserved sequences CDE-I and CDE-III that flank the A·T-rich region CDE-II.

If a centromeric sequence of DNA is responsible for segregation, any molecule of DNA possessing this sequence should move properly at cell division, while any DNA lacking it will fail to segregate. This prediction has been used to isolate centromeric DNA in the yeast, *S. cerevisiae*. Yeast chromosomes do not display visible kinetochores comparable to those of higher eukaryotes, but otherwise divide at mitosis and segregate at meiosis by the same mechanisms.

Genetic engineering has produced plasmids of yeast that are replicated like chromosomal sequences (see *Molecular Biology 4.13.6 Replication origins can be isolated in yeast*). However, they are unstable at mitosis and meiosis, disappearing from a majority of the cells because they segregate erratically. Fragments of chromosomal DNA containing centromeres have been isolated by their ability to confer mitotic stability on these plasmids.

A *CEN* fragment is identified as the minimal sequence that can confer stability upon such a plasmid. Another way to characterize the function of such sequences is to modify them *in vitro* and then reintroduce them into the yeast cell, where they replace the corresponding centromere on the chromosome. This allows the sequences required for *CEN* function to be defined directly in the context of the chromosome (for review see 188).

A *CEN* fragment derived from one chromosome can replace the centromere of another chromosome with no apparent consequence. This result suggests that centromeres are interchangeable. *They are used simply to attach the chromosome to the spindle, and play no role in distinguishing one chromosome from another.*

The sequences required for centromeric function fall within a stretch of ~120 bp. The centromeric region is packaged into a nuclease-resistant structure, and it binds a single microtubule. We may therefore look to the *S. cerevisiae* centromeric region to identify proteins that bind centromeric DNA and proteins that connect the chromosome to the spindle.

Three types of sequence element may be distinguished in the *CEN* region, as summarized in **Figure 19.25** (3174):

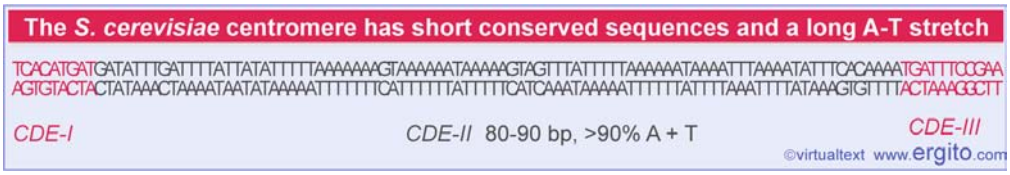

Figure 19.25 Three conserved regions can be identified by the sequence homologies between yeast *CEN* elements.

- CDE-I is a sequence of 9 bp that is conserved with minor variations at the left boundary of all centromeres.
- CDE-II is a >90% A·T-rich sequence of 80-90 bp found in all centromeres; its function could depend on its length rather than exact sequence. Its constitution is reminiscent of some short tandemly repeated (satellite) DNAs (see *Molecular Biology 1.4.12 Arthropod satellites have very short identical repeats*). Its base composition may cause some characteristic distortions of the DNA double helical structure.
- CDE-III is an 11 bp sequence highly conserved at the right boundary of all centromeres. Sequences on either side of the element are less well conserved, and may also be needed for centromeric function. (CDE-III could be longer than 11 bp if it turns out that the flanking sequences are essential.)

Mutations in CDE-I or CDE-II reduce but do not inactivate centromere function, but point mutations in the central CCG of CDE-III completely inactivate the centromere (for review see 189).

Reviews

188. Blackburn, E. H. and Szostak, J. W. (1984). *The molecular structure of centromeres and telomeres*. Annu. Rev. Biochem. 53, 163-194.

189. Clarke, L. and Carbon, J. (1985). *The structure and function of yeast centromeres*. Annu. Rev. Genet. 19, 29-56.

References

3174. Fitzgerald-Hayes, M., Clarke, L., and Carbon, J. (1982). *Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs.* Cell 29, 235-244.

CHROMOSOMES

5.19.14 The centromere binds a protein complex

Key Concepts

- A specialized protein complex that is an alternative to the usual chromatin structure is formed at CDE-II.
- The CBF3 protein complex that binds to CDE-III is essential for centromeric function.
- The proteins that connect these two complexes may provide the connection to microtubules.

Can we identify proteins that are necessary for the function of CEN sequences? There are several genes in which mutations affect chromosome segregation, and whose proteins are localized at centromeres. The contributions of these proteins to the centromeric structure are summarized in **Figure 19.26**.

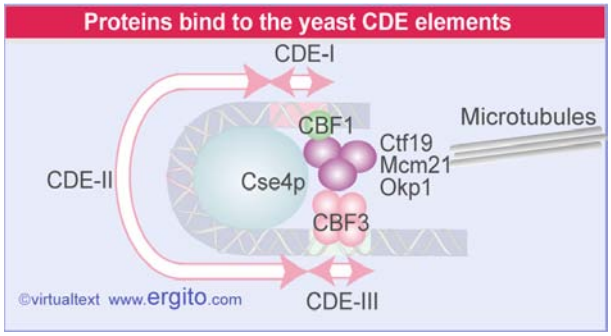

Figure 19.26 The DNA at CDE-II is wound around a protein aggregate including Cse4p, CDE-III is bound to CBF3, and CDE-I is bound to CBF1. These proteins are connected by the group of Ctf19, Mcm21, and Okp1.

A specialized chromatin structure is built by binding the CDE-II region to a protein called Cse4p, which resembles one of the histone proteins that comprise the basic subunits of chromatin (see *Molecular Biology 5.23.15 Heterochromatin depends on interactions with histones*) (624). A protein called Mif2p may also be part of this complex or connected to it (3169). Cse4p and Mif2p have counterparts that are localized at higher eukaryotic centromeres, called CENP-A and CENP-C, which suggests that this interaction may be a universal aspect of centromere construction. The basic interaction consists of bending the DNA of the CDE-II region around a protein aggregate; the reaction is probably assisted by the occurrence of intrinsic bending in the CDE-II sequence.

CDE-I is bound by the homodimer CBF1; this interaction is not essential for centromere function, but in its absence the fidelity of chromosome segregation is reduced ~10×. A 240 kD complex of four proteins, called CBF3, binds to CDE-III.

This interaction is essential for centromeric function (3175).

The proteins bound at CDE-I and CDE-III are connected to each other and also to the protein structure bound at CDE-II by another group of proteins (Ctf19, Mcm21, Okp1) (3170). The connection to the microtubule may be made by this complex.

The overall model suggests that the complex is localized at the centromere by a protein structure that resembles the normal building block of chromatin (the nucleosome). The bending of DNA at this structure allows proteins bound to the flanking elements to become part of a single complex. Some components of the complex (possibly not those that bind directly to DNA) link the centromere to the microtubule. The construction of kinetochores probably follows a similar pattern, and uses related components, in a wide variety of organisms (for review see 3425).

Last updated on 12-10-2002

Reviews

3425. Kitagawa, K. and Hieter, P. (2001). *Evolutionary conservation between budding yeast and human kinetochores.* Nat. Rev. Mol. Cell Biol. 2, 678-687.

References

624. Meluh, P. B. et al. (1998). *Cse4p is a component of the core centromere of S. cerevisiae.* Cell 94, 607-613.

3169. Meluh, P. B. and Koshland, D. (1997). *Budding yeast centromere composition and assembly as revealed by in vitro cross-linking.* Genes Dev. 11, 3401-3412.

3170. Ortiz, J., Stemmann, O., Rank, S., and Lechner,J. (1999). *A putative protein complex consisting of Ctf19, Mcm21, and Okp1 represents a missing link in the budding yeast kinetochore.* Genes Dev. 13, 1140-1155.

3175. Lechner, J. and Carbon, J. (1991). *A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere.* Cell 64, 717-725.

CHROMOSOMES

5.19.15 Centromeres may contain repetitious DNA

Key Concepts

- Centromeres in higher eukaryotic chromosomes contain large amounts of repetitious DNA.
- The function of the repetitious DNA is not known.

The length of DNA required for centromeric function is often quite long. (The short, discrete, elements of *S. cerevisiae* may be an exception to the general rule.) In those cases where we can equate specific DNA sequences with the centromeric region, they usually include repetitive sequences.

S. cerevisiae is the only case so far in which centromeric DNA can be identified by its ability to confer stability on plasmids. However, a related approach has been used with the yeast *S. pombe*. This has only 3 chromosomes, and the region containing each centromere has been identified by deleting most of the sequences of each chromosome to create a stable minichromosome. This approach locates the centromeres within regions of 40-100 kb that consist largely or entirely of repetitious DNA. It is not clear how much of each of these rather long regions is required for chromosome segregation at mitosis and meiosis.

Attempts to localize centromeric functions in *Drosophila* chromosomes suggest that they are dispersed in a large region, consisting of 200-600 kb. The large size of this type of centromere suggests that it is likely to contain several separate specialized functions, including sequences required for kinetochore assembly, sister chromatid pairing, etc. (3177).

The size of the centromere in *Arabidopsis* is comparable (922). Each of the 5 chromosomes has a centromeric region in which recombination is very largely suppressed. This region occupies >500 kb. Clearly it includes the centromere, but we have no direct information as to how much of it is required. There are expressed genes within these regions, which casts some doubt on whether the entire region is part of the centromere. At the center of the region is a series of 180 bp repeats; this is the type of structure generally associated with centromeres. It is too early to say how these structures relate to centromeric function.

The primary motif comprising the heterochromatin of primate centromeres is the α satellite DNA, which consists of tandem arrays of a 170 bp repeating unit. There is significant variation between individual repeats, although those at any centromere tend to be better related to one another than to members of the family in other locations. It is clear that the sequences required for centromeric function reside within the blocks of α satellite DNA, but it is not clear whether the α satellite sequences themselves provide this function, or whether other sequences are embedded within the α satellite arrays (3178; 3179).

Last updated on 1-10-2000

Reviews

3179. Wiens, G. R. and Sorger, P. K. (1998). *Centromeric chromatin and epigenetic effects in kinetochore assembly.* Cell 93, 313-316.

References

922. Copenhaver, G. P. et al. (1999). *Genetic definition and sequence analysis of Arabidopsis centromeres.* Science 286, 2468-2474.

3177. Sun, X., Wahlstrom, J., and Karpen, G. (1997). *Molecular structure of a functional Drosophila centromere.* Cell 91, 1007-1019.

3178. Haaf, T., Warburton, P. E., and Willard, H. F. (1992). *Integration of human alpha-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation.* Cell 70, 681-696.

CHROMOSOMES

5.19.16 Telomeres have simple repeating sequences

Key Terms

A **telomere** is the natural end of a chromosome; the DNA sequence consists of a simple repeating unit with a protruding single-stranded end that may fold into a hairpin.

Key Concepts

- The telomere is required for the stability of the chromosome end.
- A telomere consists of a simple repeat where a C+A-rich strand has the sequence $C_{>1}(A/T)_{1-4}$.

Another essential feature in all chromosomes is the **telomere**, which "seals" the end. We know that the telomere must be a special structure, because chromosome ends generated by breakage are "sticky" and tend to react with other chromosomes, whereas natural ends are stable.

We can apply two criteria in identifying a telomeric sequence:

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

The problem of finding a system that offers an assay for function again has been brought to the molecular level by using yeast. All the plasmids that survive in yeast (by virtue of possessing *ARS* and *CEN* elements) are circular DNA molecules. Linear plasmids are unstable (because they are degraded). Could an authentic telomeric DNA sequence confer stability on a linear plasmid? Fragments from yeast DNA that prove to be located at chromosome ends can be identified by such an assay. And a region from the end of a known natural linear DNA molecule – the extrachromosomal rDNA of *Tetrahymena* – is able to render a yeast plasmid stable in linear form.

Telomeric sequences have been characterized from a wide range of lower and higher eukaryotes. The same type of sequence is found in plants and man, so the construction of the telomere seems to follow a universal principle. Each telomere consists of a long series of short, tandemly repeated sequences. There may be 100-1000 repeats, depending on the organism.

All telomeric sequences can be written in the general form $C_n(A/T)_m$, where $n>1$ and m is 1-4. **Figure 19.27** shows a generic example. One unusual property of the telomeric sequence is the extension of the G-T-rich strand, usually for 14-16 bases as

a single strand. The G-tail is probably generated because there is a specific limited degradation of the C-A-rich strand (2000).

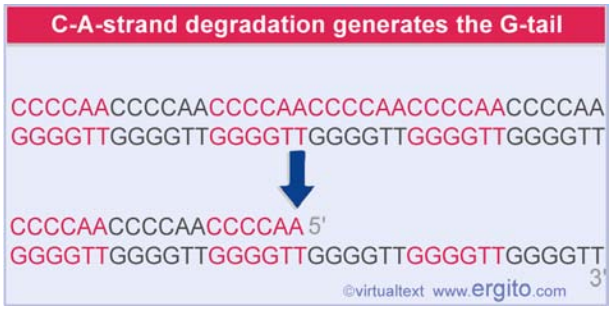

Figure 19.27 A typical telomere has a simple repeating structure with a G-T-rich strand that extends beyond the C-A-rich strand. The G-tail is generated by a limited degradation of the C-A-rich strand.

Some indications about how a telomere functions are given by some unusual properties of the ends of linear DNA molecules. In a trypanosome population, the ends are variable in length. When an individual cell clone is followed, the telomere grows longer by 7-10 bp (1-2 repeats) per generation. Even more revealing is the fate of ciliate telomeres introduced into yeast. After replication in yeast, *yeast telomeric repeats are added onto the ends of the* Tetrahymena *repeats.*

Addition of telomeric repeats to the end of the chromosome in every replication cycle could solve the difficulty of replicating linear DNA molecules discussed in *Molecular Biology 4.13.8 The ends of linear DNA are a problem for replication* . The addition of repeats by *de novo* synthesis would counteract the loss of repeats resulting from failure to replicate up to the end of the chromosome. Extension and shortening would be in dynamic equilibrium (for review see 188; 193).

If telomeres are continually being lengthened (and shortened), their exact sequence may be irrelevant. All that is required is for the end to be recognized as a suitable substrate for addition. This explains how the ciliate telomere functions in yeast.

Last updated on 9-12-2001

Reviews

188. Blackburn, E. H. and Szostak, J. W. (1984). *The molecular structure of centromeres and telomeres*. Annu. Rev. Biochem. 53, 163-194.

193. Zakian, V. A. (1989). *Structure and function of telomeres*. Annu. Rev. Genet. 23, 579-604.

References

2000. Wellinger, R. J., Ethier, K., Labrecque, P., and Zakian, V. A. (1996). *Evidence for a new step in telomere maintenance.* Cell 85, 423-433.

CHROMOSOMES

5.19.17 Telomeres seal the chromosome ends

Key Concepts

- The protein TRF2 catalyzes a reaction in which the 3′ repeating unit of the G+T-rich strand forms a loop by displacing its homologue in an upstream region of the telomere.

Isolated telomeric fragments do not behave as though they contain single-stranded DNA; instead they show aberrant electrophoretic mobility and other properties.

Guanine bases have an unusual capacity to associate with one another. The single-stranded G-rich tail of the telomere can form "quartets" of G residues (594; 595). Each quartet contains 4 guanines that hydrogen bond with one another to form a planar structure. Each guanine comes from the corresponding position in a successive TTAGGG repeating unit. **Figure 19.28** shows an organization based on a recent crystal structure (2861). The quartet that is illustrated represents an association between the first guanine in each repeating unit. It is stacked on top of another quartet that has the same organization, but is formed from the second guanine in each repeating unit. A series of quartets could be stacked like this in a helical manner. While the formation of this structure attests to the unusual properties of the G-rich sequence *in vitro*, it does not of course demonstrate whether the quartet forms *in vivo*.

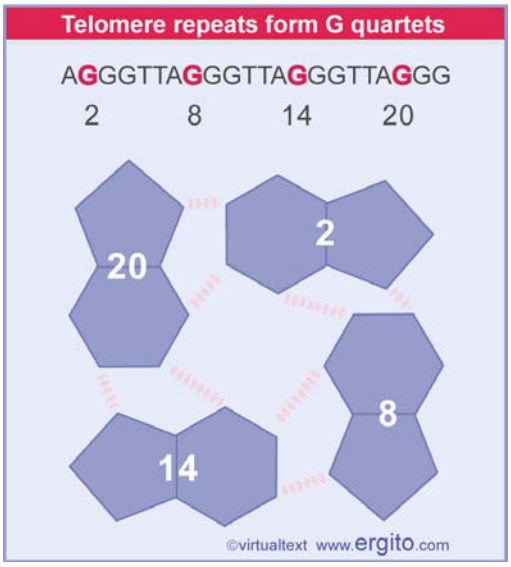

Figure 19.28 The crystal structure of a short repeating sequence from the human telomere forms 3 stacked G quartets. The top quartet contains the first G from each repeating unit. This is stacked above quartets that contains the second G (G3, G9, G15, G21) and the third G (G4, G10, G16, G22).

What feature of the telomere is responsible for the stability of the chromosome end? **Figure 19.29** shows that a loop of DNA forms at the telomere. The absence of any free end may be the crucial feature that stabilizes the end of the chromosome. The average length of the loop in animal cells is 5-10 kb.

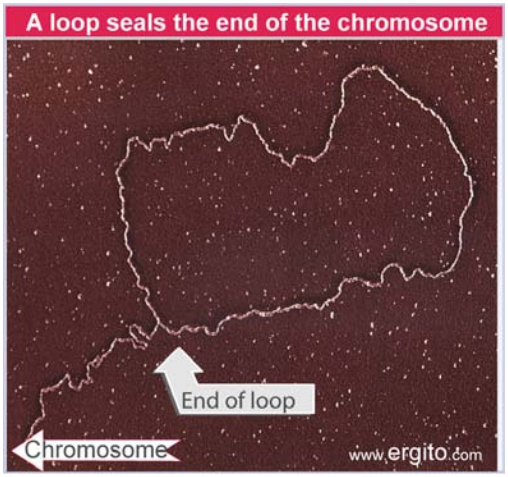

Figure 19.29 A loop forms at the end of chromosomal DNA. Photograph kindly provided by Jack Griffith.

Figure 19.30 shows that the loop is formed when the 3′ single-stranded end of the telomere $(TTAGGG)_n$ displaces the same sequence in an upstream region of the telomere. This converts the duplex region into a structure like a D-loop, where a series of TTAGGG repeats are displaced to form a single-stranded region, and the tail of the telomere is paired with the homologous strand (597).

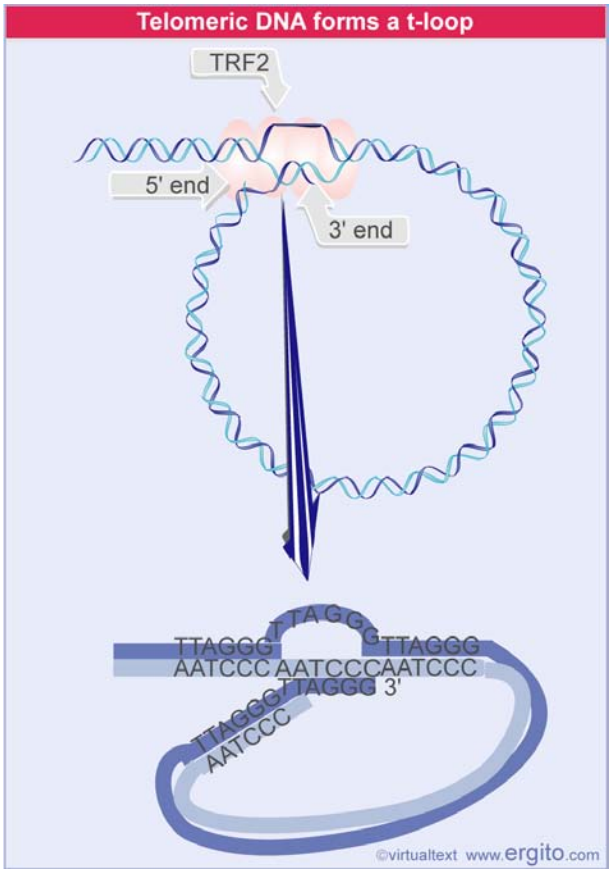

Figure 19.30 The 3 ′ single-stranded end of the telomere (TTAGGG)n displaces the homologous repeats from duplex DNA to form a t-loop. The reaction is catalyzed by TRF2.

The reaction is catalyzed by the telomere-binding protein TRF2, which together with other proteins forms a complex that stabilizes the chromosome ends. Its importance in protecting the ends is indicated by the fact the deletion of TRF2 causes chromosome rearrangements to occur (1997; 1998).

Last updated on 8-22-2002

References

594. Henderson, E., Hardin, C. H., Walk, S. K., Tinoco, I., and Blackburn, E. H. (1987). *Telomeric oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs*. Cell 51, 899-908.

595. Williamson, J. R., Raghuraman, K. R., and Cech, T. R. (1989). *Monovalent cation-induced structure of telomeric DNA: the G-quartet model*. Cell 59, 871-880.

597. Griffith, J. D. et al. (1999). *Mammalian telomeres end in a large duplex loop*. Cell 97, 503-514.

1997. van Steensel, B., Smogorzewska, A., and de Lange, T. (1998). *TRF2 protects human telomeres from end-to-end fusions.* Cell 92, 401-413.

1998. Karlseder, J., Broccoli, D., Dai, Y., Hardy, S., and de Lange, T. (1999). *p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2.* Science 283, 1321-1325.

2861. Parkinson, G. N., Lee, M. P., and Neidle, S. (2002). *Crystal structure of parallel quadruplexes from human telomeric DNA.* Nature 417, 876-880.

CHROMOSOMES

5.19.18 Telomeres are synthesized by a ribonucleoprotein enzyme

Key Terms

Telomerase is the ribonucleoprotein enzyme that creates repeating units of one strand at the telomere, by adding individual bases to the DNA 3′ end, as directed by an RNA sequence in the RNA component of the enzyme.

Key Concepts

- Telomerase uses the 3′–OH of the G+T telomeric strand to prime synthesis of tandem TTGGGG repeats.
- The RNA component of telomerase has a sequence that pairs with the C+A-rich repeats.
- One of the protein subunits is a reverse transcriptase that uses the RNA as template to synthesis the G+T-rich sequence.

The telomere has two functions:

- One is to protect the chromosome end. Any other DNA end – for example, the end generated by a double strand break – becomes a target for repair systems. The cell has to be able to distinguish the telomere.
- The second is to allow the telomere to be extended. Otherwise it would become shorter with each replication cycle (because replication cannot start at the very end).

Proteins that bind to the telomere provide the solution for both problems. In yeast, different sets of proteins solve each problem, but both are bound to the telomere via the same protein, Cdc13 (1999):

- The Stn1 protein protects against degradation (specifically against any extension of the degradation of the C-A-strand that generates the G-tail).
- A **telomerase** enzyme extends the C-A-rich strand. Its activity is influenced by two proteins that have ancillary roles, such as controlling the length of the extension.

The telomerase uses the 3′–OH of the G+T telomeric strand as a primer for synthesis of tandem TTGGGG repeats. Only dGTP and dTTP are needed for the activity. The telomerase is a large ribonucleoprotein that consists of a templating RNA (coded by *TLC1*) and a protein with catalytic activity (*EST2*). The short RNA component (159 bases long in *Tetrahymena*, 192 bases long in *Euplotes)* includes a

sequence of 15-22 bases that is identical to two repeats of the C-rich repeating sequence. This RNA provides the template for synthesizing the G-rich repeating sequence. The protein component of the telomerase is a catalytic subunit that can act only upon the RNA template provided by the nucleic acid component (for review see 195; 196; 593; 596; 986).

Figure 19.31 shows the action of telomerase. The enzyme progresses discontinuously: the template RNA is positioned on the DNA primer, several nucleotides are added to the primer, and then the enzyme translocates to begin again. The telomerase is a specialized example of a reverse transcriptase, an enzyme that synthesizes a DNA sequence using an RNA template (see *Molecular Biology 4.17.4 Viral DNA is generated by reverse transcription*). We do not know how the complementary (C-A-rich) strand of the telomere is assembled, but we may speculate that it could be synthesized by using the 3′–OH of a terminal G-T hairpin as a primer for DNA synthesis.

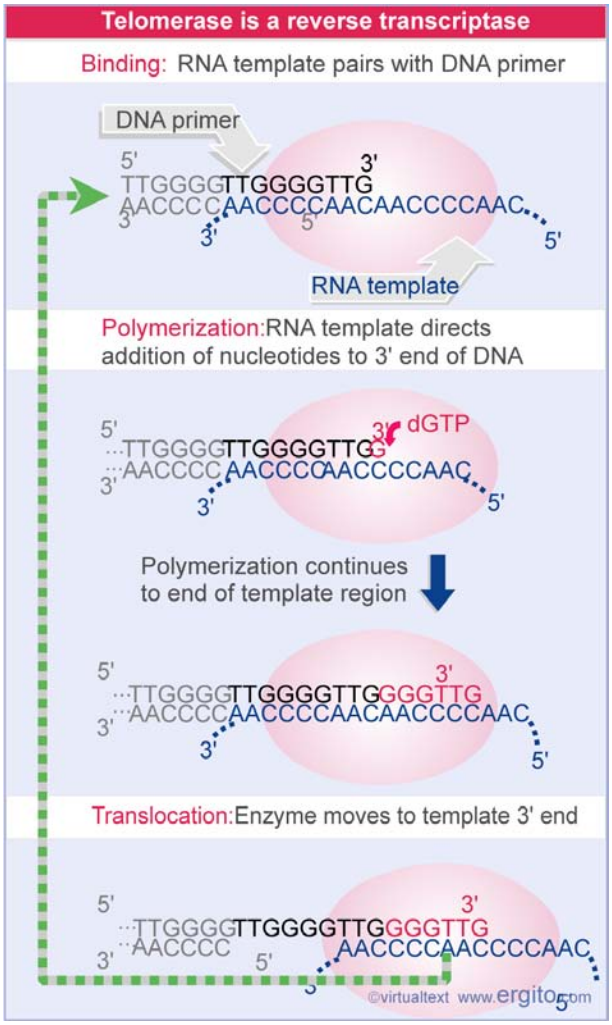

Figure 19.31 Telomerase positions itself by base pairing between the RNA template and the protruding single-stranded DNA primer. It adds G and T bases one at a time to the primer, as directed by the template. The cycle starts again when one repeating unit has been added.

Telomerase synthesizes the individual repeats that are added to the chromosome ends, but does not itself control the number of repeats. Other proteins are involved in determining the length of the telomere. They can be identified by the *EST1* and *EST3* mutants in yeast that have altered telomere lengths. These proteins may bind telomerase, and influence the length of the telomere by controlling the access of telomerase to its substrate. Proteins that bind telomeres in mammalian cells have been found similarly, but less is known about their functions (for review see 198; 199).

The minimum features required for existence as a chromosome are:

- Telomeres to ensure survival.
- A centromere to support segregation.
- An origin to initiate replication.

All of these elements have been put together to construct a yeast artificial chromosome (YAC). This is a useful method for perpetuating foreign sequences. It turns out that the synthetic chromosome is stable only if it is longer than 20-50 kb. We do not know the basis for this effect, but the ability to construct a synthetic chromosome allows us to investigate the nature of the segregation device in a controlled environment (592).

Last updated on 9-12-2001

Reviews

195. Blackburn, E. H. (1991). *Structure and function of telomeres*. Nature 350, 569-573.

196. Blackburn, E. H. (1992). *Telomerases*. Annu. Rev. Biochem. 61, 113-129.

198. Zakian, V. A. (1995). *Telomeres: beginning to understand the end*. Science 270, 1601-1607.

199. Zakian, V. A. (1996). *Structure, function, and replication of S. cerevisiae telomeres.* Annu. Rev. Genet. 30, 141-172.

986. Collins, K. (1999). *Ciliate telomerase biochemistry.* Annu. Rev. Biochem. 68, 187-218.

References

592. Murray, A., and Szostak, J. W. (1983). *Construction of artificial chromosomes in yeast*. Nature 305, 189-193.

593. Greider, C. and Blackburn, E. H. (1987). *The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity*. Cell 51, 887-898.

596. Shippen-Lentz, D. and Blackburn, E. H. (1990). *Functional evidence for an RNA template in telomerase*. Science 247, 546-552.

1999. Pennock, E., Buckley, K., and Lundblad, V. (2001). *Cdc13 delivers separate complexes to the telomere for end protection and replication.* Cell 104, 387-396.

CHROMOSOMES

5.19.19 Telomeres are essential for survival

Telomerase activity is found in all dividing cells and is generally turned off in terminally differentiated cells that do not divide. **Figure 19.32** shows that if telomerase is mutated in a dividing cell, the telomeres become gradually shorter with each cell division. An example of the effects of such a mutation in yeast are shown in **Figure 19.33**, where the telomere length shortens over ~120 generations from 400 bp to zero (1384).

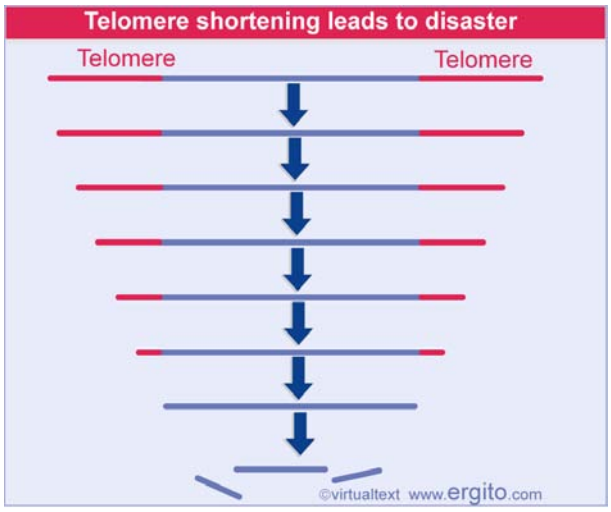

Figure 19.32 Mutation in telomerase causes telomeres to shorten in each cell division. Eventual loss of the telomere causes chromosome breaks and rearrangements.

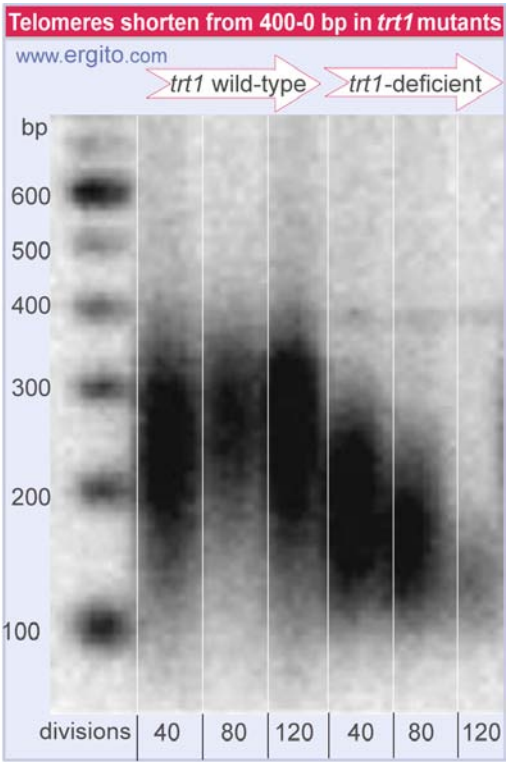

Figure 19.33 Telomere length is maintained at ~350 bp in wild-type yeast, but a mutant in the *trt1* gene coding for the RNA component of telomerase rapidly shortens its telomeres to zero length. Data kindly provided by Tom Cech and Toru Nakamura (from 1384).

Loss of telomeres has very bad effects. When the telomere length reaches zero, it becomes difficult for the cells to divide successfully. Attempts to divide typically generate chromosome breaks and translocations. This causes an increased rate of mutation (1952). In yeast this is associated with a loss of viability and the culture becomes predominantly occupied by senescent cells (elongated and nondividing, which eventually die; see 1384).

But some cells grow out of the senescing culture. They have acquired the ability to extend their telomeres by an alternative to telomerase activity. The survivors fall into groups (1929). One group have circularized their chromosomes – since they now have no telomeres, they have become independent of telomerase. The other group use unequal crossing-over to extend their telomeres (see **Figure 19.34**). Because the telomere is a repeating structure, it is possible for two telomeres to misalign when chromosomes pair. Recombination between the mispaired regions generates an unequal crossing-over, as shown previously in **Figure 4.1**, when the length of one recombinant chromosome increases, and the length of the other recombinant chromosome decreases.

Cells usually suppress unequal crossing-over because of its potentially deleterious consequences. Two systems are responsible for suppressing crossing-over between telomeres. One is provided by telomere-binding proteins. In yeast, the frequency of recombination between telomeres is increased by deletion of the gene *taz1*, which

codes for a protein that regulates telomerase activity (1929). The second is a general system that undertakes mismatch repair (see *Molecular Biology 4.15 Recombination and repair*). In addition to correcting mismatched base pairs that may arise in DNA, this system suppresses recombination between mispaired regions. As shown in **Figure 19.34**, this includes telomeres. When it is mutated, a greater proportion of telomerase-deficient yeast survive the loss of telomeres because recombination between telomeres generates some chromosomes with longer telomeres (1930).

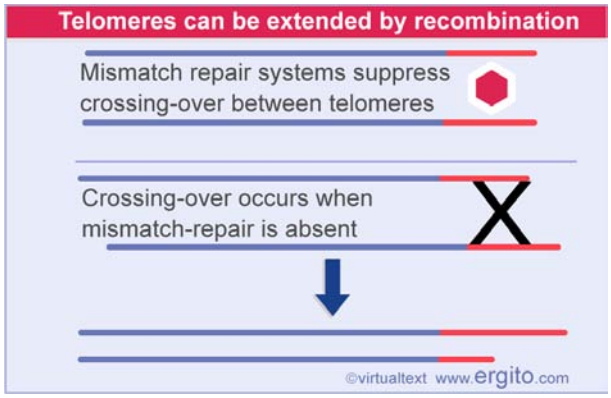

Figure 19.34 Crossing-over in telomeric regions is usually suppressed by mismatch-repair systems, but can occur when they are mutated. An unequal crossing-over event extends the telomere of one of the products, allowing the chromosome to survive in the absence of telomerase.

When eukaryotic cells are placed in culture, they usually divide for a fixed number of generations and then enter senescence. The reason appears to be a decline in telomere length because of the absence of telomerase expression (see *Molecular Biology 6.30.24 Telomere shortening causes cell senescence*). Cells enter a crisis from which some emerge, but typically with chromosome rearrangements that have resulted from lack of protection of chromosome ends. These rearrangements may cause mutations that contribute to the tumorigenic state. The absence of telomerase expression in this situation is due to failure to express the gene, and reactivation of telomerase is one of the mechanisms by which these cells then survive continued culture (this of course was not an option in the yeast experiments in which the gene had been deleted).

Last updated on 8-16-2001

References

1384. Nakamura, T. M., Morin, G. B., Chapman, K. B., Weinrich, S. L., Andrews, W. H., Lingner, J., Harley, C. B., and Cech, T. R. (1997). *Telomerase catalytic subunit homologs from fission yeast and human.* Science 277, 955-959.

1929. Nakamura, T. M., Cooper, J. P., and Cech, T. R. (1998). *Two modes of survival of fission yeast without telomerase.* Science 282, 493-496.

1930. Rizki, A. and Lundblad, V. (2001). *Defects in mismatch repair promote telomerase-independent proliferation.* Nature 411, 713-716.

1952. Hackett, J. A., Feldser, D. M., and Greider, C. W. (2001). *Telomere dysfunction increases mutation rate and genomic instability.* Cell 106, 275-286.

CHROMOSOMES

5.19.20 Summary

The genetic material of all organisms and viruses takes the form of tightly packaged nucleoprotein. Some virus genomes are inserted into preformed virions, while others assemble a protein coat around the nucleic acid. The bacterial genome forms a dense nucleoid, with ~20% protein by mass, but details of the interaction of the proteins with DNA are not known. The DNA is organized into ~100 domains that maintain independent supercoiling, with a density of unrestrained supercoils corresponding to ~1 / 100-200 bp. Interphase chromatin and metaphase chromosomes both appear to be organized into large loops. Each loop may be an independently supercoiled domain. The bases of the loops are connected to a metaphase scaffold or to the nuclear matrix by specific DNA sites.

Transcriptionally active sequences reside within the euchromatin that comprises the majority of interphase chromatin. The regions of heterochromatin are packaged ~5-10× more compactly, and are transcriptionally inert. All chromatin becomes densely packaged during cell division, when the individual chromosomes can be distinguished. The existence of a reproducible ultrastructure in chromosomes is indicated by the production of G-bands by treatment with Giemsa stain. The bands are very large regions, ~10^7 bp, that can be used to map chromosomal translocations or other large changes in structure.

Lampbrush chromosomes of amphibians and polytene chromosomes of insects have unusually extended structures, with packing ratios <100. Polytene chromosomes of *D. melanogaster* are divided into ~5000 bands, varying in size by an order of magnitude, with an average of ~25 kb. Transcriptionally active regions can be visualized in even more unfolded ("puffed") structures, in which material is extruded from the axis of the chromosome. This may resemble the changes that occur on a smaller scale when a sequence in euchromatin is transcribed.

The centromeric region contains the kinetochore, which is responsible for attaching a chromosome to the mitotic spindle. The centromere often is surrounded by heterochromatin. Centromeric sequences have been identified only in yeast *S. cerevisiae*, where they consist of short conserved elements, CDE-I and CDE-III that binds CBF1 and the CBF3 complex, respectively, and a long A·T-rich region called CDE-II that binds Cse4p to form a specialized structure in chromatin. Another group of proteins that binds to this assembly provides the connection to microtubules.

Telomeres make the ends of chromosomes stable. Almost all known telomeres consist of multiple repeats in which one strand has the general sequence $C_n(A/T)_m$, where $n>1$ and m = 1-4. The other strand, $G_n(T/A)_m$, has a single protruding end that provides a template for addition of individual bases in defined order. The enzyme telomere transferase is a ribonucleoprotein, whose RNA component provides the template for synthesizing the G-rich strand. This overcomes the problem of the inability to replicate at the very end of a duplex. The telomere stabilizes the chromosome end because the overhanging single strand $G_n(T/A)_m$ displaces its homologue in earlier repeating units in the telomere to form a loop, so there are no free ends.