

Chromosome Structure

N Patrick Higgins, *University of Alabama, Birmingham, Alabama, USA*

Genes are organized into discrete cellular structures called chromosomes that coordinate DNA replication and distribution of replicated genetic copies between two daughter cells. As vehicles of genetic transmission, chromosomes play a central role in Darwinian evolution.

Abundance

Biology is divided into three great kingdoms: Eubacteria, Archaea and Eukaryota. Bacteria (Eubacteria and Archaea) are ubiquitous in the environment, and these small single-celled organisms grow over an amazingly wide range of environmental conditions. For example, bacteria can grow at temperatures below freezing, and certain members of the archaea can grow at depths of over three miles and at temperatures exceeding 600°F and 200 atmospheres of pressure. Eukaryotes, which include the more conspicuous fungi and all plants and animals, are usually recognized as the predominant life forms on earth. They are subjects of great biological interest and they differ from bacteria by having a larger cell size and by compartmentalizing chromosomal deoxyribonucleic acid (DNA) in a nucleus, which separates it from the protein-making cytoplasm. None the less, considering the large cell numbers and wide environmental growth ranges, bacteria easily account for the majority of DNA biomass on earth.

The chromosome is the heart of a central paradox in evolution. How do species in the three kingdoms remain the same over long periods of geological time and also generate sufficient variability to produce new species, sometimes relatively rapidly? Stability versus change is a crucial dichotomy in molecular biology. The events that bring about stability and change in DNA structure involve processes of replication, transcription and recombination. Similar mechanisms operate in the three living kingdoms, but the key molecular mechanisms that control and catalyse these events are understood best in a eubacterium: *Escherichia coli*.

Chromosome Size

Free-living bacteria need genetic information to synthesize proteins for executing vital functions. Most bacteria have a single chromosome with DNA that is about 2 Mbp (mega base pairs) long (1 Mbp = 1 000 000 base pairs), but the DNA content of different species varies from 0.58 to greater than 9 Mbp of DNA, and some bacteria have multiple chromosomes. For example, *Leptospira* has two chromosomes of 4.4 and 4.6 Mbp and the largest bacterial

Introductory article

Article Contents

- Abundance
- Chromosome Size
- Plasmids
- Chromosome Shape
- Enzymes of DNA Topology
- Chromatin
- DNA Replication
- Eukaryotic Segregation
- Transcription
- Recombination
- *Chi* Sites
- Transposons
- Site-specific Recombination

genome yet analysed is that of *Myxococcus xanthus*, with 9.2 Mbp (9 200 000 bp). However, the best studied organism in nature is *E. coli*, which has a 4.6-Mbp chromosome with 4288 genes for proteins, seven operons for ribosomal ribonucleic acids (RNAs), and 86 genes for transfer RNAs.

The *E. coli* chromosome contains numerous gene families, each family having evolved from a common ancestor. The largest gene family is comprised of 96 three-component (ABC) transporters, which are membrane-bound machines that import and export a variety of small molecules and proteins. By contrast, the smallest free-living organism is *Mycoplasma genitalium* with a 0.58-Mbp genome that encodes 468 protein genes, one ribosomal RNA operon and one ABC transporter. Both *E. coli* and *M. genitalium* have complete information for the synthesis of cell walls, cell membranes and critical enzymes of intermediary metabolism, plus the RNA molecules, ribosomal proteins and a clutch of enzymes (the replisome) to replicate DNA efficiently. Putative functions for about a quarter of the genes of *E. coli* remain to be discovered. A reasonable estimate is that 150–200 protein-encoding genes would be 'essential' for a basic bacterial lifestyle (in a rich medium).

Eukaryotic organisms generally have larger chromosomes than bacteria. For example, the yeast *Saccharomyces cerevisiae* has about 6000 genes (50% more than *E. coli*), whereas mammalian cells contain 1000 times (per haploid equivalent) the DNA of an *E. coli* cell. In humans the 5000 Mbp of haploid DNA is distributed among 22 autosomes and two sex-specific chromosomes. Eukaryotic DNA is localized in a compartment, the nucleus, which is separated by a phospholipid-containing membrane from cytoplasmic ribosomes and protein translation activity. During cell division, the eukaryotic nuclear membrane

breaks down once per cell cycle to distribute the 46 diploid chromosomes equally between two daughter cells.

Plasmids

In addition to the large chromosome, many (most?) bacteria have additional DNA molecules called plasmids (or episomes.) Plasmids are separate DNA molecules that contain a replication origin which allows them to multiply independently of the host chromosome. Plasmids range in size from 1 kbp (Kilo base pair) (1000 bp) to 100 kbp, and these DNA molecules encode genetic systems for specialized functions. Some plasmids make extracellular appendages that allow bacteria to infect and colonize sensitive eukaryotic hosts. Plasmids often carry genes that confer on bacteria the ability to survive in the presence of antibiotics such as tetracycline, kanamycin and penicillin. Many plasmids also contain genes that promote DNA transfer so that plasmid genes can move into other bacterial species. Plasmid transfer has caused the emergence of bacterial pathogens that are resistant to most of the useful antibiotics in medicine, with notable examples including multidrug-resistant strains of *Staphylococcus* and *Mycobacterium tuberculosis*.

In most eukaryotes, plasmids are rare. However, *S. cerevisiae* contains a plasmid called the 2- μ circle which efficiently partitions to new daughter cells at every cell division. This DNA serves as a convenient module for gene cloning and performing genetic experiments in yeast.

Chromosome Shape

On a macroscopic scale, bacterial chromosomes are either circular or linear. Circular chromosomes are most common, at least among the best-studied bacteria. However, the causative agent of Lyme disease, *Borrelia burgdorferi*, has a 2-Mb linear chromosome plus 12 different linear plasmids. Eukaryotic chromosomes are invariably linear, and they have two ends, each carrying a special structure called a telomere, and a organized region called the centromere which allows the chromosome to attach to cellular machinery that moves it to the proper place during cell division.

One critical facet of chromosome structure is that DNA is a plectonemic helix, which means that two helical strands entwine about each other. For duplex DNA the two antiparallel strands, often referred to as the Watson and Crick (red and blue in **Figure 1**), are interwound once for every 10 bp. Because of this wound configuration, biochemical transactions that involve strand separation require chromosome movement (spin) about DNA's long axis. The processes of DNA replication, recombination and transcription all require DNA rotation, and during

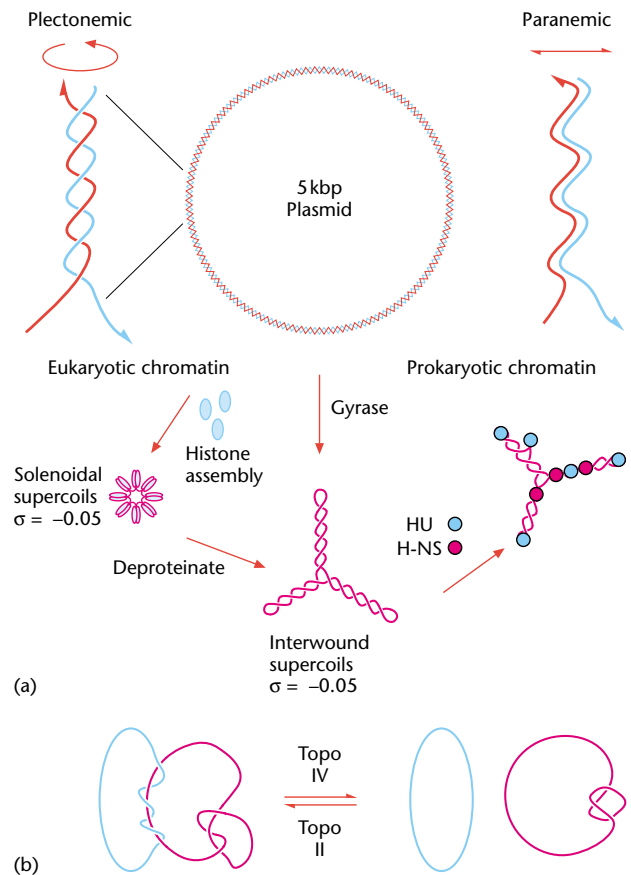


Figure 1 (a) DNA is a plectonemic helix (left) rather than a paranemic helix (right), and thus it must spin axially to undergo replication, transcription and extended pairing for homologous recombination. Supercoiling in circular bacterial chromosomes is maintained by the concerted action of DNA gyrase, which introduces negative supercoils at the expense of adenosine triphosphate (ATP) binding and hydrolysis, and TopoI plus TopoII, which remove excess negative supercoils. Negative supercoiled DNA adopts an interwound conformation. However, 'solenoidal' supercoils can be stabilized when DNA is wrapped on a protein surface. This is the mechanism of supercoil formation in mammalian cells (centre left). Most bacteria have nonspecific DNA-binding proteins such as HU and H-NS, which stabilize supercoils but are much less effective at condensing DNA compared with true histones (centre right). (b) Bacteria and eukaryotes both have an enzyme, TopoIV and TopoII respectively, designed to unknot and untangle DNA.

DNA synthesis the rotation speed approaches 6000 rpm. Chromosomal DNA molecules are very long and thin, so DNA must fold many times to fit within the confines of a bacterial cell. How does DNA twisting transpire without chromosomes getting tangled up? The problem was important enough that Watson and Crick considered a paranemic helix, which is a pair of coils that lay side by side without interwinding. Strands of a paranemic helix can separate without rotation (**Figure 1**). None the less DNA is plectonemic, and keeping chromosomes untangled requires a special class of enzymes called topoisomerases.

Enzymes of DNA Topology

The enzymes that solve the untangling problem are topoisomerases. Topoisomerases break and rejoin DNA molecules, thereby allowing individual strands to pass one through another. The number of phosphodiester bonds broken and reformed per reaction cycle divides topoisomerases into two classes. Type I enzymes, which include TopoI and TopoIII (the odd-numbered topoisomerases), break one strand per cycle, and type II enzymes (even numbered), gyrase, eukaryotic TopoII, and *E. coli* TopoIV, break two strands simultaneously. These enzymes have two important roles: they provide a swivel to allow processes such as replication and transcription to proceed unimpeded, and they untangle knots and inter-chromosome links between DNA molecules.

In all organisms DNA becomes organized into supercoils, which are turns of the double strand over the interwound twists of the Watson–Crick helix. Supercoils can be either positive or negative; negative supercoils are opposite to the handedness of the Watson–Crick turns, and positive supercoils have the same handedness. Supercoil density is defined by a term σ , which represents the number of superhelical turns divided by the Watson–Crick turns of a double helix. Supercoiling influences the Watson–Crick structure and, like the spring, the free energy of supernegative supercoils increases exponentially with quantity. Enzymes that unwind DNA to carry out their function (RNA polymerase and DNA replisomes) may sense and utilize supercoiling energy. However, once negative supercoiling has reached a critical density, the Watson–Crick structure will unwind locally, forcing DNA into alternative structures to relieve tension. Well-studied supercoil-dependent alternatives to the Watson–Crick form of DNA include left-handed Z-DNA, cruciforms, and triple-stranded or H-form DNA.

In eubacteria, gyrase, TopoI and TopoIV maintain *in vivo* supercoiling levels, $\sigma = -0.05$ to -0.075 . Gyrase is unique for its ability to introduce negative supercoils into relaxed, positively or negatively supercoiled DNA at the expense of adenosine triphosphate (ATP) binding and hydrolysis. An essential enzyme in bacteria, gyrase is critical for DNA reactions that include recombination, replication, transcription and chromosome segregation. *In vitro*, DNA gyrase can supercoil DNA to a value of $\sigma = -0.1$, a level at which many sequences adopt an alternative structure. *In vivo*, σ is buffered by the counteractive relaxing activity of TopoI and TopoIV. The second critical function of topoisomerases is untangling and unknotting DNA (Figure 1b). Eukaryotic TopoII and *E. coli* TopoIV efficiently unknot and untangle chromosomes, and after replication this activity allows the chromosomes to segregate into daughter cells.

Although an average eukaryotic nucleus is larger than an *E. coli* cell, nuclear DNA is even more concentrated than bacterial DNA. Rather than being supercoiled by an

enzyme such as gyrase, eukaryotic DNA is wrapped tightly around nucleosomes, generating solenoidal supercoils that condense DNA 8-fold (Figure 1). A nucleosome contains four subunits: histones H2A, H2B, H3 and H4. If histones are stripped from DNA by protein denaturants, eukaryotic DNA adopts the interwound structure of bacterial chromosomes, with about the same negative superhelix density (Figure 1).

Eukaryotes have three topoisomerases: TopoI and TopoIII are type 1 enzymes that break only one strand, and TopoII is the only type 2 activity that breaks both strands simultaneously. Eukaryotic TopoI removes both positive and negative supercoils from DNA, whereas TopoII carries out unknotting and decatenating reactions (see below). TopoI and TopoII are essential for eukaryotic cell viability; the functions of TopoIII in both eukaryotes and bacteria remain undefined.

Chromatin

Chromatin is a term that refers not just to DNA but to the proteins attached to a chromosome. In the dimensions of B-form DNA, *E. coli* is a sphere that is 6 kbp long (8 nm) and 4 kbp wide (2 nm), so the 4.6-Mbp chromosome must be folded many times to fit within a cell. Negative supercoiling forces DNA into an interwound configuration (Figure 1). Interwound supercoiling is produced at the expense of ATP by the enzyme gyrase, and supercoiling produces two important consequences. First, the DNA molecule doubles back on itself so that length is halved relative to that of the linear form. Second, supercoiled branches are dynamic so that opposing DNA sites in the interwound network are constantly changing. A protein bound to DNA in one supercoiling domain interacts more than 100 times more frequently with other proteins in the same domain than it does with proteins bound to a different domain. When DNA is liberated from cells by breaking the peptidoglycan coat, chromosomes form bundled loops that represent domains. Such preparations (called nucleoids) behave as discrete bodies.

Many reactions of the chromosome require the formation of intricate DNA–protein machines to replicate, transcribe or recombine DNA at specific sequences. A group of ‘chromosome-associated’ proteins assists in the formation of these complexes by shaping DNA. These proteins are sometimes described as histone-like, although they share no structural similarity with the eukaryotic histones (Figure 1). Chromosome-associated proteins include HU, H-NS, integration host factor (IHF) and factor for inversion stimulation (FIS).

HU is encoded by two genes, *hupA* and *hupB*, which are closely related to each other and to the genes encoding the two IHF subunits. HU is the most abundant double-stranded DNA-binding protein in *E. coli*, although it also

binds RNA. Although HU shows a preference for structurally ‘bendable’ sequences, it binds DNA relatively nonspecifically, forming complexes that stabilize negative supercoils in double-stranded DNA (Figure 1). In phage Mu transposition reactions, HU binds to a flexed DNA structure called a transpososome with high affinity and specificity.

H-NS is a second relatively nonspecific DNA-binding protein that forms dimers, tetramers and possibly higher-order associations when bound to DNA. Like HU, H-NS constrains negative supercoils. H-NS binds ‘bendable’ DNA, which include AT-rich sequences often found near bacterial promoters; H-NS also modulates the transcription of numerous operons in *E. coli* and of lysogenic phages.

FIS protein forms homodimers and binds DNA at specific sites. FIS was discovered nearly simultaneously in two related site-specific recombination systems: the Hin inversion reaction of *Salmonella typhimurium* and the Gin inversion reaction of phage Mu. FIS concentration rises in early log phase when the protein has regulatory roles in replication initiation at *oriC* and transcription of ribosomal RNA operons, and then FIS abundance declines in stationary phase.

IHF protein is composed of two subunits (encoded by the *himA* and *hip* genes) that are about 100 amino acids long and which make stable heterodimers. IHF protein binds to a consensus sequence that was first discovered in phage λ , where it stimulates insertional and excisional recombination *in vitro* by a factor of 10^4 . IHF bends DNA severely (more than 90°); such bends are critical for site-specific recombination in λ , for repression and activation of phage Mu, and for repression–activation in a number of cellular operons, including *ilv*, *oxyR* and *nifA*.

In the eukaryotic nucleus, enzymes such as RNA polymerase gain access to DNA, which remains histone bound throughout most biochemical transactions. The way in which multiple proteins interact with nucleosome-coated DNA is not completely understood; however, protein access to DNA is modulated by histone structure. Histone phosphorylation and acetylation change chromatin structure. Genes that are transcribed contain nucleosomes with acetylated subunits of histones H3 and H4, whereas inactive genes are bound by histones lacking acetyl groups. In addition to the histones, eukaryotic chromosomes contain regulatory proteins that are much less abundant. One class of proteins, the high mobility group (HMG) family of DNA-binding proteins, bends DNA much like bacterial proteins IHF and FIS.

DNA Replication

The central problem in chromosome replication is generating two high-fidelity DNA copies and distributing

them precisely to compartments that become the daughter cells. To control chromosomal replication temporally, and to integrate DNA metabolism with other aspects of the cell growth such as membrane synthesis and cell wall expansion, *E. coli* makes a master regulator, the DnaA protein, which controls the onset of replication at a site called *oriC*.

The bacterial replication cycle can be described by progression through four stages. Less detail is available for eukaryotes, but many points are similar. As cells grow, the mass of protein and membrane increases to a critical point that triggers the initiation of replication. Two replisomes are associated in a ‘factory’ that moves to the cell centre during chromosomal elongation. As DNA strands are pulled to the centre, replicated sister chromosomes migrate toward opposite cell poles. Completion of DNA synthesis occurs as the DNA that is pulled into the factory reaches the terminus. At this point, chromosomes are tangled together (catenated), and all physical connections must be removed to allow final separation. When physical separation is complete, cell wall synthesis forms two daughter cells.

Initiation

DNA replication starts at *oriC* near position -85 of the standard *E. coli* genetic map (Figure 2). Initiation involves the DnaA protein-assisted assembly of two replisomes, each dedicated to replicate half a chromosome. Replisome components include: the DnaB helicase, which is an ATP-dependent ring that moves along a single strand breaking open the double helix for polymerase access; a dimeric DNA polymerase molecule with accessory cofactors; and a complex called the primosome that triggers RNA-primed initiation of new DNA strands on one side of the growing fork. Each replisome has a pair of polymerase subunits: one synthesizes DNA continuously along the strand with 5′–3′ polarity, and the other subunit synthesizes DNA discontinuously in small fragments called Okazaki pieces. Thus DNA replication is semidiscontinuous.

In eukaryotic replication, initiation steps are not as fully understood as they are in *E. coli*. Initiation occurs at *ars* sites, so called because they are autonomously replicating sequences (Figure 3). Initiation is controlled by a group of proteins called ORC (origin replication complex). Typically, a chromosome has many ORC-binding sites, and bidirectional semidiscontinuous forks move out from several *ars* sites on each chromosome until they converge with other forks. Eukaryotic replication also occurs in ‘factories’, and most of the chromosome is replicated in a semiconservative and semidiscontinuous mode, as in *E. coli*. The eukaryotic replication fork behaves very similarly to that found in *E. coli*.

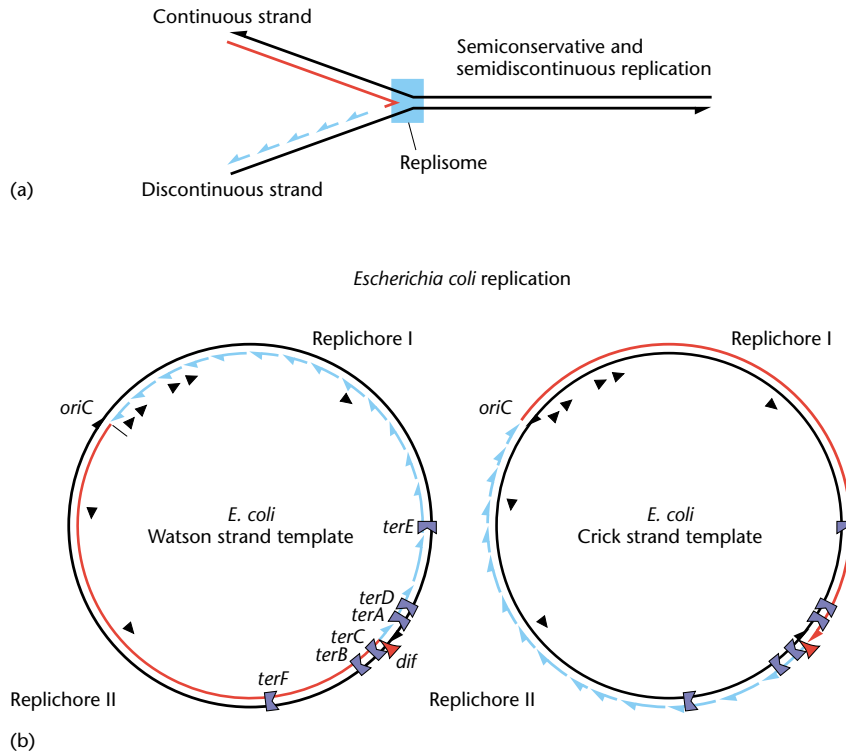


Figure 2 (a) A replication fork is shown with a replisome as a blue box, moving and synthesizing DNA from left to right. Replication is semiconservative, meaning that a new strand is synthesized on each of the two parental template strands, and semidiscontinuous, meaning that the strand that follows the fork with 5' to 3' polarity is made as one continuous piece while the strand that follows with overall 3' to 5' polarity is made discontinuously as Okazaki pieces. (b) The two circular chromosomes show the products of bidirectional replication. Initiation starts at *oriC* (– 85) and proceeds clockwise in the zone called replichore I and counterclockwise through the zone called replichore II. Replication forks meet at a terminus near the *dif* site (– 34). Strands made in the continuous mode are shown in red, and discontinuous strands in blue. Also included in the map are the seven ribosomal RNA operons (arrowheads), five of which reside in replichore I and two of which are found in replichore II, six *ter* sites and the *dif* site. Note that the two daughters are synthesized with different semidiscontinuous styles in replichores II and I so that the Watson and Crick strands are isomers.

Elongation

After initiation, duplication of the entire 4.6-Mbp *E. coli* chromosome takes 40 min. One replisome copies DNA in the clockwise direction (replichore I), while the other replisome copies DNA counterclockwise (replichore II) (Figure 2). DNA polymerase synthesizes new strands at a staggering rate of 300 nucleotides per second. Replication is semiconservative, meaning that one newly synthesized strand is made to match each parental template. Because DNA polymerases synthesize DNA only in the 5' to 3' direction, and because replication forks copy both templates at nearly the same time, there is asymmetry in the mechanism of chain growth. One strand is made as a single piece while the other strand is synthesized in a series of 1-kb segments, with each segment being initiated with a small RNA molecule. These are called Okazaki pieces, and they are stitched together after removal of the RNA primers. This semidiscontinuous pattern of DNA replication is a recurring theme in organisms as diverse as plants and humans.

Because replication is semidiscontinuous, the sister chromosomes (shown at the bottom of Figure 2) are replication isomers. The template strand with 5' to 3' polarity in the clockwise direction (the Watson strand template in Figure 2) always contains DNA replichore I, which is made discontinuously, and DNA replichore II, which is made continuously; the Crick strand template always has the opposite style. Mutation and recombination rates differ between discontinuous and continuous replication modes, so the sisters are not always equivalent.

As replication proceeds, positive supercoils build up in front of the fork, and the daughter chromosomes become entangled behind the fork (Figure 4). These two topological problems are solved in similar ways in the two kingdoms. For *E. coli*, DNA gyrase removes positive supercoils ahead of the fork, and TopoIV removes the links between the daughter chromosomes, which are called catenanes. Eukaryotic TopoI eliminates the positive supercoils formed ahead of the fork, and TopoII decatenates the chromosomes for segregation (Figure 4).

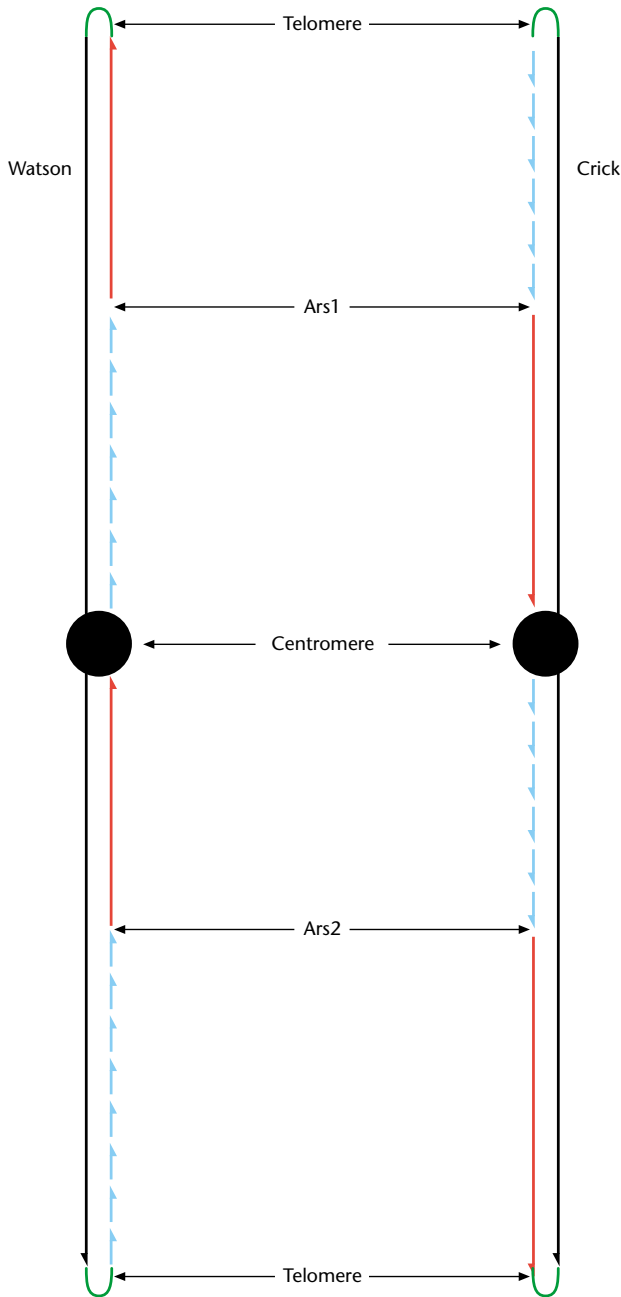


Figure 3 Structure and replication pattern of a eukaryotic chromosome. Eukaryotic chromosomes are linear structures which have special structures at each end called telomeres (green) and an organizer centre called the centromere which attaches the chromosome to the spindle during chromosome segregation. Replication is initiated at *ars* sites, and replication is carried out semidiscontinuously so that the two strands are replication isomers.

One segment of a eukaryotic chromosome that is different from prokaryotic chromosomes is the tip of the chromosome, the telomere (**Figure 3**). Telomeres are

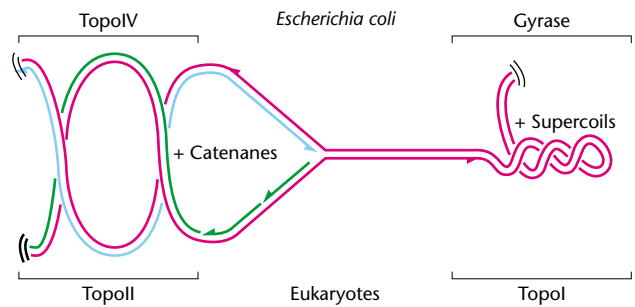


Figure 4 Topology of DNA replication. Movement of a replication fork produces positive supercoiling ahead of the fork and results in entanglements of the sister chromosomes, called catenanes, behind the fork. Positive supercoils are removed by gyrase in bacteria and by TopoI in eukaryotes, whereas TopoIV resolves catenanes in bacteria and TopoII in eukaryotes.

replicated by a special DNA polymerase called telomerase, which is related to the reverse transcriptase of retroviruses. Telomerase synthesizes a simple repeat sequence that is added on to every chromosome using an RNA template that is part of the enzyme. In most organisms telomerase is not expressed after cell differentiation, and consequently the telomere sequences shorten with age, eventually causing cell senescence and death.

Termination

In *E. coli*, replication is completed in a region of the chromosome called the terminus, which is 180° around the circular genetic map from *oriC*. Two special sites are found near the terminus. First, the terminus is surrounded by at least six *ter* sites (**Figure 2**, blue boxes) which bind the Tus protein. A *ter*–Tus complex induces replication forks to stop or pause in a unidirectional fashion by impeding the movement of DnaB helicase. *Ter* sites work in only one orientation, and they are ordered so that replication forks can proceed to the terminus but not continue in the direction opposite to normal fork movement. Thus, *ter* sites ensure that replication produces no more than one round of synthesis per initiation event. The second special site at the terminus is called *dif*, which functions to monomerize dimeric chromosomes during segregation (see below in site-specific recombination). The processes involved in terminating eukaryotic replication forks are not yet known.

Segregation

At the conclusion of replication of a circular chromosome, all physical barriers must be eliminated before daughter chromosomes completely separate. Replication leaves sister chromosomes interlocked by multiple catenane links (**Figure 1**). To uncuff the DNA molecules, both gyrase and DNA topoisomerase work together. Once chromosomes

separate, cell wall synthesis can form a permanent barrier between daughter cells.

Bacteria can constrict the cell cycle to allow doubling times that are shorter than the 40-min period necessary to make one complete chromosomal copy. Under conditions of rapid growth, initiation at *oriC* speeds up so that daughter cells inherit one fully replicated chromosome plus two or more partially replicated chromosomes at the time of cell division (dichotomous replication).

Eukaryotic Segregation

Following DNA replication in eukaryotes, which occurs in a part of the growth cycle called the S phase, cells move through a mechanical cycle called mitosis to distribute the replicated chromosomes to each daughter cell (**Figure 5**). Mitosis proceeds through four stages. The first is the prophase in which, after replication, each chromosome becomes condensed. Stage 2 is metaphase, where two

changes occur: each pair of replicated chromosomes moves to the cell centre and then the nuclear membrane begins to dissolve. In stage 3, anaphase, one centromere of each pair of chromosomes is attached to a set of fibres called the spindle, and molecular motors pull one of each chromosome pair to opposite cell poles. In stage 4, telophase, the nuclear membrane is reformed and daughter cells are separated by the synthesis of a new septum.

Transcription

A fundamental problem in biology is understanding how organisms respond to the wide variety of environmental conditions that a cell or a population of cells encounters. As noted above, different bacterial species grow over an amazing range of temperatures and environmental niches, but even a single organism can adapt to widely varying growth states, from aerobic growth on rich nutrients to anaerobic growth in minimal salts and a single carbon and

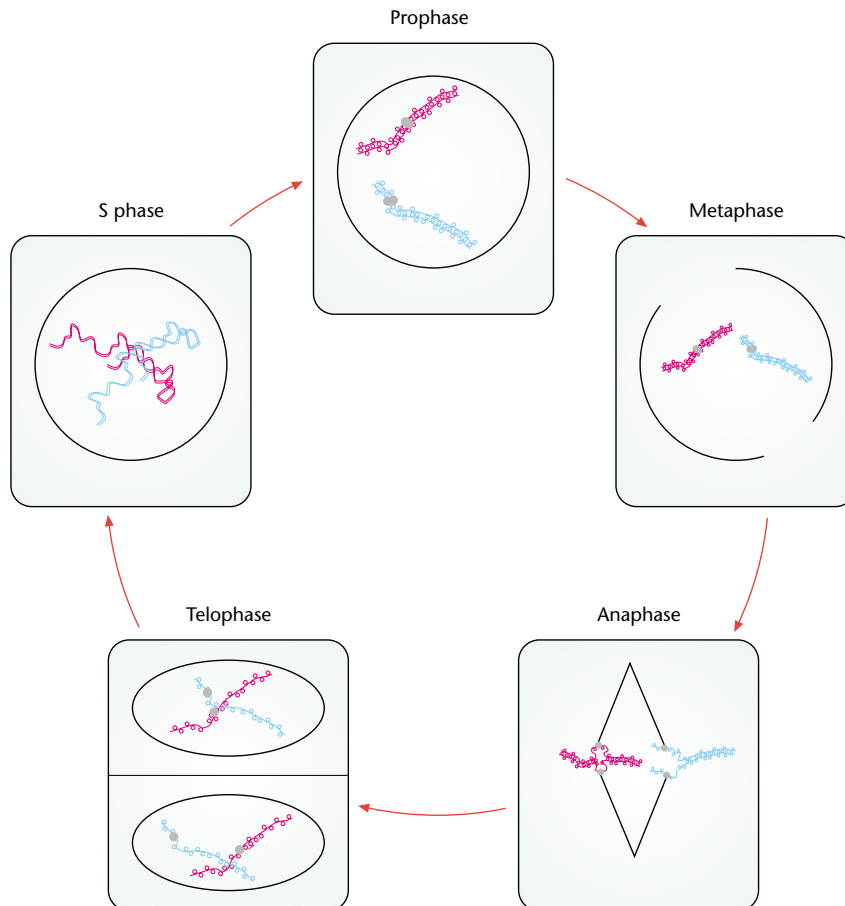


Figure 5 Chromosome segregation in eukaryotes is completed in the mitotic cycle. Mitosis proceeds through four stages, prophase, metaphase, anaphase and telophase, as described in the text.

nitrogen source. One key to efficient growth is control of gene expression at the level of RNA transcription. The process of making of an RNA complement to a DNA gene is called transcription.

To accommodate complex regulatory scenarios, chromosomes are divided into discrete transcriptional modules called operons. An operon has two functional components: the operator (or control region) and the coding sequence, which encodes the protein or RNA product. Although early studies focused on operons such as *lac*, *trp* and *his*, which produce several proteins from a single operator, 73% of the 2600 different *E. coli* operons encode only one protein. The operator includes at least one binding site for RNA polymerase, which is the enzyme that makes a complementary RNA copy of DNA. RNA polymerase has a core subunit, plus a specificity factor called sigma (σ). Sigma factors direct RNA polymerase to bind a specific subset of promoters. In *E. coli* there are four sigma factors: σ^{70} , which controls many housekeeping genes and most intermediary metabolism operons; σ^E , which is specific for nitrogen fixation genes; σ^S , which controls genes that are turned on in stationary phase; and *fliA*, which is specific for flagellar genes.

In addition to RNA polymerase, promoters contain both positive and negative regulatory proteins. Negative regulatory proteins are called repressors, and they block transcription by either impeding access of RNA polymerase or by inhibiting its ability to initiate RNA chains. The LacI repressor blocks expression of the lactose genes when no lactose is available for metabolism. There are also positively acting proteins called activators, which stimulate RNA polymerase transcription, either by stabilizing its binding to a promoter or by stimulating the ability of RNA polymerase to initiate RNA chains. One example of a well-characterized activator protein is the cyclic adenosine monophosphate (cAMP)-binding protein (CAP), which binds to a large number of operator sites and stimulates transcription when cAMP levels are high.

In bacteria, replication and transcription can occur simultaneously. Replication requires all attached, double strand-specific DNA-binding proteins to be temporarily displaced during complementary strand unwinding and synthesis. The proteins that have to be removed include the chromosome-associated proteins (see section on Chromatin above), all repressors and operon activator proteins, and the RNA polymerase. Transcription presents a special problem for the replisome. Because transcription and replication occur simultaneously, two situations arise where transcription machinery and replication machinery collide. One case is where transcription and replication move in the same direction. Because a replisome pulls in DNA at 500 bp per second and RNA polymerase transcribes DNA at 50 bp per second, replication forks rapidly overtake RNA polymerase, even when both enzymes are headed in the same direction. The more striking encounter occurs when DNA replication meets a

transcribing RNA polymerase molecule moving opposite to the direction of replication fork movement. In both cases, the replisomes can pass RNA polymerases without causing them to release transcripts or dissociate from DNA. However, most highly transcribed genes are transcribed in the same direction as replication forks move. This includes all seven of the ribosomal RNA operons (black arrowheads in **Figure 2b**) and 53 of the 86 transfer RNA genes.

In eukaryotic chromosomes, the RNA polymerase responsible for transcribing most genes is remarkably similar at the structural level to the *E. coli* RNA polymerase. However, regulatory mechanisms are different. Eukaryotic genes have a region called the promoter which is where RNA polymerase binds and starts transcription. However, polymerase binding and its ability to initiate transcription is influenced by sites called transcription enhancers that can be upstream or downstream of the promoter (relative to the direction of transcription). Enhancers act over very large distances, and so DNA looping is required to bring enhancers into contact with RNA polymerase at promoters. In addition to enhancers, there are proteins called co-activators that must bind to RNA polymerase to stimulate transcription. While there are many questions about how co-activators work, one important known function is to modify histones by acetylating the H3 and H4 subunits.

Recombination

Chromosomes undergo three types of genetic recombination: homologous recombination, site-specific recombination and transposition. Homologous recombination involves exchanges between regions of identical or nearly identical sequences that are 300 bp or longer, with the efficiency of recombination increasing up to several kbp. Several repetitive sequences, including the seven copies of the ribosomal RNA operon in the *E. coli* genome, are large enough to stimulate efficient recombination (**Figure 2b**). In *E. coli*, the pathway of homologous recombination is carried out and regulated by four genes: *recA*, *recB*, *recC* and *recD*. Homologous recombination is responsible for the introduction of new information into the bacterial genome through mechanisms of transformation, conjugation and transduction. However, another important function of recombination is DNA repair. Recombination allows replication forks that have stalled or fallen apart to restart DNA synthesis, in addition to allowing one damaged chromosome to be repaired using information from a sister chromosome. Thus, in addition to being an agent of change, recombination helps chromosomes remain the same.

Homologous recombination is critical for repair of DNA damage caused by chemical modification by

alkylating agents, by ultraviolet light, and X-ray-induced chromosome breaks. DNA damage is also caused by internal oxygen metabolism, which generates free radicals. Replication errors also cause DNA breaks. The RecABC pathway corrects a damaged copy of a chromosome by using identical sequences in a sister chromosome in the same cell. As noted above, when rapid growth is underway, cells have more than one chromosome. RecA protein serves two roles in repairing chromosome damage. First, it binds to single-stranded regions of a broken chromosome and facilitates a search for the homologous sequence, and then forms a hybrid molecule that starts recombination. RecA also regulates gene expression. When DNA damage occurs, RecA protein inactivates a repressor called LexA, which results in the expression of over a dozen DNA repair proteins. After DNA repair is complete, RecA protein stops inactivating LexA repressor and DNA metabolism returns to normal.

Recombination is a critical repair pathway in mammalian chromosomes as well. Proteins that carry out biochemical reactions similar to the *E. coli* RecABC system have been identified. A protein called p53 coordinates the activity of many DNA repair proteins. Repair enzymes are stored at chromosome telomeres, and after a signal from p53 these proteins migrate to sites of DNA damage to restore chromosome function. Mutations in DNA repair genes have been discovered to be responsible for several human genetic syndromes that result in premature ageing and high spontaneous rates of cancer.

Chi Sites

One sequence in *E. coli* influences the efficiency and direction of recombination: the *chi* sequence, GCTGGTGG. A *chi* site stimulates recombination directionally, by triggering the pairing of chromosome sequences near *chi*. Like *ter* sites in replication, *chi* sites act in only one direction. About 75% of all *chi* sites are arranged in the clockwise direction in replicore I and counter-clockwise in replicore II, so that they instruct recombination to proceed according to the direction of DNA replication.

Transposons

Transposons are discrete genetic elements that move from one site or from one chromosome to another, often with little regard to any specific DNA target sequence. Transposons sometimes carry with them genes for resistance to antibiotics such as tetracycline (Tn10) and kanamycin (Tn5). A transposon generally has inverted repeats at the ends of the element, and encodes at least one

protein called the transposase, which binds the ends and stimulates transposon movement. The chromosome of *E. coli* has 42 different transposons that have invaded by horizontal transfer from other bacteria or viruses.

In eukaryotes, transposons (usually called retrotransposons because of their similarity to retroviruses and their dependence on reverse transcriptase for replication) make up a large fraction of total chromosomal DNA. In human cells, sequences called short interspersed nuclear elements (SINEs), which are about 300 bp long are present in about 10^6 copies and represent 5% of the mass of DNA in a haploid genome. One particular SINE, called AluI, is present on average once every 5000 bp in every human chromosome. There are also long interspersed nuclear elements (LINEs) of about 6 kb that are present in about 10^5 copies and represent 15% of the haploid chromosomal mass. What function, if any, these sequences provide for the host organism is questionable, but many genetic mutations have been attributed to gene disruption caused by recent transposon insertions.

Site-specific Recombination

Site-specific recombination provides an efficient mechanism for rearranging DNA at sites less than 100 bp in length (below the level required for homologous recombination). Site-specific recombination systems cause insertion and excision of different lysogenic viruses, inversion of regions flanked by inverted sites, and deletions when two sites are directly repeated in a chromosome. Bacteriophage λ contains the best-studied site-specific recombination system – the Int/Xis system – which allows a prophage to integrate into the bacterial chromosome at one point called *attB*, or subsequently to excise and replicate in the lytic mode. Site-specific systems use a protein recombinase to bind and recombine short sequences, usually about 20 bp long.

One site-specific recombination system in *E. coli* plays a crucial role in chromosome segregation. DNA synthesis may generate breaks (often on the discontinuous strand) that stimulate recombination between daughter chromosomes. If an odd number of crossovers occurs between daughters, the chromosomes will be dimerized at the completion of DNA replication. In *E. coli*, about 15% of the normal replication cycles produce dimeric chromosomes. The site-specific recombination system that resolves these molecules is composed of the *dif* site at the terminus (**Figure 2**) and recombination proteins XerC and XerD, which efficiently separate daughter chromosomes just before cell division.

In eukaryotes, site-specific recombination also has many critical roles in biological development. One spectacular example is the mammalian immune system, where different chromosomal segments called V, D and J genes are cut and

spliced together to produce antibodies. The mammalian immune system can make antibodies to a diverse array of different structures by cutting and pasting genes together in an immense number of different ways. The antibody system is essential for life, because it protects an organism from invasion by bacterial and viral pathogens. Thus, site-specific recombination has evolved as one key mechanism that changes the genome in a small subset of differentiated cells, and this ultimately allows an organism to survive and stay the same.

With multiple transposons, efficient homologous and site-specific recombination systems, the occurrence of frequent chromosome breaks, and the presence of multiple long inverted and direct repeats, it would seem impossible for chromosomes to remain constant through time. Surprisingly, the genetic maps of *E. coli* and *S. typhimurium* are very similar after 140 million years of separation from a common ancestor. Thus, the efficiency of enzyme

systems designed to promote chromosome stability is nearly equal to the forces that promote change.

Further Reading

- Casjens S (1998) Bacterial genome structure. *Annual Review of Genetics* **32**: 339–377.
- Charlebois RL (ed) (1999) *Organization of the Prokaryotic Genome*. Washington DC: ASM Press.
- Cozzarelli NR and Wang JC (1990) *DNA Topology and its Biological Effects*. Cold Spring Harbor, New York: Cold Spring Harbor Press.
- Kornberg A and Baker T (1991) *DNA Replication*, 2nd edn. New York: WH Freeman.
- Kuzminov A (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage λ . *Microbiology and Molecular Biology Reviews* **63**: 751–813.
- Neidhardt FC (ed.) (1996) *Escherichia coli and Salmonella*. Washington DC: ASM Press.