PART ONE

Basic Principles of Human Genetics
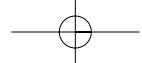
1

DNA Structure and Function

INTRODUCTION

The 20th century will likely be remembered by historians of biological science for the discovery of the structure of DNA and the mechanisms by which information coded in DNA is translated into the amino acid sequence of proteins. Although the story of modern human genetics begins about 50 years before the structure of DNA was elucidated, we will start our exploration here. We do so because everything we know about inheritance must now be viewed in the light of the underlying molecular mechanisms. We will see here how the structure of DNA sets the stage both for its replication and for its ability to direct the synthesis of proteins. We will also see that the function of the system is tightly regulated, and how variations in the structure of DNA can alter function. The story of human genetics did not begin with molecular biology, and it will not end there, as knowledge is now being integrated to explain the behavior of complex biological systems. Molecular biology, however, remains a key engine of progress in biological understanding, so it is fitting that we begin our journey here.

KEY POINTS

- DNA consists of a double helical sugar-phosphate structure with the two strands held together by hydrogen bonding between adenine and thymine or cytosine and guanine bases.
- DNA replication involves local unwinding of the double helix and copying a new strand from the base sequence of each parental strand. Replication proceeds bidirectionally from multiple start sites in the genome.
- DNA is complexed with proteins to form a highly compacted chromatin fiber in the nucleus.
- Genetic information is copied from DNA into messenger RNA in a highly regulated process that involves activation or repression of individual genes. mRNA molecules are extensively processed in the nucleus, including removal of introns and splicing together of exons, prior to export to the cytoplasm for translation into protein.
- The base sequence of mRNA is read in triplet codons to direct the assembly of amino acids into protein on ribosomes.
- Some genes are permanently repressed by methylation of some cytosine bases. These include most genes on one of two X chromosomes in cells of females and one of the two copies of genes that are said to be imprinted.

DEOXYRIBONUCLEIC ACID

Mendel described dominant and recessive inheritance before the concept of the "gene" was introduced, and long before the chemical basis of inheritance was known. Cell biologists during the late 19th and early 20th centuries had established the cell nucleus as the likely location of the genetic material, and DNA was long known to be a major chemical constituent. As the chemistry of DNA came to be understood, for a long time it was considered to be too simple a molecule – consisting of just four chemical building blocks, the bases **adenine**, **guanine**,
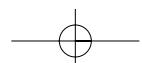
Methods 1.1

Mendelian inheritance in man

Dr. Victor McKusick and his colleagues at Johns Hopkins School of Medicine began to catalog genes and human genetic traits in the 1960s. The first edition of the catalog *Mendelian Inheritance in Man* was published in 1969. Multiple, subsequent print editions have appeared, and now the catalog is maintained on the world wide web by the National Center for Biotechnology Information (NCBI) as *Online Mendelian Inheritance in Man* (OMIM). The url is: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

OMIM is recognized as the authoritative source of information about human genes and genetic traits. The catalog can be searched by gene, phenotype, gene locus and many other features. The catalog provides a synopsis of the gene or trait, including a summary of clinical features associated with mutations. There are links to other databases, providing access to gene and amino acid sequences, mutations, etc. Each entry has a unique, six digit number, the MIM number. Autosomal dominant traits have entries beginning with 1, recessive traits with 2, X-linked with 3, and mitochondrial with 5. Specific genes have MIM numbers that start with 6.

Throughout this book, genes or genetic traits will be annotated with their corresponding MIM number to remind the reader that more information is available on OMIM and to facilitate access to the entry.

---

thymine, and **cytosine**, along with sugar and phosphate – to account for the complexity of genetic transmission. Credit for recognition of the role of DNA in inheritance goes to the landmark experiments by Avery *et al.*, who demonstrated that a phenotype of smooth or rough colonies of the bacterium *Pneumococcus* could be transmitted from cell to cell through DNA alone. Elucidation of the structure of DNA by Watson and Crick in 1953 opened the door to understanding the mechanisms whereby this molecule functions as the agent of inheritance (Methods 1.1).

DNA Structure

DNA consists of a pair of strands of a sugar-phosphate backbone attached to a set of **pyrimidine** and **purine** bases (Figure 1.1). The sugar is **deoxyribose** – ribose missing an oxygen atom at its 2′ position. Each DNA strand consists of alternating deoxyribose molecules connected by phosphodiester bonds from the 5′ position of one deoxyribose to the 3′ position of the next.

What is the structure of DNA?

Cytosine   Guanine
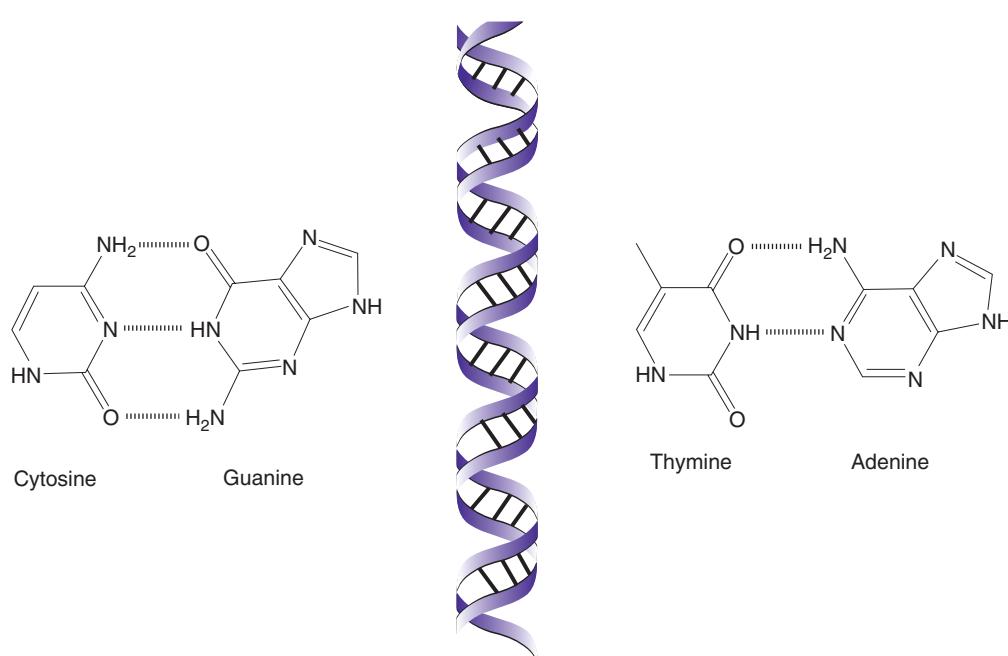
Thymine   Adenine

Figure 1.1 • Double helical structure of DNA (center). The sugar-phosphate helices are held together by hydrogen bonding between adenine and thymine bases, or guanine and cytosine bases.
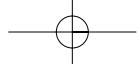
Methods 1.2

Isolation of DNA

DNA, or in some cases RNA, is the starting point for most experiments aimed at study of gene structure or function. DNA can be isolated from any cell that contains a nucleus. The most commonly used tissue for human DNA isolation is peripheral blood, where white blood cells provide a readily accessible source of nucleated cells. Other commonly used tissues include cultured skin fibroblasts, epithelial cells scraped from the inner lining of the cheek and fetal cells obtained by amniocentesis or chorionic villus biopsy. Peripheral blood lymphocytes can be transformed with Epstein–Barr virus into immortalized cell lines, providing permanent access to growing cells from an individual.

Nuclear DNA is complexed with proteins, which must be removed in order for the DNA to be analyzed. For some experiments it is necessary to obtain highly purified DNA, which involves digestion or removal of the proteins. In other cases, relatively crude preparations suffice. This is the case, for example, with DNA isolated from cheek scrapings. The small amount of DNA isolated from this source is usually released from cells with minimal effort to remove proteins. This preparation is adequate for limited analysis of specific gene sequences. Crude DNA preparations can be obtained from very minute biological specimens, such as drops of dried blood, skin cells, or hair samples isolated from crime scenes for forensic analysis.

Isolation of RNA involves purification of nucleic acid from the nucleus and/or cytoplasm. This RNA can be used to study the patterns of gene expression in a particular tissue. RNA tends to be less stable than DNA, requiring special care during isolation to avoid degradation.

The strands are bound together by hydrogen bonds between adenine and thymine bases and between guanine and cytosine bases. Together these strands form a double helix. The two strands run in opposite (antiparallel) directions, so that one extends 5′ to 3′ while the other goes 3′ to 5′.

The key feature of DNA, wherein resides its ability to encode information, is in the sequence of the four bases (Methods 1.2). The number of adenine bases (A) always equals the number of thymines (T), and the number of cytosines (C) always equals the number of guanines (G). This is because A on one strand is always paired with T on the other, and C on one strand is always paired with G. The pairing is noncovalent, due to hydrogen bonding between complementary bases. G–C base pairs form three hydrogen bonds, whereas A–T pairs form two, making the G–C pairs slightly more thermodynamically stable. Because the pairs always include one purine base (A or G) and one pyrimidine base (C or T), the distance across the helix remains constant.

DNA Replication

How are DNA molecules replicated?

The complementarity of A to T and G to C provides the basis for DNA replication, a point that was recognized by Watson and Crick in their paper describing the structure of DNA. DNA replication proceeds by a localized unwinding of the double helix, with each strand serving as a template for replication of a new sister strand (Figure 1.2). Wherever a G base is found on one strand a C will be placed on the growing strand; wherever a T is found an A will be placed, etc. Bases are positioned in the newly synthesized strand by hydrogen bonding, and new phosphodiester bonds are formed in the growing strand by action of the enzyme **DNA polymerase**. This is referred to as **semiconservative replication**, because the newly synthesized DNA double helices are hybrid molecules that consist of one parental strand and one new "daughter" strand. Unwinding of the double helix is accomplished by another enzyme system, called **helicase**.

DNA replication requires growth of a strand from a pre-existing "primer" sequence. The primer sequences are provided by a process of transcription, in which a short RNA molecule is synthesized from the DNA template. We will focus on transcription in the next section when we look at the means by which genetic information is used to synthesize protein. RNA is a single-stranded nucleic acid, similar to DNA, except that the sugar molecules are ribose rather than deoxyribose, and uracil substitutes for thymine (and pairs with adenine). These short RNA primers are extended by DNA polymerase (Figure 1.3). DNA is synthesized in a 5′ (exposed phosphate on 5′ carbon of the ribose molecule) to 3′ (exposed hydroxyl on the 3′ carbon) direction.

Structure of a replication fork

DNA ligase

DNA polymerase α

DNA primase subunit

Old RNA primer

New RNA primer

Okazaki fragment

Single strand DNA binding protein

Lagging strand template

5′

3′

DNA helicase

5′

Parental DNA

3′

Leading strand template

DNA polymerase δ
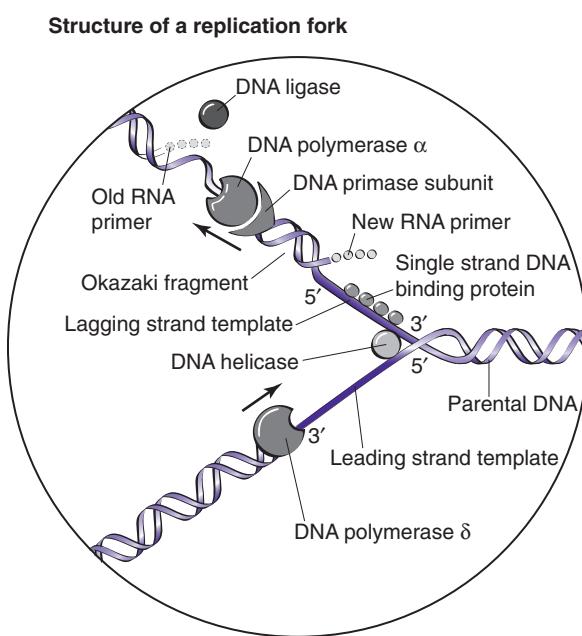
For one strand, referred to as the **leading strand**, this can be accomplished continuously as the DNA unwinds. The other strand, called the **lagging strand**, is replicated in short segments, called **Okazaki fragments**, which are then enzymatically ligated together by DNA ligase. Two distinct polymerases, δ (leading strand) and α (lagging strand) replicate the DNA. The short RNA primers are ultimately removed and replaced with DNA to complete the replication process.

The human genome consists of over 3 billion base pairs of DNA packaged onto 23 pairs of chromosomes. Each chromosome consists of a single, continuous DNA molecule, encompassing tens to hundreds of millions of base pairs. If the DNA on each chromosome were to be replicated in a linear manner from one end to another the process would go on interminably – certainly too long to sustain the rates of cell division that must occur. In fact, the entire genome can be replicated in a matter of hours because replication occurs simultaneously at multiple sites along a chromosome. These origins of replication are bubble-like structures from which DNA replication proceeds bidirectionally until an adjacent bubble is reached (Figure 1.4).

Replication of human DNA

Replication origins

5′   3′

3′   Parental DNA   5′

Replication bubbles   Replication forks

Okazaki fragments

5′   3′

3′   5′
5′   3′

3′   5′

Daughter strands of DNA
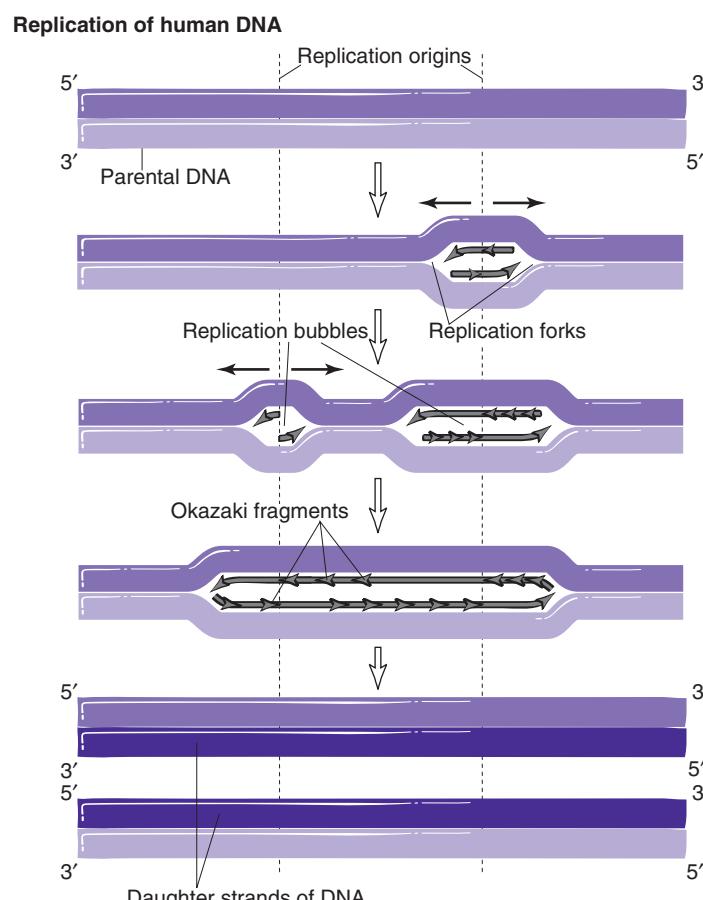
Figure 1.4 • DNA replication proceeds bidirectionally from multiple start sites. From Pritchard & Korf (2003) Medical Genetics at a Glance. Blackwell Publishing, Oxford.

Extend parental strand using
RNA template in telomerase

Parental Strand   G G G T T A G G G T T A G G G T T A – 3′

Lagging Strand   C C C A A T – 5′   AUCCCAAU

telomerase

Complete lagging strand with
DNA polymerase

Parental Strand   G G G T T A G G G T T A G G G T T A G G G T T A – 3′

Lagging Strand   C C C A A T – 5′   T C C C A A T C C C – 5′
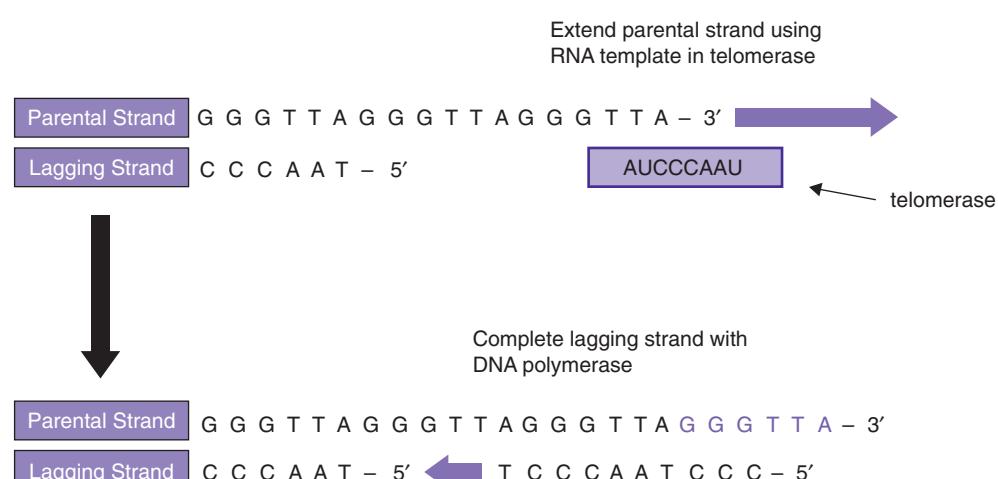
Figure 1.5 • The enzyme telomerase adds a terminal sequence to the ends of chromosomes.

One special case in DNA replication is the replication of the ends of chromosomes. Removal of the terminal RNA primer from the lagging strand at the end of a chromosome would result in shortening of the end, since there is no upstream primer for DNA polymerase to replace the short RNA primer. This problem is circumvented by action of an enzyme called **telomerase**, which uses an RNA template intrinsic to the enzyme to add a stretch of DNA onto the 3′ end of the lagging strand (Figure 1.5). The DNA sequence of the telomere is determined by the

CLINICAL SNAPSHOT 1.1

■ Dyskeratosis congenita

Eddy is a 4-year-old boy brought in by his parents because of recurrent cough. He has had two bouts of pneumonia, which were treated with antibiotics, over the past 2 months. Now he is sick again, having never stopped coughing since the last episode of pneumonia. He has also been noted by his parents to have lacked energy over the past several weeks. His examination shows a fever of 39°C and rapid respirations with frequent coughing. His breath sounds are abnormal on the right side of his chest. He also has hyperkeratotic skin with streaky hyperpigmentation. His finger and toe nails are thin and broken at the ends and his hair is sparse. A blood count shows anemia and a reduced number of white blood cells. A bone marrow aspirate is obtained, and it shows a generalized decrease in all cell lineages. A clinical diagnosis of dyskeratosis congenita is made.

Dyskeratosis congenita consists of reticulated hyperpigmentation of the skin, dystrophic hair and nails, and generalized bone marrow failure (Figure 1.6). It usually presents in childhood, often with signs of pancytopenia. There is an increased rate of spontaneous chromosome breakage seen in peripheral blood lymphocytes. Dyskeratosis congenita can be inherited as an X-linked recessive (MIM 305000), autosomal dominant (MIM 127550), or autosomal recessive (MIM 224230) trait. The X-linked form is due to mutation in a gene that encodes the protein dyskerin (MIM 300126). Dyskerin is involved in the synthesis of ribosomal RNA and also interacts with telomerase. The autosomal dominant form is due to mutation in the gene *hTERC* (MIM 602322). *hTERC* encodes the RNA component of telomerase. The gene encoding the autosomal recessive form is not yet known. The X-linked recessive form is more severe and earlier in onset than the dominant form. Both forms are associated with defective telomere functioning, leading to shortened telomeres. This likely leads to premature cell death and also explains the spontaneous chromosome breakage. The phenotype of the X-linked form may also be due, in part, to defective rRNA processing.
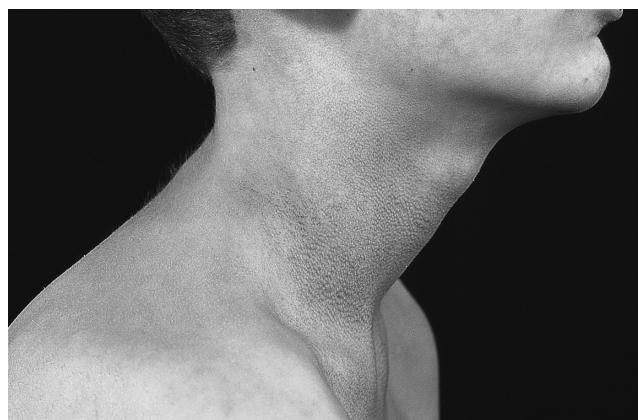
Figure 1.6 • Skin changes in an individual with dyskeratosis congenita. [Reproduced with permission from T. Burns, S. Breathnach, N. Cox, and C. Griffiths, eds, 7th edn, Rook's Textbook of Dermatology, Blackwell Publishing, Oxford.]

RNA sequence in the enzyme; for humans the sequence is GGGTTA. Each chromosome end has a tandem repeat of thousands of copies of the telomere sequence that is replicated during early development. Somatic cells may replicate without telomerase activity, resulting in a gradual shortening of the ends of the chromosomes with successive rounds of replication. This may be one of the factors that limits the number of times a cell can divide before it dies, a phenomenon known as **senescence** (Clinical Snapshot 1.1).

Chromatin

The DNA within each cell nucleus must be highly compacted to accommodate the entire genome in a very small space. The enormous stretch of DNA that composes each chromosome is actually a highly organized structure (Figure 1.7). The DNA double helix measures approx-

How is DNA packaged in the cell nucleus?
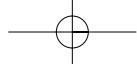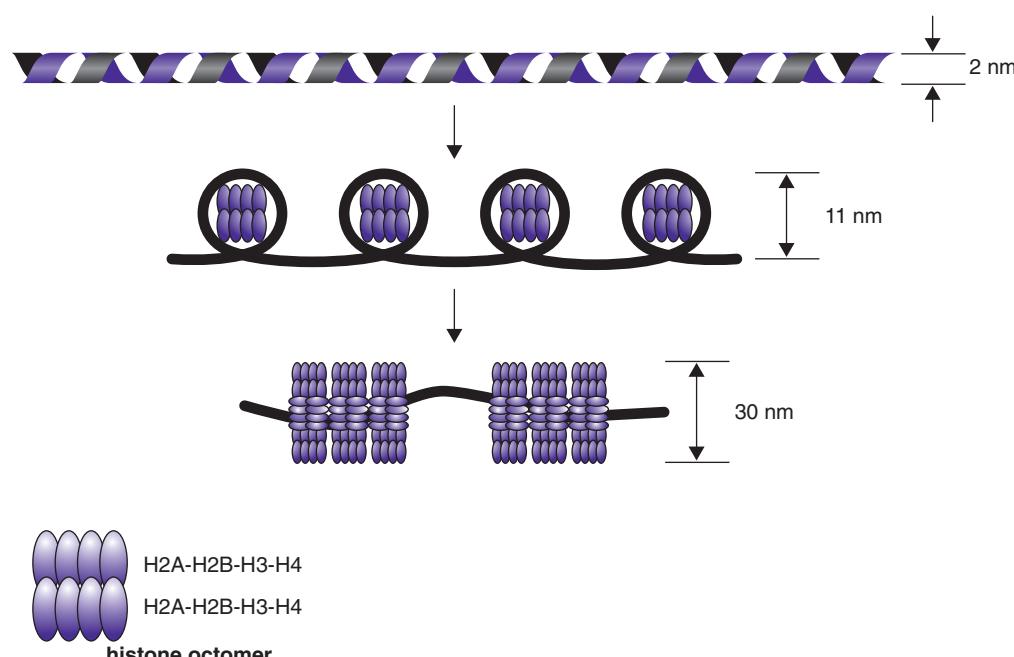
Figure 1.7 • Levels of organization of chromatin. The DNA double helix has a width of approximately 2 nm. The fundamental unit of chromatin is the nucleosome, which consists of 146 base pairs of DNA wound around a core consisting of two copies of each of the four histone proteins (H2A, H2B, H3, and H4). These are arrayed as beads on a string. The diameter of a nucleosome is 11 nm. The nucleosomes are, in turn, wound into a structure measuring 30 nm. This is further coiled and condensed to compose a metaphase chromosome.

imately 2 nm in diameter, but DNA does not exist in the nucleus in a "naked" form. It is complexed with a set of lysine- and arginine-rich proteins called histones. Two molecules of each of four major histone types – H2A, H2B, H3, and H4 – associate together with about every 146 base pairs to form a structure known as the **nucleosome**, which results in an 11-nm thick fiber. Nucleosomes are separated from one another by up to 80 base pairs, like beads on a string. This is more or less the conformation of actively transcribed chromatin but, during periods of inactivity, some regions of the genome are more highly compacted. The next level of organization is the coiling of nucleosomes into a 30-nm thick chromatin fiber held together by another histone, H1, and other nonhistone proteins. Chromatin is further compacted into the highly condensed structures comprising each chromosome, with the maximum condensation occurring during the metaphase stage of mitosis (see Chapter 6).

GENE FUNCTION

The basic tenet of molecular genetics – often referred to as "the central dogma"—is that DNA encodes RNA, which in turn encodes the amino acid sequence of proteins. It is now clear that this is a simplified view of the function of the genome. As will be seen in Chapter 4, much of the DNA sequence does not encode protein. A large proportion of the genome consists of noncoding sequences, such as repeated DNA, or encodes RNA that is not translated into protein. Nevertheless, the central dogma remains a critical principle of genome function. We will explore here the flow of information from DNA to RNA to protein.

Transcription

What is the role of RNA in conveying the sequence of DNA from the nucleus to the cytoplasm?

The process of copying the DNA sequence of a gene into **messenger RNA** (**mRNA**) is referred to as **transcription**. Some genes are expressed nearly ubiquitously. These are referred to as **housekeeping genes**. They include genes necessary for cell replication or metabolism. For other genes, expression is tightly controlled, with particular genes being turned on or off in particular cells at specific times in development or in response to physiological signals.

Gene expression is regulated by proteins that bind to DNA and either activate or repress
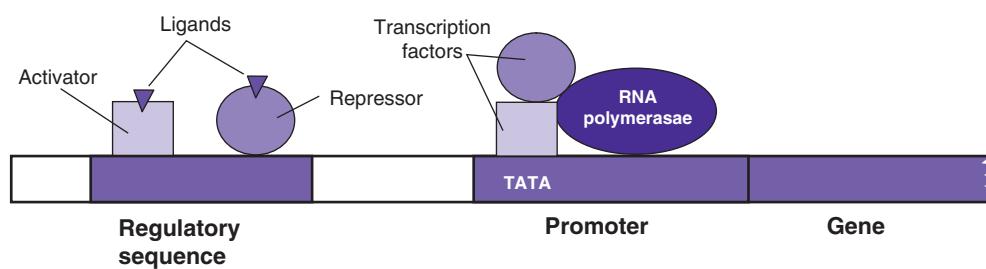
Figure 1.8 • Cis-acting elements regulating gene expression. Transcription starts at the promoter by binding of an RNA polymerase. Control of gene expression occurs via binding of transcription factors upstream of the transcription start site at the TATA box. Upstream regulatory sequences bind repressor or activator proteins, whose function is regulated by binding specific ligands.

transcription. The anatomy of elements that regulate gene transcription is shown in Figure 1.8. The **promoter region** is immediately adjacent to the transcription start site, usually within 100 base pairs. Most promoters include a base sequence of T and A bases called the TATA box. In some cases there may be multiple, alternative promoters at different sites in a gene that respond to different regulatory factors in different tissues. Regulatory sequences may occur adjacent to the promoter, or may be located thousands of base pairs away. These distantly located regulatory sequences are known as enhancers. Enhancer sequences function regardless of their orientation with respect to the gene.

DNA-binding proteins may serve as repressors or activators of transcription, and may bind to the promoter, to upstream regulatory regions, or to more distant enhancers. Activator or repressor proteins are regulated by binding of specific ligands. Ligand binding changes the confirmation of the transcription factor and may activate it or inactivate it. The ligand is typically a small molecule, such as a hormone. Many transcription factors work as a duet to form dimers. These may be homodimers of two identical proteins, or heterodimers of two different proteins. There may also be **corepressor** or **coactivator** proteins. Some transcription factors stay in the cytosol until the ligand or some other activation process occurs, at which time they move to the nucleus for activation of their target gene. In other situations, the transcription factors reside in the nucleus most of the time and may even be located at the response element sequences, but without the ligand they are inactive or even repress transcription.

Transcription begins with the attachment of the enzyme RNA polymerase to the promoter (Figure 1.9). There are three major types of RNA polymerase, designated types I, II, and III. Most gene transcription is accomplished by RNA polymerase II. Type I is involved in transcription of **rRNA** that resides in the ribosome and type III transcribes **transfer RNA** (**tRNA**) (see below). The polymerase reads the sequence of the DNA template strand, copying a complementary RNA molecule, which grows from the 5′ to the 3′ direction. The resulting mRNA is an exact copy of the DNA sequence, except that uracil takes the place of thymine in RNA. Soon after transcription begins, a 7-methyl guanine residue is attached to the 5′-most base, forming the "cap." Transcription proceeds through the entire coding sequence. Some genes include a sequence near the 3′ end that signals RNA cleavage at that site and enzymatic addition of 100 to 200 adenine bases, the "**poly-A tail**." Polyadenylation is characteristic of housekeeping genes, which are expressed in most cell types. Both the 5′ cap and the poly-A tail appear to function to stabilize the mRNA molecule and facilitate its export to the cytoplasm.

The DNA sequence of most genes far exceeds the length required to encode their corresponding proteins. This is accounted for by the fact that the coding sequence is broken up into segments, called **exons**, which are interrupted by segments called **introns**. Some exons may

Transcription by RNA polymerase II

Initiation of transcription

**Elongation of transcript**

Polyadenylation
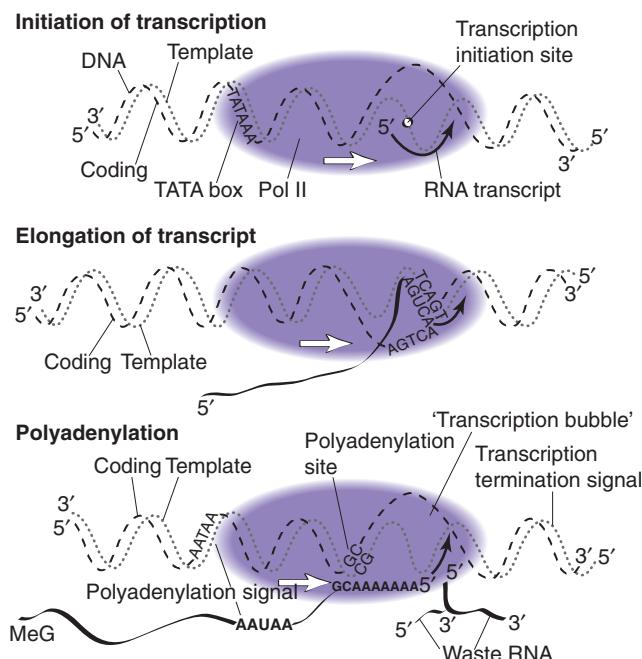
Figure 1.9 • Transcription involves copying an RNA from one stand of DNA. The reaction is catalyzed by RNA polymerase. A 7-methylguanosine cap (MeG) is added to the 5′ end of most mRNA molecules before transcription is completed. From Pritchard & Korf (2003) Medical Genetics at a Glance. Blackwell Publishing, Oxford.

Intron excision and exon splicing
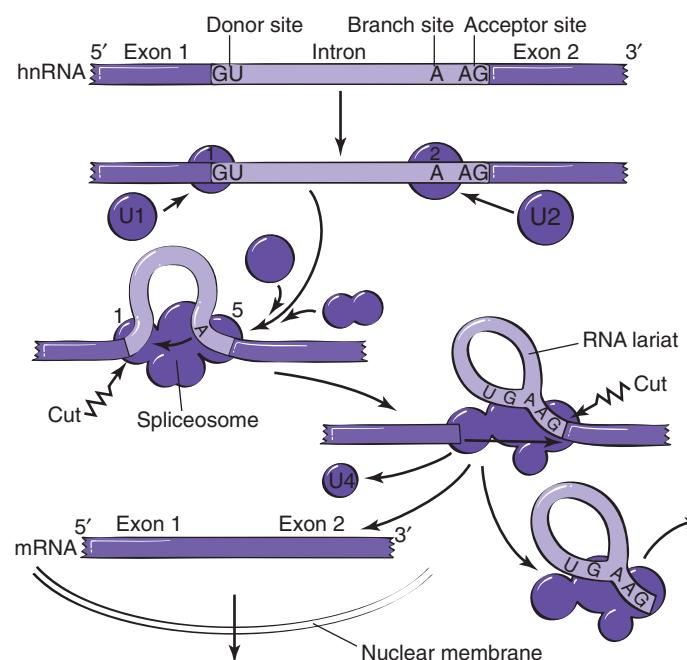
Figure 1.10 • RNA splicing begins with binding of specific ribonucleoproteins (U1 and U2) to the splice donor and acceptor. These two sites are then brought together by other components of the splicosome. The donor site is then cut and the free end of the intron binds to the branch point within the intron to form a lariat structure. Then the acceptor site is cleaved, releasing the lariat, and the exons at the two ends are ligated together. From Pritchard & Korf (2003) Medical Genetics at a Glance. Blackwell Publishing, Oxford.

be under a hundred bases long, whereas introns can be several thousand bases in length. Therefore, much of the length of a gene may be devoted to noncoding introns. The number of exons in a gene may be as few as one or two, or may number in the dozens. The processing of the RNA transcript into mature mRNA requires the removal of the introns and splicing together of the exons (Figure 1.10). This is carried out by an enzymatic process that occurs in the nucleus. The 5′ end of an intron always consists of the two bases GU, following by a con-
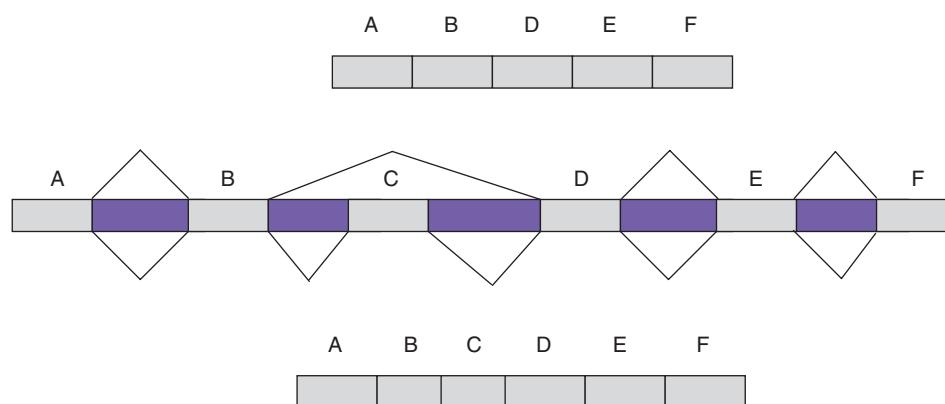
Figure 1.11 • Alternative splicing. Splicing out each intron results in inclusion of exons A–F in the mRNA. Alternatively, a splice can be made directly between exons B and D, skipping exon C. This results in production of a distinct protein, missing the amino acids encoded by exon C.

sensus sequence that is similar, but not identical, in all introns. This is the **splice donor**. The 3′ end, the **splice acceptor**, ends in AG, preceded by a consensus sequence.

The splicing process requires a complex machinery comprised of both proteins and small RNA molecules (**small nuclear RNA**, or **snRNA**), consisting of fewer than 200 bases. snRNA is also transcribed by RNA polymerase II. The splice is initiated by binding of a protein–RNA complex to the splice donor, at a point within the intron called the **branch point**, and the splice acceptor. First the DNA is cleaved at the donor site and this is attached in a 5′–2′ bond to the branch point. Then the acceptor site is cleaved, releasing a lariat structure that is subsequently degraded, and the 5′ and 3′ ends are ligated together. The splicing process also requires the function of proteins, **SR proteins**, which are involved in selecting sites for the initiation of splicing. These proteins interact with sequences known as **splice enhancers** or **silencers**. The splicing process is vulnerable to disruption by mutation, as might be predicted from its complexity.

The RNA splicing process offers a point of control of gene expression. Under the influence of control molecules present in specific cells, particular exons may be included or not included in the mRNA due to differential splicing (Figure 1.11). This results in the potential to produce multiple, different proteins from the same gene, adding greatly to the diversity of proteins encoded by the genome. Specific exons may correspond with particular functional domains of proteins, leading to the production of multiple proteins with diverse functions from the same gene. Some mRNAs are subject to RNA editing, in which a specific base may be enzymatically modified. For example, the protein apolipoprotein B exists in two forms, a 48-kDa form made in the intestine and a 100-kDa form in the liver. Both forms are the product of the same gene. In the intestine, however, the enzyme cytidine deaminase alters a C to a U at codon 2153, changing the codon from CAA (encoding glutamine) to UAA (a stop codon). This truncates the peptide, accounting for the 48-kDa form. Recently, another mechanism of post-transcritional regulation, called RNA interference, has also been identified (Hot Topics 1.1).

Translation

The mature mRNA is exported to the cytoplasm for translation into protein. During translation, the mRNA sequence is read into the amino acid sequence of a protein (Figure 1.13). The translational machinery consists of a protein–RNA complex called the **ribosome**. Ribosomes consist of a complex of proteins and specialized ribosomal RNA molecules (rRNA). The eukaryotic ribosome is comprised of two subunits, designated 60S and 40S (the "S" is a measurement of density, the Svendborg unit, reflecting how the complexes were initially characterized).

How is mRNA translated into protein?
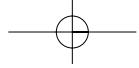
Hot Topic 1.1   R N A   I N T E R F E R E N C E

Gene regulation is not limited to control at the level of gene transcription. There is another level of control that occurs post-transcriptionally, referred to as RNA interference (RNAi). RNAi was discovered in plants, but appears to play a role in animals, including vertebrates. Its function in gene regulation is only beginning to come into focus.

The mechanisms of RNAi are illustrated in Figure 1.12. In experimental systems, and perhaps in some viral infections, RNAi begins with introduction of double-stranded RNA molecules (dsRNA), which are cleaved by the enzyme Dicer into short interfering RNA (siRNA) molecules. siRNA are 21 to 23 nucleotide, double-stranded RNAs with two unpaired bases at both ends. siRNAs separate into single strands and associate with specific proteins to form the RNA-induced silencing complex (RISC). The single-stranded siRNA binds to homologous sequences in mRNA and the RISC cleaves that RNA, thereby inactivating it.

The endogenous RNAi mechanism in animals begins with the transcription of genes that produce micro-RNAs (miRNAs), which are short RNA molecules containing segments with complementary bases that allow the molecule to form hairpin structures. The enzyme Drosha cleaves the hairpins, which are then exported to the cytoplasm, where Dicer further cleaves the hairpins into miRNA molecules that are similar to siRNAs. These associate with proteins and bind to homologous sequences, often in a region 3′ to the stop codon, referred to as the 3′ untranslated region (3′ UTR). Binding of the ribonucleoprotein complexes then inhibits translation, by as yet unknown mechanisms.

RNAi has been studied extensively in invertebrates, such as the flatworm *Caenorhabditis elegans* and the fruit fly *Drosophila*. In these organisms, interfering RNAs are involved in silencing genes during normal development. The role of RNAi in vertebrates, including humans, is just beginning to be explored, but here, too, it is likely to be involved in gene regulation. RNAi is also being exploited as an experimental tool and as an approach to therapy. dsRNA molecules can be introduced that are cleaved to produce siRNAs homologous to any gene of interest. This provides a means of selectively silencing genes in cells or tissues, allowing the function of these genes to be studied. As therapeutic tools, siRNA are being designed to silence viral genes, or to turn off genes that are activated by mutation, in genetic disorders or cancer.

Each subunit includes proteins and an rRNA molecule. The 60S subunit includes a 28S rRNA and the 40S subunit an 18S rRNA. Ribosomes can be free or associated with the **endoplasmic reticulum** (**ER**), also known as the "rough ER."

The mRNA sequence is read in triplets, called **codons**, beginning at the 5′ end of the mRNA, which is always AUG, encoding methionine (although this methionine residue is often later cleaved off). Each codon corresponds with a particular complementary **anticodon**, which is part of another RNA molecule called **transfer RNA** (**tRNA**). tRNA molecules bind specific amino acids defined by their anticodon sequence (Table 1.1). Protein translation therefore consists of binding of a specific tRNA to the appropriate codon, which juxtaposes the next amino acid in the growing peptide, which is enzymatically linked by an amide bond to the peptide. The process ends when a stop codon is reached (UAA, UGA, or UAG). The peptide is then released from the ribosome for transport to the appropriate site within the cell, or for secretion from the cell. A leader peptide sequence may direct the protein to its final destination in the cell; this peptide is cleaved off upon arrival. Post-translational modification, such as glycosylation, begins during the translation process and continues after translation is complete.

The process of translation consists of three phases, referred to as initiation, elongation, and termination. Initiation involves the binding of the first amino acyl tRNA, which always carries methionine, to the initiation codon, always AUG. A set of proteins, referred to as **elongation factors**, are involved in the process, which also requires ATP and GTP. The ribosome binds to the mRNA at two successive codons. One is designated the **P site** and carries the growing peptide chain. The other is the next codon, designated the **A site**. Elongation involves the binding of the next amino acyl tRNA to its anticodon at the A site. This delivers the next amino acid in the peptide chain, which is attached to the growing peptide, with peptide bond formation catalyzed by **peptidyl transferase**. The ribosome then moves on to the next codon under the action of a **translocase**, with energy provided by GTP. When a stop codon is reached, a release factor protein–GTP complex binds and the peptidyl transferase adds an OH to the end of the peptide, which is then released from the ribosome under the influence of proteins called **release factors**.

Hot Topic 1.1   RNA INTERFERENCE continued

Cytoplasm   Nucleus

Primary miRNA
transcript

Drosha

Experimental
system
viral RNA

miRNA precursor

Dicer

ATP   DCR-1

siRNA

RISC

miRNP

7mG   AAAA   7mG   AAAA

mRNA cleavage   Translational repression
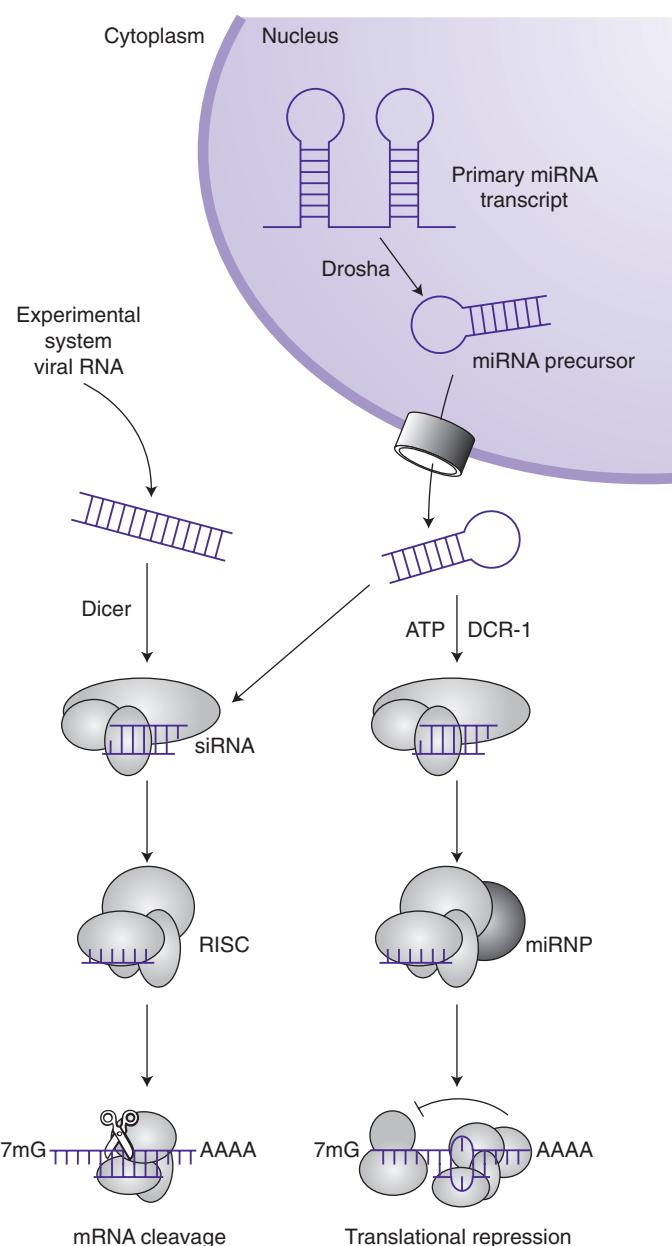
Figure 1.12 • Mechanisms of RNA interference. dsRNA may be introduced by viral infection or experimentally. siRNA is produced from dsRNA through cleavage by the enzyme Dicer. Single-stranded siRNA complexes with proteins to form the RISC, which then binds to mRNA through homologous pairing with the siRNA and cleaves the mRNA. Endogenous miRNA is transcribed and cleaved into hairpin structures in the nucleus. These are exported to the cytoplasm, where they are processed into miRNA by Dicer. The miRNA complexes with proteins and binds to mRNA, often in the 3′ UTR, and inhibits translation. (Adapted by permission from Macmillan Publishers Ltd: Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. Nature 2004;431:343–349.)
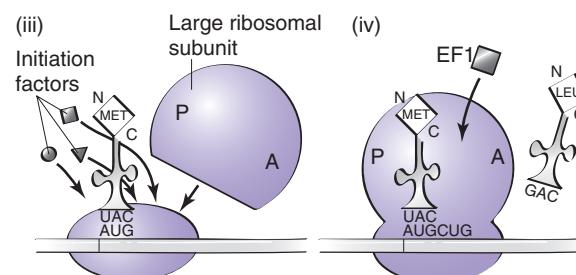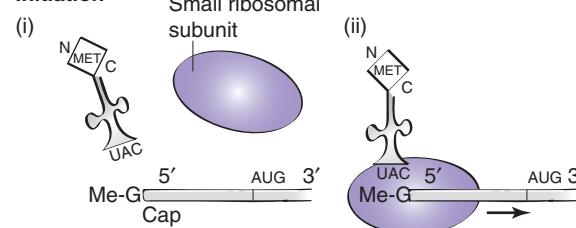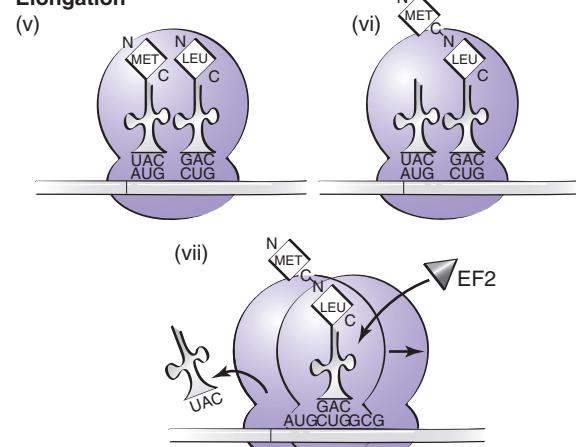
Translation
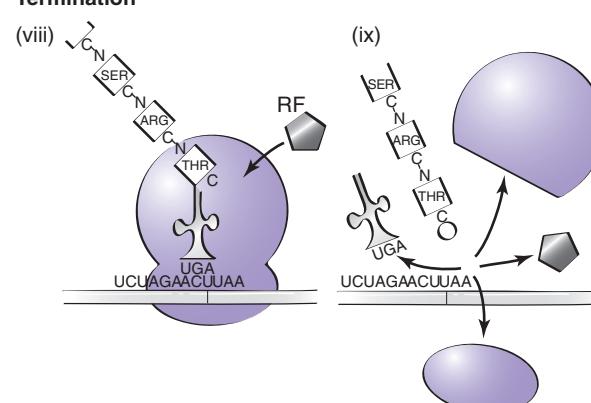
Figure 1.13 • The process of protein translation. Translation takes place at the ribosome, which binds to the mRNA. Specific amino acyl tRNA molecules bind to the mRNA by base pair complementarity between a triplet codon on the mRNA and an anticodon on the tRNA. A peptide bond is formed between the growing peptide and the next amino acyl tRNA, transferring the growing peptide and elongating it by one amino acid. This continues until a stop codon is reached. EF: elongation factor, RF: release factor. From Pritchard & Korf (2003) Medical Genetics at a Glance. Blackwell Publishing, Oxford.

How can genes be permanently inactivated?

Gene Inactivation and Imprinting

Individual genes may be reversibly activated or repressed, but there are some situations where genes or sets of genes are permanently silenced. This occurs on one of the two copies of the X chromosome in females, and on the maternal or paternal copy of a set of genes that are said to be **imprinted**. Gene silencing is accompanied by methylation of cytosine bases to

TABLE 1.1   The genetic code. A triplet codon is read from the left column, to the top row, to the full triplet in each box. Each codon corresponds with a specific amino acid, except for the three stop codons (labeled "Ter"). Most amino acids are encoded by more than one codon

|	T	C	A	G
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)
	TTC ″	TCC ″	TAC	TGC
	TTA Leu (L)	TCA ″	TAA **Ter**	TGA **Ter**
	TTG ″	TCG ″	TAG **Ter**	TGG Trp (W)
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)
	CTC ″	CCC ″	CAC ″	CGC ″
	CTA ″	CCA ″	CAA Gln (Q)	CGA ″
	CTG ″	CCG ″	CAG ″	CGG ″
A	ATT I1e (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)
	ATC ″	ACC ″	AAC ″	AGC ″
	ATA ″	ACA ″	AAA Lys (K)	AGA Arg (R)
	ATG Met (M)	ACG ″	AAG ″	AGG ″
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)
	GTC ″	GCC ″	GAC ″	GGC ″
	GTA ″	GCA ″	GAA Glu (E)	GGA ″
	GTG ″	GCG ″	GAG ″	GGG ″
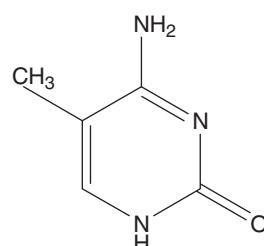
Figure 1.14 • Structure of 5-methylcytosine.

5-methylcytosine (Figure 1.14). This occurs in regions where cytosine is following by guanine (5′–CpG–3′) near the promoter, sites referred to as **CpG islands**. Methylated sites bind protein complexes that remove acetyl groups from histones, leading to transcriptional repression. The silencing is continued from cell generation to generation because the enzymes responsible for methylation recognize the 5-methylcytosine on the parental strand of DNA and methylate the cytosine on the newly synthesized daughter strand (Figure 1.15).

X-chromosome inactivation provides a mechanism for equalization of gene dosage on the X chromosome in males, who have one X, and females, who have two. Most genes on one of the two X chromosomes in each cell of a female are permanently inactivated early in development (Figure 1.16). The particular X inactivated in any cell is determined at random, so that in approximately 50% of cells one X is inactivated and in the other 50% the other X is inactivated. Regions of homology between the X and Y at the two ends of the X escape inactivation. These are referred to as **pseudoautosomal regions**. The inactive X remains condensed through most of the cell cycle, and can be visualized as a densely staining body during interphase, referred to as the **Barr body**.

Initiation of inactivation is controlled from a region called the **X-inactivation center** (**Xic**). A gene within this region, known as *Xist*, is expressed on one of the two X chromosomes early in development. *Xist* encodes a 25-kb RNA that is not translated into protein, but appears to bind to sites along the X to be inactivated. Subsequently, CpG islands on this chromosome are methylated, and changes occur to histones, particularly deacetylation.
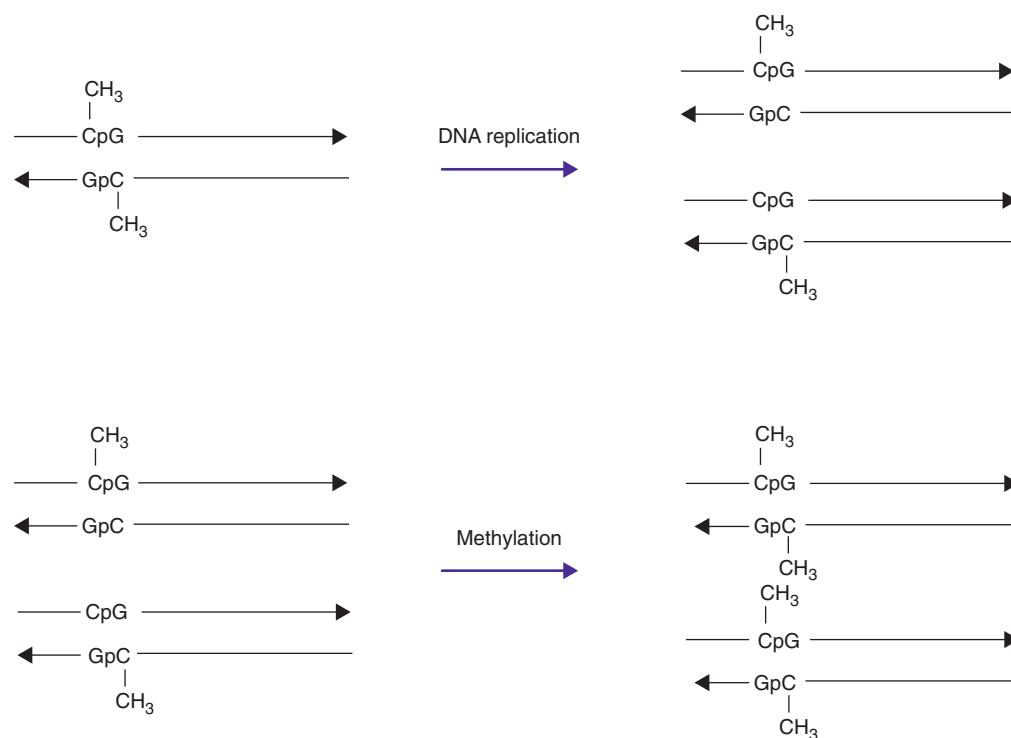
Figure 1.15 • Cytosine residues adjacent to guanines may be methylated near the 5′ ends of some genes. When the DNA is replicated, only one strand will be methylated, but then an enzyme recognizes the single methylated strand and methylates the cytosines on the opposite strand.
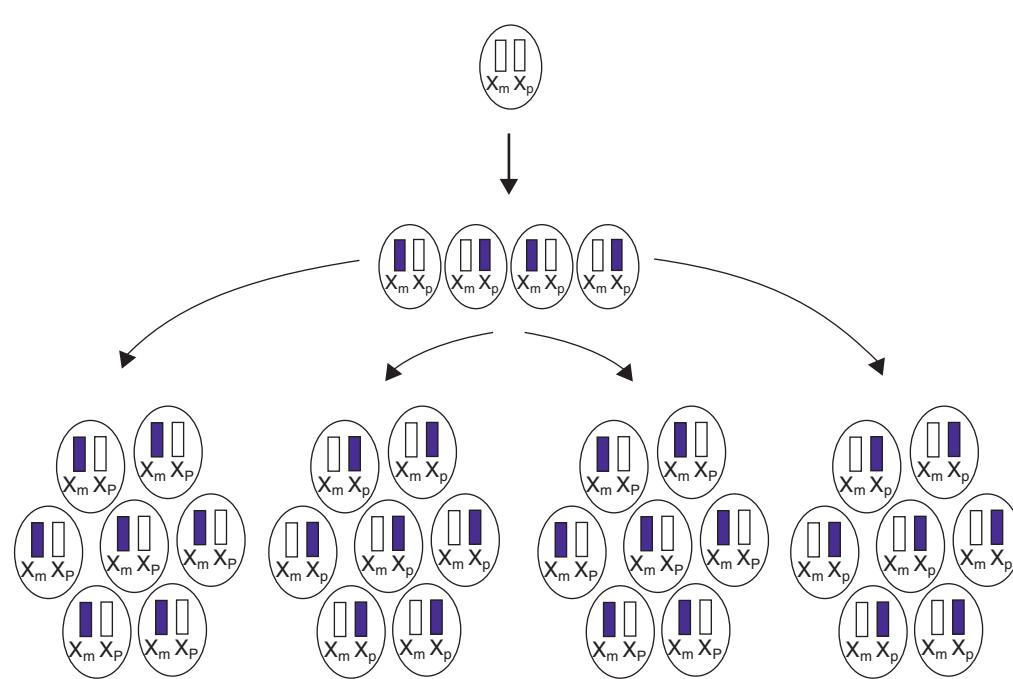

Figure 1.16 • X chromosome inactivation. In the zygote, both the maternally and paternally derived X chromosomes (X_m and X_p) are active. Early in development, one of the two X chromosomes in each cell is inactivated (indicated as the dark chromosome). This X chromosome remains inactive in all the descendants of that cell.

Genomic imprinting involves the silencing of either the maternal or paternal copy of a gene during early development (Figure 1.17). Like X-chromosome inactivation, imprinting is probably accomplished through methylation of specific chromosome regions. The methylation "imprint" is erased in germ cells, so the specific gene copy to be inactivated is always determined by the parent of origin, regardless of whether that particular gene copy was active or inactive in the previous generation. Genomic imprinting appears to apply only to a small subset of genes, although the full extent of imprinting is not yet known.
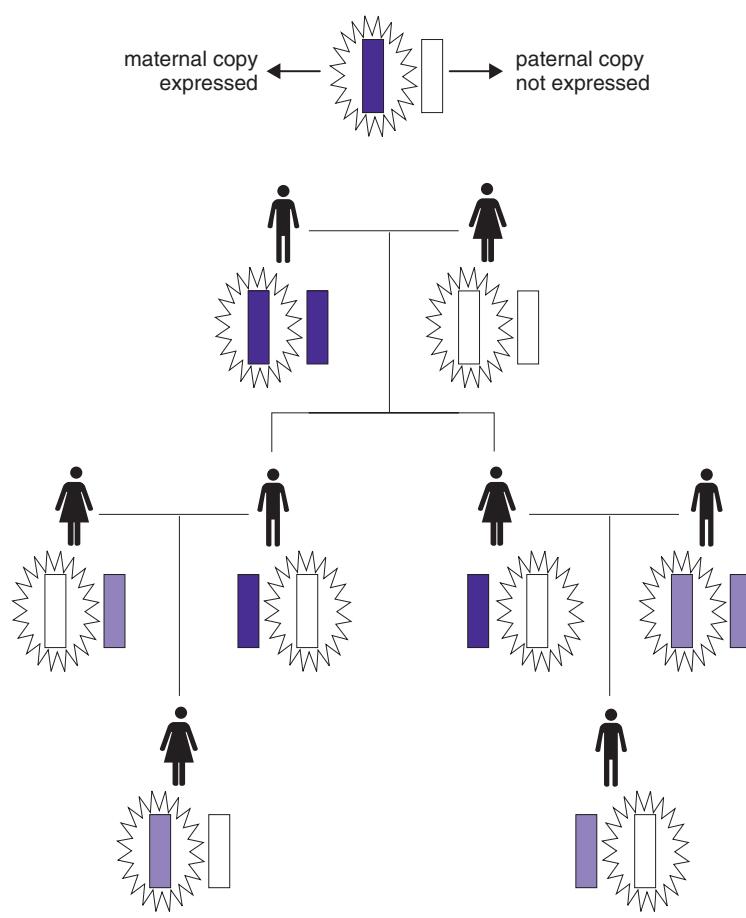
maternal copy
expressed

paternal copy
not expressed

CONCLUSION

More than half a century of research in molecular biology has resulted in a detailed picture of the mechanisms of gene structure and function. Much of the remainder of this book will be devoted to exploration of the implications of dysfunction at the level of the gene or groups of genes and their interactions with the environment. We will see also that genetics research is moving to a new level of integration of basic molecular mechanisms, towards formation of a picture of how entire cells and organisms function. It is important to realize, however, that some fundamental molecular mechanisms, such as the role of small RNAs and genomic imprinting, have been discovered only within the past decade or so. Even as the effort towards larger-scale integration goes forward, there remains much to be learned about the fundamental molecular mechanisms at the level of the gene.

REVIEW QUESTIONS

1.1 The two strands of DNA separate when heated, and the temperature at which separation occurs is dependent on base content. Specifically, DNA with a higher proportion of G–C base pairs tends to "melt" at a higher temperature than molecules with a higher A–T content. Why is this?

1.2 What is the role of transcription in DNA replication?

1.3 Consider the gene sequence below. What is the base sequence of the mRNA that would be transcribed from this gene, and what is the amino acid sequence of the peptide that would be translated?

5′ – promoter – ATG GTT GAT AGT CGT TGC CGC GGG CTG TGA – 3′

3′ – promoter – TAC CAA CTA TCA GCA ACG GCG CCC GAC ACT – 5′
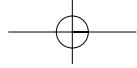
1.4 There are more proteins than there are genes. What are some of the mechanisms that account for this discrepancy?

1.5 A woman is heterozygous for an X-linked trait that leads to expression of two different forms of an enzyme. The two forms are separable as two distinct bands when the enzyme protein is run through an electric field by electrophoresis. If you were to test cultured skin fibroblasts and isolate enzyme, what would you expect to see? If you could isolate single fibroblasts and grow them into colonies before extracting the enzyme and subjecting it to electrophoresis what would you expect to see?

FURTHER READING

General References
Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell, 4th edn., 2002, New York: Garland Science.

Lewin B. Genes VIII, 2003, Upper Saddle River, NJ: Prentice Hall.

Lodish H, Berk A, Zipursky L, Matsudaira P, Baltimore D, Darnell J. Molecular Cell Biology, 5th edn., 2004, New York: Freeman.

Chromatin Structure
Dehghani H, Dellaire G, Bazett-Jones, DP. Organization of chromatin in the interphase mammalian cell. Micron 2005;36:95–108.

X Chromosome Inactivation
Latham KE. X chromosome imprinting and inactivation in preimplantation mammalian embryos. Trends Genet 2005;21:120–127.

Imprinting
Miyoshi N, Barton SC, Kaneda M, Hajkova P, Surani MA. The continuing quest to comprehend genomic imprinting. Cytogenet Genome Res 2006;113:6–11.

Paulsen M, Ferguson-Smith AC. DNA methylation in genomic imprinting, development, and disease. J Pathol 2005;195:97–110.

Clinical Snapshot 1.1 Dyskeratosis Congenita
Bessler M, Wilson DB, Mason PJ. Dyskeratosis congenita and telomerase. Curr Opin Pediatr 2004;16:23–28.

Methods 1.1 Mendelian Inheritance in Man
Online Mendelian Inheritance in Man: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

Hot Topic 1.1 RNA Interference
Hannon GJ, Rossi JJ. Unlocking the potential of the human genome with RNA interference. Nature 2004;431:371–378.