F1 DNA STRUCTURE

Key Notes

Bases

In DNA there are four bases: adenine (abbreviated A), guanine (G), thymine (T) and cytosine (C). Adenine and guanine are purines; thymine and cytosine are pyrimidines.

Nucleosides

A nucleoside is a pyrimidine or purine base covalently bonded to a sugar. In DNA, the sugar is deoxyribose and so this is a deoxynucleoside. There are four types of deoxynucleoside in DNA; deoxyadenosine, deoxyguanosine, deoxythymidine and deoxycytidine.

Nucleotides

A nucleotide is base + sugar + phosphate covalently bonded together. In DNA, where the sugar is deoxyribose, this unit is a deoxynucleotide.

3′5′ phosphodiester bonds

In DNA the nucleotides are covalently joined together by 3′5′ phosphodiester bonds to form a repetitive sugar–phosphate chain which is the backbone to which the bases are attached.

DNA sequence

The DNA sequence is the sequence of A, C, G and T along the DNA molecule which carries the genetic information.

DNA double helix

In a DNA double helix, the two strands of DNA are wound round each other with the bases on the inside and the sugar–phosphate backbones on the outside. The two DNA chains are held together by hydrogen bonds between pairs of bases; adenine (A) always pairs with thymine (T) and guanine (G) always pairs with cytosine (C).

Related topics

DNA replication in bacteria (F3)
DNA replication in eukaryotes (F4)
RNA structure (G1)
Transcription in prokaryotes (G2)
Transcription in eukaryotes: an overview (G4)

Bases

The bases in DNA have carbon–nitrogen ring structures; because of the nitrogen atoms they are called nitrogenous bases. There are two types of ring structure. **Adenine** and **guanine** are **purines** (*Fig. 1a*), each having two joined carbon–nitrogen rings but with different side-chains. **Thymine** and **cytosine** are **pyrimidines** (*Fig. 1a*); each has only one carbon–nitrogen ring and again they differ in their side-chains.

Nucleosides

In RNA, the nucleosides have ribose as the sugar component (see Topic G1) and so are ribonucleosides. In DNA the sugar is deoxyribose (*Fig. 1b*) (i.e. the 2′-OH group in ribose is replaced by a hydrogen atom; hence 'deoxy') and so the nucleosides are **deoxynucleosides**. For DNA these are **deoxyadenosine, deoxyguanosine, deoxythymidine** and **deoxycytidine**. In each case, the C-1 of the sugar is joined to the base via one of its nitrogen atoms. If the base is a pyrimidine, the

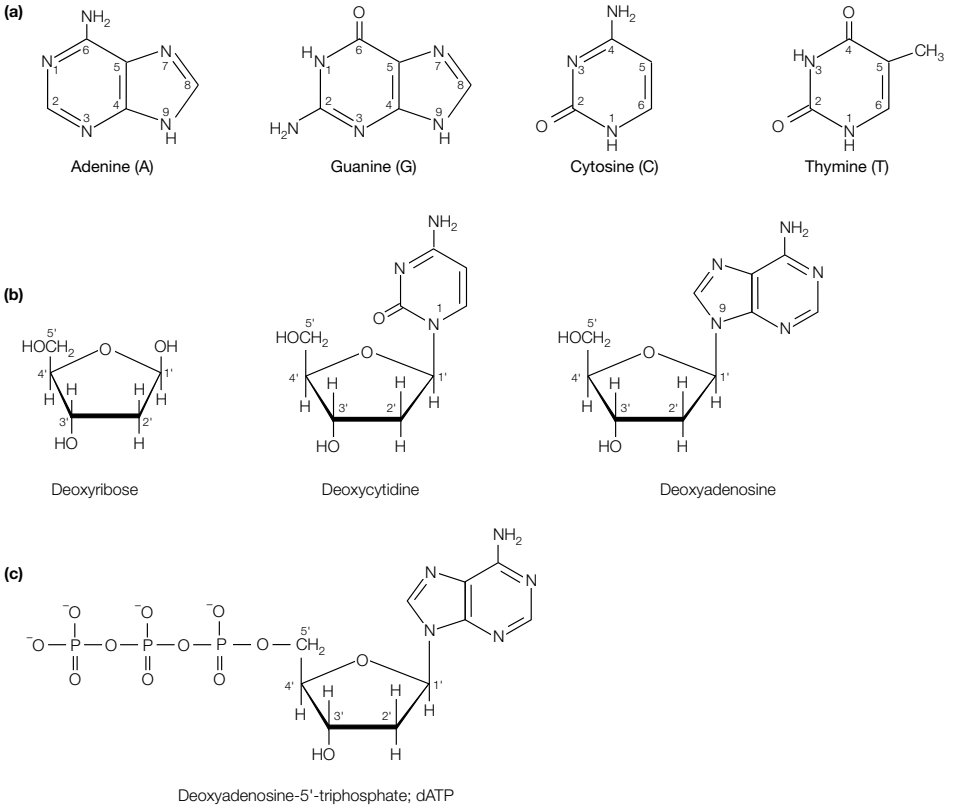


Fig. 1. (a) The purines, adenine and guanine, and the pyrimidines, thymine and cytosine; (b) deoxyribose and two deoxynucleosides, deoxycytidine and deoxyadenosine; (c) a deoxynucleotide, deoxyadenosine 5′-triphosphate (dATP).

nitrogen at the 1 position (i.e. N-1) is involved in bonding to the sugar. If the base is a purine, the bonding is to the N-9 position of the base (*Fig. 1b*).

Nucleotides

A nucleotide is a phosphate ester of a nucleoside. It consists of a phosphate group joined to a nucleoside at the hydroxyl group attached to the C-5 of the sugar, that is it is a **nucleoside 5′-phosphate** or a **5′-nucleotide**. The primed number denotes the atom of the sugar to which the phosphate is bonded. In DNA the nucleotides have deoxyribose as the sugar and hence are called **deoxynucleotides**. Deoxynucleotides may have a single phosphate group (**deoxynucleoside 5′-monophosphates**, dNMPs), two phosphate groups (**deoxynucleoside 5′-diphosphates**, dNDPs) or three phosphate groups (**deoxynucleoside 5′-triphosphates**, dNTPs). Deoxynucleoside triphosphates are the precursors for DNA synthesis. These are deoxyadenosine 5′-triphosphate (dATP) (*Fig. 1c*), deoxyguanosine 5′-triphosphate (dGTP), deoxycytidine 5′-triphosphate (dCTP) and deoxythymidine 5′-triphosphate (dTTP). In each case the 'd' in the abbreviation (for example in dATP) indicates that the sugar in the nucleotide is

deoxyribose. During DNA synthesis (see Topics F3 and F4), two of the phosphates of each deoxynucleotide are split off (as pyrophosphate) so that only a single phosphate (the α phosphate) is incorporated into DNA.

3′5′ phosphodiester bonds

In a DNA molecule, the different nucleotides are covalently joined to form a long polymer chain by covalent bonding between the phosphates and sugars. For any one nucleotide, the phosphate attached to the hydroxyl group at the 5′ position of the sugar is in turn bonded to the hydroxyl group on the 3′ carbon of the sugar of the next nucleotide. Since each phosphate–hydroxyl bond is an ester bond, the linkage between the two deoxynucleotides is a **3′5′ phosphodiester bond** (*Fig. 2*). Thus, in a DNA chain, all of the 3′ and 5′ hydroxyl groups are involved in phosphodiester bonds except for the first and the last nucleotide in the chain. The first nucleotide has a 5′ phosphate not bonded to any other nucleotide and the last nucleotide has a free 3′ hydroxyl. Thus each DNA chain has **polarity;** it has a **5′ end** and a **3′ end**.

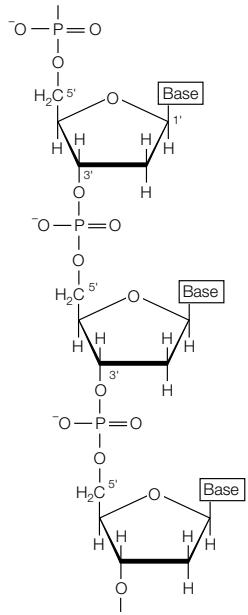


Fig. 2. 3′5′ phosphodiester bonds formed between nucleotides in a DNA molecule.

DNA sequence

Each nucleotide can be thought of as a single letter in an alphabet that has only four letters, A, G, C and T. Different genes have different sequences of these four nucleotides and so code for different biological messages. Since the deoxynucleotides in DNA differ only in the bases they carry, this sequence of deoxynucleotides can be recorded simply as a **base sequence**. For example, ACTTTCAGACC is part of the base sequence of one gene and codes for part of one protein whereas TGGAACCGTCA is part of the base sequence of a different gene coding for a different protein. Traditionally the base sequence is written in

the order from the 5′ end of the DNA strand to the 3′ end, i.e. it is **written in the 5′→3′ direction**. Given that there are four types of nucleotide, the number of different possible sequences (or messages) in a DNA strand n nucleotides long is 4^n. DNA molecules are typically many thousands of nucleotides long so that the number of possible messages is enormous.

DNA double helix

In 1953, Watson and Crick worked out the three-dimensional structure of DNA, starting from X-ray diffraction photographs taken by Franklin and Wilkins. They deduced that DNA is composed of two strands wound round each other to form a double helix, with the bases on the inside and the sugar–phosphate backbones on the outside. In the double helix (*Fig.*3), the two DNA strands are organized in an **antiparallel** arrangement (i.e. the two strands run in opposite directions, one strand is orientated 5′→3′ and the other is orientated 3′→5′). The bases of the two strands form hydrogen bonds to each other; A pairs with T and G pairs with C. This is called **complementary base pairing** (*Fig. 4*). Thus a large two-ringed purine is paired with a smaller single-ringed pyrimidine and the two bases fit neatly in the gap between the sugar–phosphate strands and maintain the correct spacing. There would be insufficient space for two large purines to pair and too much space for two pyrimidines to pair, which would be too far apart to bond. The G:C and A:T base pairing also maximizes the number of effective hydrogen bonds that can form between the bases; there are three hydrogen bonds between each G:C base pair and two hydrogen bonds between each A:T base pair. Thus A:T and G:C base pairs form the most stable conformation both from steric considerations and from the point of view of maximizing hydrogen bond formation.

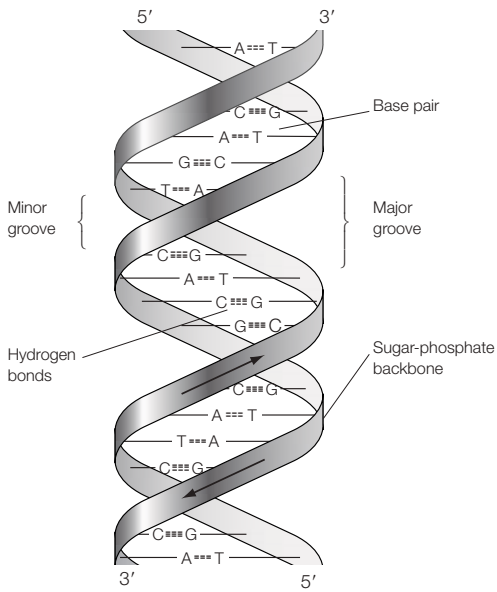
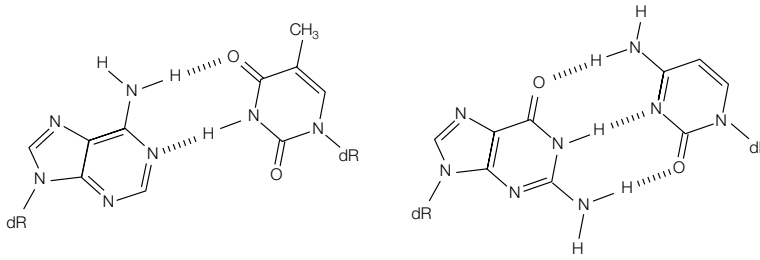


Fig. 3. The DNA double helix.



Adenine : thymine

Guanine : cytosine

Fig. 4. The DNA base pairs. Hydrogen bonds are shown as dashed lines. dR, deoxyribose.

F2 GENES AND CHROMOSOMES

Key Notes

Concept of the gene

The original concept of a gene was a region of DNA that encoded a single polypeptide (or RNA) product. However, post-transcriptional mechanisms in eukaryotes are now known that can generate multiple (sequence-related) polypeptides from a single RNA transcript so the working definition needs to be modified to accommodate this new knowledge.

Prokaryotic chromosomes

The DNA in a bacterium is a supercoiled double-stranded circular molecule that is packaged in the nucleoid region of the cell. The DNA is negatively supercoiled, complexed to several histone-like proteins (mainly proteins HU, HSP-1 and H-NS) and organized into about 50 domains bound to a protein scaffold.

Eukaryotic chromosomes

Eukaryotic cells contain much more DNA than prokaryotes. In the nucleus, the DNA is packaged into chromosomes that consist mainly of DNA and proteins called histones although other nonhistone proteins (NHP) are also present. Each chromosome contains a single linear double-stranded DNA molecule.

Nucleosomes

The chromosomal DNA is complexed with five types of histone (H1, H2A, H2B, H3 and H4). These are very basic proteins, rich in arginine and lysine. The amino acid sequences of histones are highly conserved in evolution. The DNA is wound round a histone octamer (two molecules each of H2A, H2B, H3 and H4) to form a nucleosome. The DNA between neighboring nucleosomes (linker DNA) binds histone H1. The packing ratio of nucleosomes is about 7.

30 nm fiber

Nucleosomes are organized into a 30 nm fiber. The exact arrangement of nucleosomes in the 30 nm fiber is unclear; possibilities are a higher order helix (called a solenoid) or a zigzag arrangement. The overall packing ratio is about 40.

Radial loops

The 30 nm fiber is attached to a central protein scaffold in each chromosome in a series of radial loops.

Related topics

DNA structure (F1)

DNA replication in eukaryotes (F4)

Concept of the gene

By the 1960s, the gene was clearly defined as the region of DNA that gives rise to a single polypeptide (or to a single RNA for genes whose final product is RNA not protein, e.g. ribosomal RNA genes). The existence of operons in prokaryotes (see Topic G3) did not challenge this concept since, although several clustered genes produced a single polycistronic mRNA, one could still

identify single DNA regions as genes based on the distinct polypeptides they encoded. The concept even accommodated the discovery that many protein-coding genes in eukaryotes comprise coding regions (exons) separated by noncoding sequences (introns) (see Topic G5) since, again, only one polypeptide was encoded by this region of DNA. More recently, however, other mechanisms have come to light in eukaryotic cells that can lead to a variety of polypeptides being produced from a single DNA sequence; for example, alternative RNA splicing, alternative polyadenylation sites and RNA editing (see Topic G7). Nevertheless, in each of these cases, the protein products are closely related by sequence and all are derived from the same single region of DNA. Thus the original definition perhaps needs tweaking to indicate that a protein-coding gene is a region of DNA that encodes a single polypeptide or a set of closely-related polypeptides, but otherwise the definition is intact. The alternative scenario, to regard a single DNA sequence that gives rise to, say, 10 closely-related polypeptides by post-transcriptional processing as representing 10 genes, would certainly not fit in with accepted practice.

The genome of an organism encompasses all of the genes of that organism. Even in a bacterial cell such as *Escherichia coli* (*E. coli*), the amount of DNA required is substantial (4.6 million base pairs) and so this DNA must be packaged. In a eukaryotic cell, the problem is even greater. A typical human cell, for example, contains about 1000 times more DNA than an *E. coli* cell. The rest of this Topic describes how DNA is packaged in both prokaryotes and eukaryotes.

Prokaryotic chromosomes

The DNA of a bacterial cell, such as *Escherichia coli*, is a circular double-stranded molecule often referred to as the **bacterial chromosome**. The circular DNA is packaged into a region of the cell called the **nucleoid** (see Topic A1) where it is organized into 50 or so **loops** or **domains** that are bound to a central **protein scaffold**, attached to the cell membrane. *Fig.1a* illustrates this organization, although only six loops are shown for clarity. Within this structure, the DNA is actually not a circular double-stranded DNA molecule such as that shown in *Fig.1b* but is negatively **supercoiled**, that is, it is twisted upon itself (*Fig. 1c*) and is also complexed with several **DNA-binding proteins**, the most common of which are proteins **HU**, **HLP-1** and **H-NS**. These are **histone-like proteins** (see below for a description of histones).

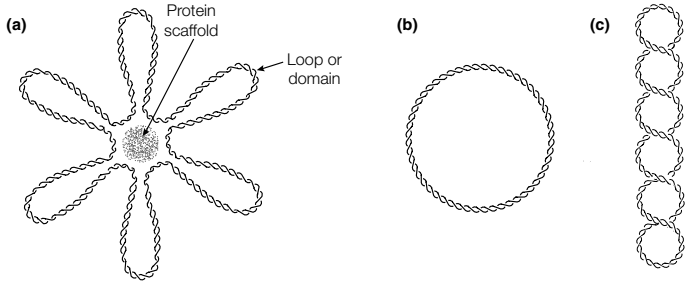


Fig. 1. (a) The association of circular bacterial DNA with a protein scaffold; (b) a circular double-stranded DNA molecule; (c) supercoiled DNA.

Eukaryotic chromosomes

The large amount of genomic DNA in a eukaryotic cell is tightly packaged in chromosomes contained within a specialized organelle, the nucleus. This very large amount of eukaryotic nuclear DNA is tightly packaged in chromosomes. With the exception of the sex chromosomes, diploid eukaryotic organisms such as humans have two copies of each chromosome, one inherited from the father and one from the mother. Chromosomes contain both DNA and protein. Most of the protein on a weight basis is **histones**, but there are also many thousands of other proteins found in far less abundance and these are collectively called **nonhistone proteins** (NHP). This nuclear DNA–protein complex is called **chromatin**. The mitochondria and chloroplasts of eukaryotic cells also contain DNA but, unlike the nuclear DNA, this consists of double-stranded circular molecules resembling bacterial chromosomes.

In the nucleus, each chromosome contains a single linear double-stranded DNA molecule. The length of the packaged DNA molecule varies. In humans, the shortest DNA molecule in a chromosome is about 1.6 cm and the longest is about 8.4 cm. During the **metaphase** stage of mitosis, when the chromosomes align on the mitotic spindle ready for segregation, they are at their most condensed and range in size from only 1.3 μm to 10 μm long. Thus the **packing ratio**, that is the ratio of the length of the linear DNA molecule to the length of the metaphase chromosome, is about 10^4. In the time period between the end of one mitosis and the start of the next (i.e. **interphase**), the chromatin is more disperse. Here the packing ratio is in the range 10^2–10^3. Overall, the extensive packaging of DNA in chromosomes results from three levels of folding involving nucleosomes, 30 nm filaments and radial loops.

Nucleosomes

The first level of packaging involves the binding of the chromosomal DNA to histones. Overall, in chromosomes, the ratio of DNA to histones on a weight basis is approximately 1:1. There are five main types of histones called H1, H2A, H2B, H3 and H4. Histones are very basic proteins; about 25% of their amino acids are lysine or arginine so histones have a large number of positively charged amino acid side-chains. These positively charged groups therefore bind to the negatively charged phosphate groups of DNA. Not surprisingly given their importance in packaging DNA, the amino acid sequences of histones have been highly conserved in evolution. The most conserved are histones H3 and H4; for example, H3 and H4 from peas and cows differ in only four and two amino acids respectively! Histone H1 is the least conserved histone, which reflects its somewhat different role in packaging DNA compared with the other histones (see below). In sperm heads, DNA is particularly highly condensed and here the histones are replaced with small basic proteins called **protamines**.

When chromosomes are gently 'decondensed', they have the appearance under the electron microscope of 'beads on a string' (*Fig. 2*). The 'beads' are called **nucleosomes** and consist of DNA complexed with histones. The 'string' is linear double-stranded DNA between adjacent nucleosomes and is called **linker DNA** (*Fig. 2*). The average distance between nucleosomes, that is the length of the linker DNA, is typically about 55 base pairs (bp) but varies greatly from organism to organism. Even in a single nucleus, the distance between adjacent nucleosomes varies depending on, for example, the presence of other sequence-specific DNA-binding proteins. If a chromatin preparation is incubated with micrococcal nuclease, an enzyme that degrades DNA, the linker DNA is destroyed leaving **nucleosome core particles** in which the histones protect the associated DNA from digestion. Each nucleosome core particle contains a

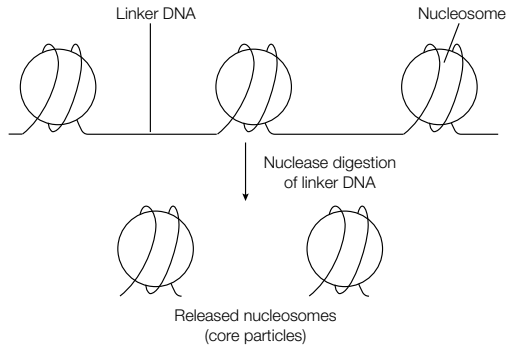


Fig. 2. 'Beads-on-a-string' structure of chromatin.

double-stranded DNA fragment 146 bp long bound to a complex of eight histones, the **histone octamer**, consisting of two molecules each of histones H2A, H2B, H3 and H4 (*Fig. 3*). The DNA is wound round the outside of the histone octamer in about 1.65 turns of a left-handed supercoil. DNA–histone contacts are made along the inside face of this superhelix. Overall the packing ratio is about 7, that is the DNA length is shortened about seven-fold by winding around the nucleosome.

30 nm fiber

If nuclei are lysed very gently, the chromatin is seen to exist as a 30 nm diameter fiber. This diameter is much larger than a single nucleosome (which is about 11 nm) and suggests that the nucleosomes are organized into a higher order structure. The fiber is formed by a histone H1 molecule binding to the linker DNA of each nucleosome at the point where it enters and leaves the nucleosome (*Fig. 3*). The histone H1 molecules interact with each other, pulling the nucleosomes together. Exactly how the nucleosomes are organized to form the 30 nm fiber is not known; one possibility is that the nucleosomes wind up into a higher order helix with six nucleosomes per turn to form a **solenoid** (*Fig. 4*). This would give a fiber three nucleosomes wide, which is indeed the diameter observed. In such

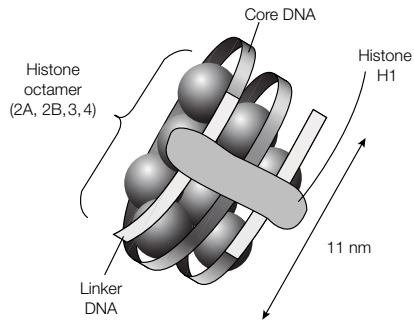


Fig. 3. Schematic diagram of a nucleosome consisting of the DNA double helix wound 1.8 times round a histone octamer (two molecules each of histones H2A, H2B, H3 and H4).

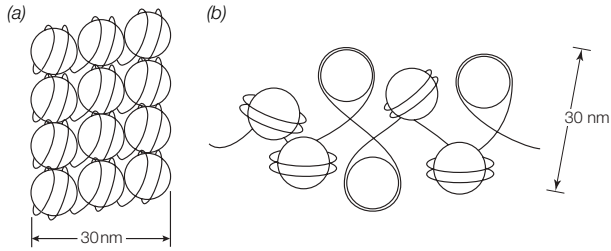


Fig. 4. (a) The proposed solenoid model of chromatin to yield a 30 nm fiber. The structure consists of six nucleosomes per turn of the helix and hence would be three nucleosomes wide. In the diagram, only three nucleosomes of each turn are visible; the other three nucleosomes per turn are hidden from view. (b) Zigzag model. For each nucleosome, the DNA is wound round the outside (1.65 turns of a left-handed supercoil). However in this simple diagram, because of the orientation of some nucleosomes, not all of the DNA supercoil is shown.

a solenoid the linear length of the DNA has been reduced by a further factor of 6 (equivalent to six nucleosomes per turn of the solenoid). Coupled with the packing ratio of 7 for the nucleosome itself (see above), this gives a packing ratio for the solenoid of approximately 6×7 (i.e. about 40). An alternative model is the **zigzag model** (*Fig. 4b*).

Radial loops

When chromosomes are depleted of histones, they are seen to have a central fibrous '**protein scaffold**' (or **nuclear matrix**) to which the DNA is attached in loops (*Fig. 5*). Therefore, *in vivo* it seems likely that the next order of packaging involves the attachment of the 30 nm fiber to multiple locations on this central protein scaffold in a series of radial loops. Little is known as to how this structure is organized.

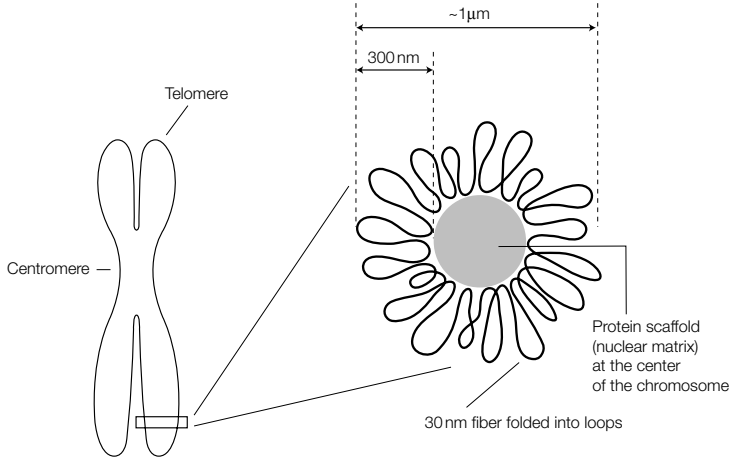


Fig. 5. Attachment of the 30nm fiber to a central protein scaffold with the loops arranged radially around the scaffold. The diagram on the right shows a representational cross-section through a chromosome.

F3 DNA REPLICATION IN BACTERIA

Key Notes

DNA polymerases

E. coli DNA polymerase I requires all four deoxynucleoside 5′-triphosphates (dNTPs) as precursors, Mg^{2+}, a DNA template and a primer with a 3′-OH end. DNA synthesis occurs in a 5′ → 3′ direction. DNA polymerase I also has a 3′ → 5′ exonuclease (proof-reading) activity and a 5′ → 3′ exonuclease activity. *E. coli* DNA polymerases II and III lack the 5′ → 3′ exonuclease activity.

Replication forks

Replication starts at a single origin, is bi-directional and semi-conservative. Each replication bubble (or eye) consists of two replication forks.

Okazaki fragments

DNA synthesis proceeds in a 5′ → 3′ direction on each strand of the parental DNA. On the strand with 3′ → 5′ orientation (the leading strand) the new DNA is synthesized continuously. On the strand that has 5′ → 3′ orientation (the lagging strand) the DNA is synthesized discontinuously as a series of short Okazaki fragments that are then joined together.

RNA primer

DNA replication requires an RNA primer that is synthesized by an RNA polymerase called primase. This is extended by DNA polymerase III, which makes the DNA for both the leading and lagging strands. DNA polymerase degrades the primer and replaces it with DNA. DNA ligase then joins DNA ends.

Accessory proteins

A helicase unwinds the DNA double helix and single-stranded DNA-binding (SSB) protein stabilizes the single-stranded regions during replication. DNA topoisomerase I is needed to allow the helix to unwind without causing extensive rotation of the chromosome. DNA topoisomerase II separates the two daughter DNA circles following replication.

Related topics

DNA structure (F1)

DNA replication in eukaryotes (F4)

DNA polymerases

DNA polymerase I from *E. coli* catalyzes the stepwise addition of deoxyribonucleotides to the 3′-OH end of a DNA chain:

$$(DNA)_{n\ \text{residues}} + dNTP \rightarrow (DNA)_{n+1} + PP_i$$

The enzyme has the following requirements:

- all four dNTPs (dATP, dGTP, dTTP and dCTP) must be present to be used as precursors; Mg^{2+} is also required;

- a DNA **template** is essential, to be copied by the DNA polymerase;
- a **primer** with a free 3′-OH that the enzyme can extend.

DNA polymerase I is a template-directed enzyme, that is it recognizes the next nucleotide on the DNA template and then adds a complementary nucleotide to the 3′-OH of the primer, creating a 3′5′ phosphodiester bond, and releasing pyrophosphate. The reaction is shown in *Fig. 1*. It involves nucleophilic attack of the 3′-OH of the primer on the α-phosphate group of the incoming nucleotide. The primer is extended in a 5′ → 3′ direction.

DNA polymerase I also corrects mistakes in DNA by removing mismatched nucleotides (i.e. it has **proof-reading activity**). Thus, during polymerization, if the nucleotide that has just been incorporated is incorrect (mismatched), it is removed using a 3′ → 5′ exonuclease activity. This gives very high fidelity; an error rate of less than 10^{-8} per base pair. DNA polymerase also has a 5′ → 3′ exonuclease activity; it can hydrolyze nucleic acid starting from the 5′ end of a chain. This activity plays a key role in removing the RNA primer used during replication (see below). Thus, overall, DNA polymerase I has three different active sites on its single polypeptide chain; 5′ → 3′ polymerase, 3′ → 5′ exonuclease and 5′ → 3′ exonuclease. As well as its role in DNA replication, DNA polymerase I is involved in **DNA repair**, for example removing UV-induced alterations such as pyrimidine dimers.

E. coli also contains two other DNA polymerases, **DNA polymerase II** and **DNA polymerase III**. As with DNA polymerase I, these enzymes also catalyze the template-directed synthesis of DNA from deoxynucleotidyl 5′-triphosphates, need a primer with a free 3′-OH group, synthesize DNA in the 5′ → 3′ direction, and have 3′ → 5′ exonuclease activity. Neither enzyme has 5′ → 3′ exonuclease activity.

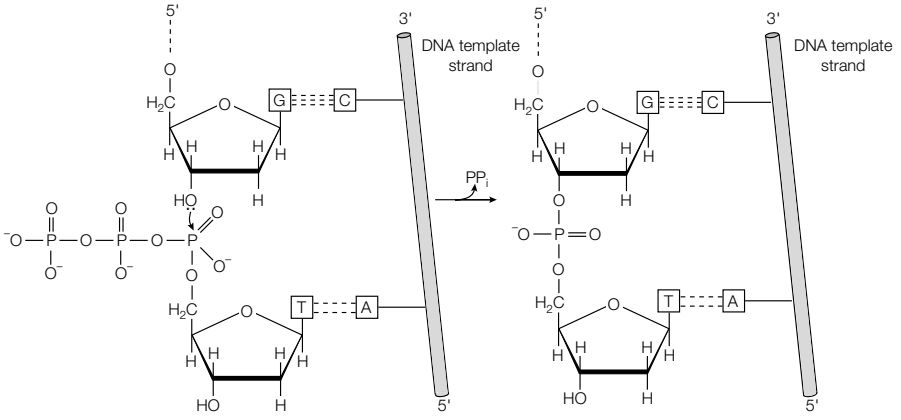


Fig. 1. DNA synthesis. In this schematic diagram, the incoming dTTP hydrogen bonds with the adenine on the template DNA strand and a 3′5′ phosphodiester bond is formed, releasing pyrophosphate.

Replication forks When the bacterial circular chromosome is replicated, replication starts at a **single origin**. The double helix opens up and both strands serve as template for the synthesis of new DNA. DNA synthesis then proceeds outward in both direc-

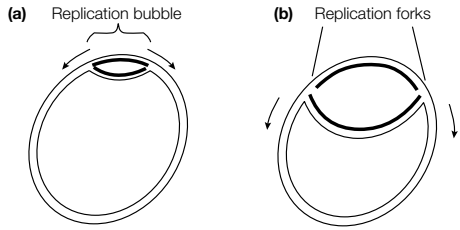


Fig. 2. Replication of the bacterial circular chromosome. Replication starts from a single origin and proceeds bi-directionally (a) moving around the chromosome with time (b). The two replication forks eventually meet and fuse. The two circular daughter DNA molecules produced each have one original template DNA strand (thin line) and one new strand (thick line).

tions from the single origin (i.e. it is **bi-directional**; *Fig. 2*). The products of the reaction are two daughter double-stranded DNA molecules each of which has one original template strand and one strand of newly synthesized DNA. Thus, replication is **semi-conservative**. The region of replicating DNA associated with the single origin is called a **replication bubble** or **replication eye** and consists of two **replication forks** moving in opposite directions around the DNA circle (*Fig. 2*).

Okazaki fragments

Double-stranded DNA is **antiparallel** (see Topic F1); one strand runs 5′ → 3′ and the complementary strand runs 3′ → 5′. As the original double-stranded DNA opens up at a replication fork, new DNA is made against each template strand. Superficially, therefore, one might expect new DNA to be made 5′ → 3′ for one daughter strand and 3′ → 5′ for the other daughter strand. However, all DNA polymerases make DNA only in the 5′ → 3′ direction and never in the 3′ → 5′ direction. What actually happens is that on the template strand with 3′ → 5′ orientation, new DNA is made in a continuous piece in the correct 5′ → 3′ direction. This new DNA is called the **leading strand** (*Fig. 3*). On the other template strand (that has a 5′ → 3′ orientation), DNA polymerase synthesizes short pieces of new DNA (about 1000–2000 nucleotides long) in the 5′ → 3′ direction (*Fig. 3*) and then joins these pieces together. The small fragments are called **Okazaki fragments** after their discoverer. The new DNA strand which is made by this discontinuous method is called the **lagging strand**.

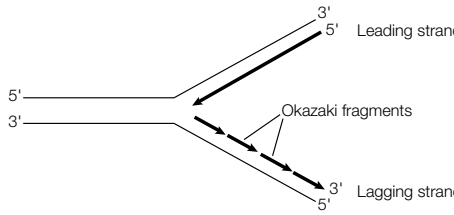


Fig. 3. Synthesis of DNA at a replication fork. As the parental DNA (thin line) opens up, each of the two parental strands acts as a template for new DNA synthesis (thick lines). The leading strand is synthesized continuously but the lagging strand is synthesized as short (Okazaki) DNA fragments that are then joined together.

RNA primer

DNA polymerase cannot start DNA synthesis without a primer. Even on the lagging stand, each Okazaki fragment requires an RNA primer before DNA synthesis can start. The primer used in each case is a short piece of RNA and is synthesized by an RNA polymerase called **primase** (*Fig. 4a*). Primase can make RNA directly on the single-stranded DNA template because, like all RNA polymerases, it does not require a primer to begin synthesis. The RNA primer made by primase (*Fig. 4b*) is then extended by DNA polymerase III (*Fig. 4c*). DNA polymerase III synthesizes DNA for both the leading and lagging strand. After DNA synthesis by DNA polymerase III, DNA polymerase I uses its 5′ → 3′ exonuclease activity to remove the RNA primer and then fills the gap with new DNA (*Fig. 4e* and *f*). DNA polymerase III cannot carry out this task because it lacks the 5′ → 3′ activity of DNA polymerase I. Finally, DNA ligase joins the ends of the DNA fragments together (*Fig. 4g*).

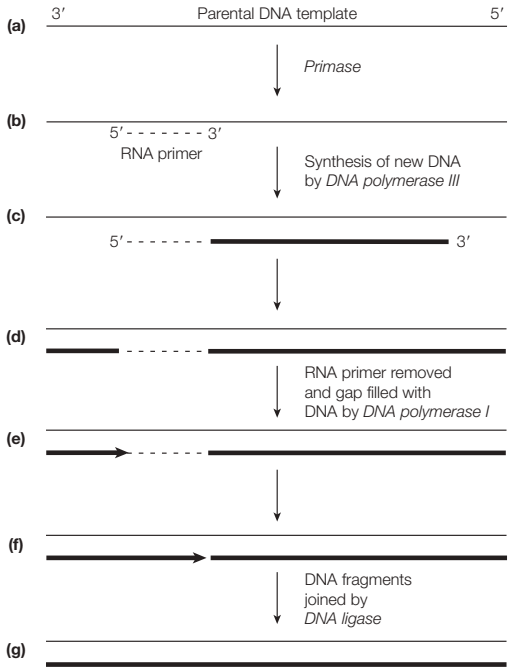


Fig. 4. Details of DNA replication. (a) Primase binds to the DNA template strand (thin line) and (b) synthesizes a short RNA primer (dotted line); (c) DNA polymerase III now extends the RNA primer by synthesizing new DNA (thick line); (d) during synthesis of the lagging strand, adjacent Okazaki fragments are separated by the RNA primers; (e) the RNA primers are now removed and the gaps filled with DNA by DNA polymerase I (f) generating adjacent DNA fragments that are then (g) joined by DNA ligase.

Accessory proteins

DNA polymerases I and III, primase and DNA ligase are not the only proteins needed for replication of the bacterial chromosome. The DNA template is a double helix with each strand wound tightly around the other and hence the two strands must be unwound during replication. How is this **unwinding**

problem solved? A **DNA helicase** (**Dna B helicase**) is used to unwind the double helix (using ATP as energy source) and **SSB (single-stranded DNA-binding) protein** prevents the single-stranded regions from base pairing again so that each of the two DNA strands is accessible for replication. In principle, for a replication fork to move along a piece of DNA, the DNA helix would need to unwind ahead of it, causing the DNA to rotate rapidly. However, the bacterial chromosome is circular and so there are no ends to rotate. The solution to the problem is that an enzyme called **topoisomerase I** breaks a phosphodiester bond in one DNA strand (a single-strand break) a small distance ahead of the fork, allowing the DNA to rotate freely (swivel) around the other (intact) strand. The phosphodiester bond is then re-formed by the topoisomerase.

After the bacterial circular DNA has been replicated, the result is two double-stranded circular DNA molecules that are interlocked. **Topoisomerase II** separates them as follows. This enzyme works in a similar manner to topoisomerase I but causes a transient break in each strand (a double-strand break) of a double-stranded DNA molecule. Thus topoisomerase II binds to one double-stranded DNA circle and causes a transient double-strand break that acts as a 'gate' through which the other DNA circle can pass (*Fig. 5*). Topoisomerase II then re-seals the strand breaks.

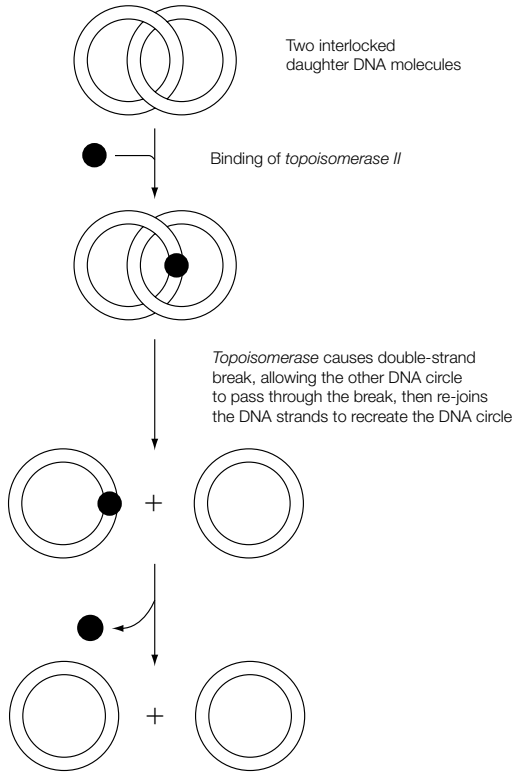


Fig. 5. Separation of daughter DNA circles by topoisomerase II.

F4 DNA REPLICATION IN EUKARYOTES

Key Notes

Cell cycle

In eukaryotes, the cell cycle consists of G_1, S, G_2 and M phases. Most differences in the cycle times of different cells are due to differences in the length of the G_1 phase. Quiescent cells are said to be in the G_0 phase.

Multiple replicons

DNA replication occurs only in the S phase. It occurs at many chromosomal origins, is bi-directional and semi-conservative. Sets of 20–80 replicons act as replication units that are activated in sequence.

Five DNA polymerases

DNA polymerases α and δ replicate chromosomal DNA, DNA polymerases β and ε repair DNA, and DNA polymerase γ replicates mitochondrial DNA.

Leading and lagging strands

DNA polymerase α and δ synthesize the lagging strand, via Okazaki fragments. The RNA primers are synthesized by DNA polymerase α which carries a primase subunit. DNA polymerase δ synthesizes the leading strand.

Telomere replication

Telomerase, a DNA polymerase that contains an integral RNA that acts as its own primer, is used to replicate DNA at the ends of chromosomes (telomeres).

Replication of chromatin

Nucleosomes do not dissociate from the DNA during DNA replication; rather they must open up to allow the replication apparatus to pass. Both daughter DNA molecules have old histones bound to them but new histones must also be synthesized to allow all the DNA to be packaged correctly in nucleosomes.

Related topics DNA structure (F1) DNA replication in bacteria (F3)

Cell cycle

The life of a eukaryotic cell can be defined as a **cell cycle** (*Fig. 1*). Mitosis and cell division occur in the **M phase** which lasts for only about 1 h. This is followed by the **G_1 phase** (G for gap), then the **S phase** (S for synthesis), during which time the chromosomal DNA is replicated, and finally the **G_2 phase** in which the cells prepare for mitosis. Eukaryotic cells in culture typically have cell cycle times of 16–24 h but the cell cycle time can be much longer (> 100 days) for some cells in a multicellular organism. Most of the variation in cell cycle times occurs by differences in the length of the G_1 phase. Some cells *in vivo*, such as neurons, stop dividing completely and are said to be quiescent, locked in a **G_0 phase**.

Multiple replicons

In eukaryotes, replication of chromosomal DNA occurs only in the S phase of the cell cycle. As for bacterial DNA (see Topic F3), eukaryotic DNA is replicated

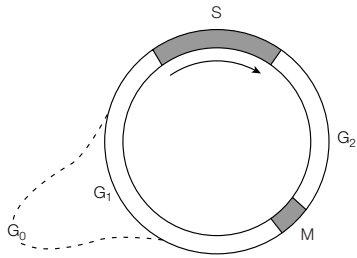


Fig. 1. The eukaryotic cell cycle. The S phase is typically 6–8 h long, G_2 is a phase in which the cell prepares for mitosis and lasts for 2–6 h, mitosis itself (M) is short and takes only about 1 h. The length of G_1 is very variable and depends on the cell type. Cells can enter G_0, aquiescent phase, instead of continuing with the cell cycle.

semi-conservatively. Replication of each linear DNA molecule in a chromosome starts at **many origins**, one every 30–300 kb of DNA depending on the species and tissue, and proceeds **bi-directionally** from each origin. The use of multiple origins is essential in order to ensure that the large amount of chromosomal DNA in a eukaryotic cell is replicated within the necessary time period. At each origin, a **replication bubble** forms consisting of two **replication forks** moving in opposite directions. The DNA replicated under the control of a single origin is called a **replicon**. DNA synthesis proceeds until replication bubbles merge together (*Fig. 2*).

All of the regions of a chromosome are not replicated simultaneously. Rather, many replication eyes will be found in one part of the chromosome and none in

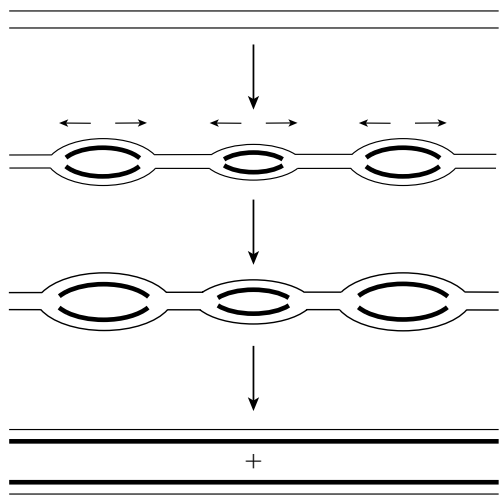


Fig. 2. Replication of eukaryotic chromosomal DNA. Replication begins at many origins and proceeds bi-directionally at each location. Eventually the replication eyes merge together to produce two daughter DNA molecules, each of which consists of one parental DNA strand (thin line) and one newly synthesized DNA strand (thick line).

another section. Thus replication origins are activated in clusters, called **replication units**, consisting of 20–80 origins. During S phase, the different replication units are activated in a set order until eventually the whole chromosome has been replicated. Transcriptionally-active DNA appears to be replicated early in S phase, whilst chromatin that is condensed and not transcriptionally active is replicated later.

Five DNA polymerases

Eukaryotic cells contain five different DNA polymerases; α, β, γ, δ and ε. The DNA polymerases involved in replication of chromosomal DNA are α and δ. DNA polymerases β and ε are involved in DNA repair. All of these polymerases except DNA polymerase γ are located in the nucleus; DNA polymerase γ is found in mitochondria and replicates mitochondrial DNA.

Leading and lagging strands

The basic scheme of replication of double-stranded chromosomal DNA in eukaryotes follows that for bacterial DNA replication (see Topic F3); a leading strand and a lagging strand are synthesized, the latter involving discontinuous synthesis via Okazaki fragments. The RNA primers required are made by DNA polymerase α which carries a **primase subunit**. DNA polymerase α initiates synthesis of the lagging strand, making first the RNA primer and then extending it with a short region of DNA. DNA polymerase δ then synthesizes the rest of the Okazaki fragment. The leading strand is synthesized by DNA polymerase δ. The δ enzyme has 3′ → 5′ exonuclease activity and so can proofread the DNA made, but DNA polymerase α has no such activity.

Telomere replication

The replication of a linear DNA molecule in a eukaryotic chromosome creates a problem that does not exist for the replication of bacterial circular DNA molecules. The normal mechanism of DNA synthesis (see above) means that the 3′ end of the lagging strand is not replicated. This creates a gap at the end of the chromosome and therefore a shortening of the double-stranded replicated portion. The effect is that the chromosomal DNA would become shorter and shorter after each replication. Various mechanisms have evolved to solve this problem. In many organisms the solution is to use an enzyme called **telomerase** to replicate the chromosome ends (**telomeres**).

Each telomere contains many copies of a repeated hexanucleotide sequence that is G-rich; in *Tetrahymena* it is GGGTTG. Telomerase carries, as an integral part of its structure, an RNA molecule, part of which is complementary to this G-rich sequence. The exact mechanism of action of telomerase is not clear; *Fig. 3* shows one possible model. The RNA molecule of telomerase is envisaged to hydrogen-bond to the telomere end. Then, using the RNA as a template, telomerase copies the RNA template (hence this enzyme is a reverse transcriptase; see Topic I4) and adds six deoxynucleotides to the telomere DNA end. Telomerase then dissociates from the DNA, re-binds at the new telomere end and repeats the extension process. It can do this hundreds of times before finally dissociating. The newly extended DNA strand can then act as a template for normal DNA replication (lagging strand synthesis by DNA polymerase α) to form double-stranded chromosomal DNA. The two processes, of the DNA ends shortening through normal replication and of lengthening using telomerase, are very roughly in balance so that each chromosome stays approximately the same length.

Replication of chromatin

When a chromosome is replicated, the replication machinery passes through the nucleosomes without removing the histones from the DNA. How this occurs is

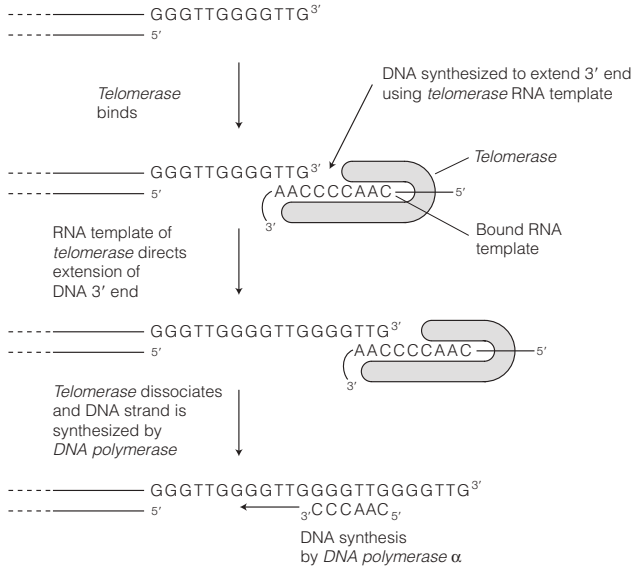


Fig. 3. Replication of telomeric DNA. Telomerase has a bound RNA molecule that is used as template to direct DNA synthesis and hence extension of the ends of chromosomal DNA.

still not fully understood. Both of the daughter DNA molecules that result from replication have old histones bound to them but overall, since the amount of DNA has now doubled, more histones are needed to package the DNA correctly into nucleosomes. Not surprisingly, therefore, the S phase of the cell cycle is also the time when large amounts of histones are synthesized.