医学整合课程系列教材 · 原版影印

Integrated Genetics

整合遗传学

（第2版）

LINDA R. ADKISON

北京大学医学出版社

医学整合课程系列教材

整合遗传学

Integrated Genetics

（第2版）

Linda R. Adkison, PhD

Professor of Genetics
Associate Dean for Curricular Affairs
Kansas City University of Medicine and Biosciences
Kansas City, Missouri

北京大学医学出版社
Peking University Medical Press

图书在版编目（CIP）数据

整合遗传学 = Elsevier's Integrated Review Genetics : 第2版 : 英文 / （美）阿基森（Adkison, L. R.）编. -- 北京 : 北京大学医学出版社, 2014.8
ISBN 978-7-5659-0855-2

Ⅰ. ①整… Ⅱ. ①阿… Ⅲ. ①遗传学－医学院校－教材－英文 Ⅳ. ①Q3

中国版本图书馆CIP数据核字(2014)第090708号

北京市版权局著作权合同登记号：图字：01－2014－2489

Elsevier (Singapore) Pte Ltd.
3 Killiney Road, #08-01 Winsland House I, Singapore 239519
Tel: (65) 6349-0200 Fax: (65) 6733-1817
First Published 2014
2014年初版

整合遗传学（第2版）

编　　写：Linda R. Adkison
出版发行：北京大学医学出版社
电　　话：发行部：010-82802230　图书邮购：010-82802495
地　　址：（100191）北京市海淀区学院路38号　北京大学医学部院内
网　　址：http://www.pumpress.com.cn
E-mail：booksale@bjmu.edu.cn
印　　刷：北京圣彩虹制版印刷技术有限公司
经　　销：新华书店
责任编辑：冯智勇　　**责任印制**：张京生
开　　本：889mm×1194mm　1/16　　**印张**：16.5　　**字数**：511千字
版　　次：2014年8月第1版　2014年8月第1次印刷
书　　号：ISBN 978-7-5659-0855-2
定　　价：105.00元

出版说明

知识整合是当前医学教育改革的一项重要内容。目前国内基础医学各门课程的教材基本上是以学科为单位单独编写的，缺乏学科之间知识的联系。为了推动医学教育改革，借鉴国外医学教材的编写模式，北京大学医学出版社经过充分调研，引进出版了世界著名医学出版集团Elsevier公司的“Integrated”系列教材。

在编写上，该系列书最大的特色就是在保持本学科知识体系完整的同时插入大量的“整合框”。这些“整合框”出现在需要链接到其他学科相关知识的位置，每个学科都有独特的标识。例如在《病理学》的细胞损伤一节，讲述缺氧时，会插入一个“生物化学整合框”，介绍生物化学中糖酵解的知识；在感染一节，出现NK细胞的时候，会插入一个“免疫学整合框”，介绍免疫学中NK细胞的知识；在凝血一节，则是插入一个“临床医学整合框”，介绍临床上凝血的实验室评估方面的知识……这些分布在各本书中的“整合框”，把各学科之间知识点连接起来，不但方便了读者学习，更是体现了学科整合的理念。

该系列书包括：

●整合生理学 ●整合病理学

●整合药理学 ●整合生物化学

●整合遗传学 ●整合免疫学与微生物学

该系列书可作为国内医学生整合课程教材、双语教学教材及来华留学生教材，也有利于医学教师拓展知识，方便备课；同时也是美国医师执照考试的优秀参考用书。

北京大学医学出版社

Series Preface

How to Use This Book

The idea for Elsevier's Integrated Series came about at a seminar on the USMLE Step 1 exam at an American Medical Student Association (AMSA) meeting. We noticed that the discussion between faculty and students focused on how the exams were becoming increasingly integrated—with case scenarios and questions often combining two or three science disciplines. The students were clearly concerned about how they could best integrate their basic science knowledge.

One faculty member gave some interesting advice: "read through your textbook in, say, biochemistry, and every time you come across a section that mentions a concept or piece of information relating to another basic science—for example, immunology—highlight that section in the book. Then go to your immunology textbook and look up this information, and make sure you have a good understanding of it. When you have, go back to your biochemistry textbook and carry on reading."

This was a great suggestion—if only students had the time, and all of the books necessary at hand, to do it! At Elsevier we thought long and hard about a way of simplifying this process, and eventually the idea for Elsevier's Integrated Series was born.

The series centers on the concept of the *integration box*. These boxes occur throughout the text whenever a link to another basic science is relevant. They're easy to spot in the text—with their color-coded headings and logos. Each box contains a title for the integration topic and then a brief summary of the topic. The information is complete in itself—you probably won't have to go to any other sources—and you have the basic knowledge to use as a foundation if you want to expand your knowledge of the topic.

You can use this book in two ways. First, as a review book …

When you are using the book for review, the integration boxes will jog your memory on topics you have already covered. You'll be able to reassure yourself that you can identify the link, and you can quickly compare your knowledge of the topic with the summary in the box. The integration boxes might highlight gaps in your knowledge, and then you can use them to determine what topics you need to cover in more detail.

Second, the book can be used as a short text to have at hand while you are taking your course …

You may come across an integration box that deals with a topic you haven't covered yet, and this will ensure that you're one step ahead in identifying the links to other subjects (especially useful if you're working on a PBL exercise). On a simpler level, the links in the boxes to other sciences and to clinical medicine will help you see clearly the relevance of the basic science topic you are studying. You may already be confident in the subject matter of many of the integration boxes, so they will serve as helpful reminders.

At the back of the book we have included case study questions relating to each chapter so that you can test yourself as you work your way through the book.

Online Version

An online version of the book is available on our Student Consult site. Use of this site is free to anyone who has bought the printed book. Please see the inside front cover for full details on the Student Consult and how to access the electronic version of this book.

In addition to containing USMLE test questions, fully searchable text, and an image bank, the Student Consult site offers additional integration links, both to the other books in Elsevier's Integrated Series and to other key Elsevier textbooks.

Books in Elsevier's Integrated Series

The nine books in the series cover all of the basic sciences. The more books you buy in the series, the more links that are made accessible across the series, both in print and online.


Anatomy and Embryology


Histology


Neuroscience


Biochemistry


Physiology


Pathology


Immunology and Microbiology


Pharmacology


Genetics

Preface

Though the youngest of all the medical specialties, genetics embodies the essence of all normal and abnormal development and all normal and disease states. Perhaps because of its recent recognition as a discipline and perhaps because of its derivation from research in several areas, it is easier for genetics to be an "integrated" discipline. Approaching genetics as "a particular gene located on a specific chromosome and inherited in a specific manner" loses the appreciation of spatial and temporal dimensions of expression and the many, many factors affecting every single aspect of development, survival, and even death.

Every medical discipline is connected to human well-being through the mechanisms of gene expression, environmental influences, and inheritance. Genetics underscores the many biochemical pathways, physiologic processes, and pathologic mechanisms presented in other volumes of this series. It explains better the morphologic variation observed in embryologic development and anatomic presentation. It provides better insight into susceptibility to infection and disease. It offers insight into neurologic and behavioral abnormalities. It is defining the strategies for gene therapy and pharmacogenomics. For these reasons, it has been exciting to put this book together.

This text focuses on well-known and better described diseases and disorders that students and practitioners are likely to read about in other references. Many of these do not occur at a high frequency in populations, but they underscore major mechanisms and major concepts associated with many other medical situations. It is my hope that this text will be as stimulating to read as it was to write.

Linda R. Adkison, PhD

Editorial Review Board

I have been fortunate to have excellent mentors during my career in academics. I have learned a great deal about my own learning through this journey. My goals as a teacher are to help students become challenged by the fascination of learning, visualize what they cannot necessarily see, and describe what they see with integration of thought broadly across disciplines. This textbook is dedicated to the many wonderful students and colleagues who constantly challenge the boundaries of learning – theirs and mine. Finally, without the support and understanding of my family, especially my children, Emily and Seth, this project could not have been completed.

Linda R. Adkison, PhD

Contents

Case Studies and Case Study Answers are available online on Student Consult www.studentconsult.com

Basic Mechanisms 1

CONTENTS

The essence of genetics is an understanding of the hereditary material within a cell and the influence it has on survival of the cell through every function and response the cell and its organelles undertake. Without these fundamental concepts, no aspect of human development and well-being can be adequately explained.

CHROMATIN

One of the finest triumphs of modern science has been the elucidation of the chemical nature of chromatin and its role in the transfer of information from nucleic acids into proteins, known as the central dogma. James Watson built on his earlier work, which outlined the fundamental unit and chemical composition of the complex molecule composing chromatin deoxyribonucleic acid (DNA). Briefly stated, the central dogma "oversimplifies" the mechanism whereby the chemical message held in DNA is transferred to ribonucleic acid (RNA) through transcription and this RNA blueprint is translated into protein: DNA → RNA → protein. Other proteins associated with DNA contribute to its structure and many play roles in regulating functions. In its simplest form, chromatin is composed of DNA and histone proteins.

Histones are small, highly conserved, positively charged proteins that bind to DNA and to other histones. The five major histones are H1, H2A, H2B, H3, and H4. The presence of 20% to 30% lysine and arginine accounts for the positive charge of histones and distinguishes these from most other proteins. All histones except H1 are highly conserved among eukaryotes.

DNA is packaged into the nucleus by winding the double helix twice around an octamer of histones; this DNA-histone structure is called a nucleosome (Fig. 1-1). Each nucleosome is composed of two of each histone except H1 and approximately 150 nucleotide pairs wrapped around the histone core. H1 histone anchors the DNA around the core. This structure leads to a superhelix of turns upon turns upon turns called a solenoid structure. In the solenoid structure, each helical turn contains 6 nucleosomes and approximately 1200 nucleotide pairs. Additional turns form minibands that, when tightly stacked upon each other, give the structure recognized as a chromosome. In each nucleus, chromatin is organized into 46 chromosomes. In a fully relaxed configuration, DNA is approximately 2 nm in diameter; chromatids are approximately 840 nm in diameter. Twisting and knotting are extremely effective at compacting DNA within the nucleus (Fig. 1-2).

A DNA molecule comprises two long chains of nucleotides arranged in the form of a double helix. Its shape may be compared to a twisted ladder in which the two parallel supports of the ladder are made up of alternating deoxyribose sugars and phosphate molecules. Each rung of the ladder is composed of one pair of nitrogenous bases, held together by specific hydrogen bonds. Hydrogen bonds are weak bonds; however, the total number of hydrogen bonds between the strands assures that the strands of the double helix are firmly associated with each other under conditions commonly found in living cells.

BIOCHEMISTRY

DNA Configuration

There are three basic three-dimensional configurations of DNA. The most common is the B form in which DNA is wound in a right-handed direction with 10 bp per turn. Within the turned structure are a major groove and a minor groove, where proteins can bind. The A form also has a right-handed turn and is composed of 11 bp per turn. This form is seen in dehydrated DNA such as in oligonucleotide fibers or crystals. The third form, Z-form DNA, was named for its zigzag appearance and has a left-handed turn composed of 12 bp per turn. This form occurs in regions of DNA with alternating pyrimidines-purines: CGCGCG.

The molar concentration of adenine equals thymine and that of guanine equals cytosine. This information is best accommodated in a stable structure if the double-ring purines (adenine or guanine) lay opposite the smaller, single-ring pyrimidines (thymine or cytosine). The combination of one purine and one pyrimidine to make up each cross-connection

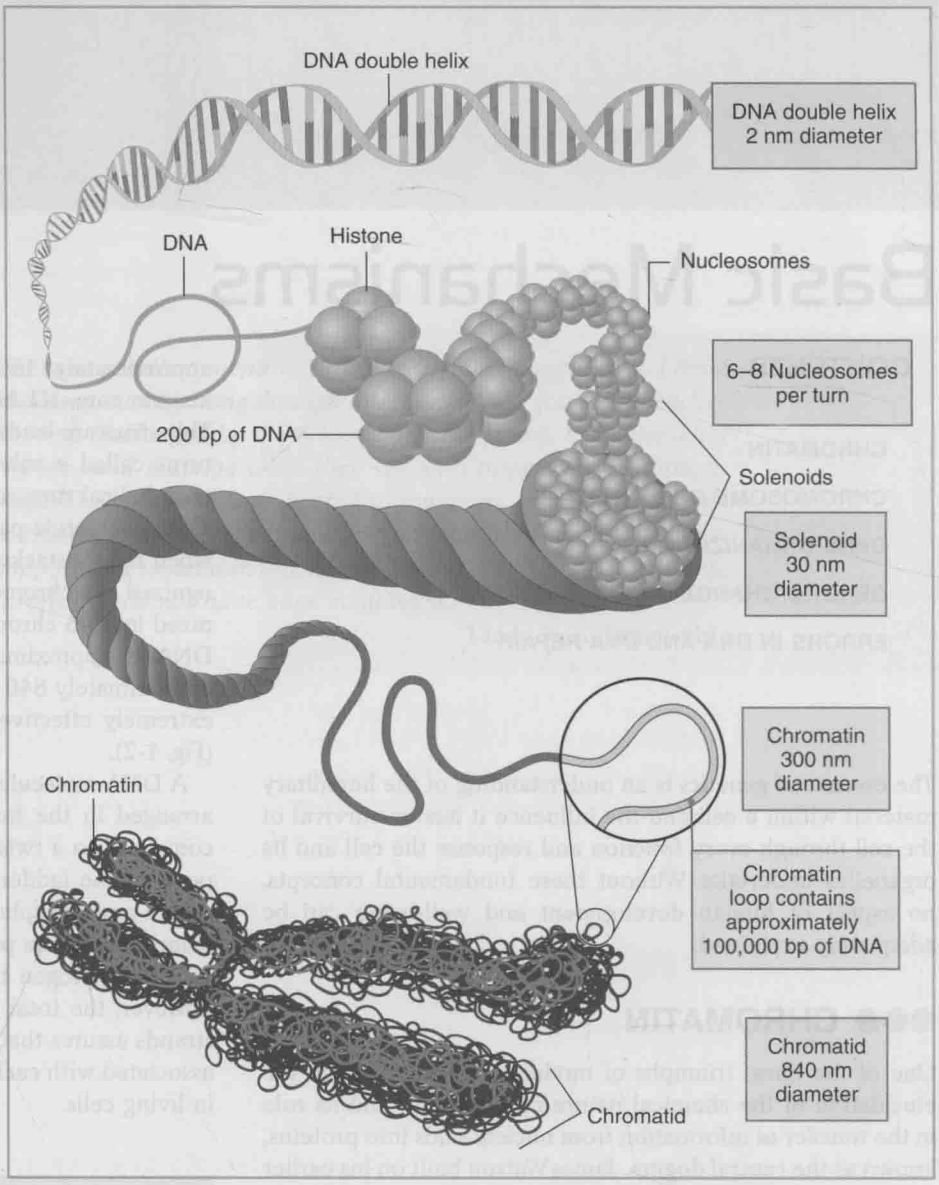

Figure 1-1. Chromosome organization. The shortening, or condensation, of chromatin results in a diminished volume of each chromosome and a reduction in the exposed chromosome surface. This is a dynamic process beginning with the least condensed form, the DNA double helix, and proceeding to chromatin visible in interphase and prophase. The level of greatest condensation occurs at metaphase.

is conveniently called a *base pair* (bp). In a DNA base pair, adenine (A) forms two hydrogen bonds with thymine (T), and guanine (G) and cytosine (C) share three hydrogen bonds. The sequence of one strand of DNA automatically implies the sequence of the opposite strand because of the precise pairing rule A = T and C = G.

BIOCHEMISTRY

Nitrogenous Bases

Purines are adenosine (A) and guanine (G). Pyrimidines are cytosine (C) and thymine (T). In the double helix structure, A binds to T with two hydrogen bonds; C binds to G with three hydrogen bonds.

Uracil (U) is found in RNA in place of T in DNA. The structure of U is T without the methyl group at carbon 5. Hypoxanthine is found in certain tRNAs.

Because of the configuration of phosphodiester bonds between the 3′ and 5′ positions of adjacent deoxyribose molecules, every linear polynucleotide can have a free, unbounded 3′ hydroxyl group at one pole of the polynucleotide (3′ end) and a free 5′ hydroxyl at the other pole (5′ end). There are theoretically two possible ways for the two polynucleotides to be oriented in a double helix. They could have the same polarity—that is, be parallel, with both strands having 3′ ends at one pole and 5′ ends at the other pole. Or, by rotating one strand 180 degrees with respect to the other, they could have opposite polarity—that is, be antiparallel—with a 3′ and a 5′ end at one pole of the double helix and a 5′ and a 3′ end at the other pole of the double helix. Only the antiparallel orientation actually occurs. The antiparallel nature of the double helix dictates that a new DNA chain being replicated must be copied in the opposite direction from the template (Fig. 1-3).

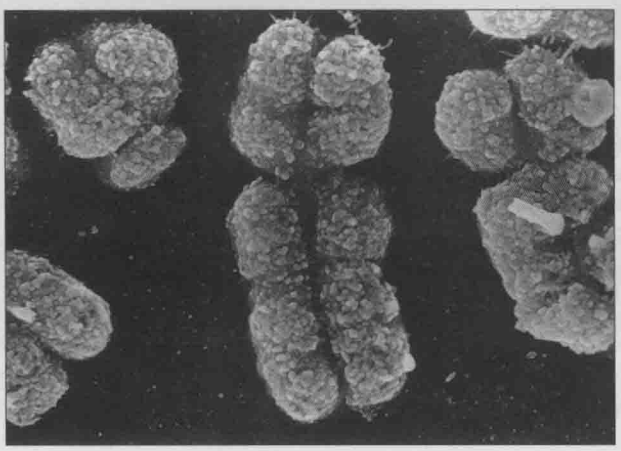

Figure 1-2. Generally, chromosomes are shown as in this photograph—in a highly condensed stage known as metaphase. This structure, however, represents one chromosome that has been replicated and is composed of two identical sister chromatids. At a later stage, the sister chromatids will separate at the centromere, and two chromosomes will exist. Note: when in doubt about the number of chromosomes present, count the number of centromeres present!

CHROMOSOME ORGANIZATION

DNA of eukaryotes is repetitive—that is, there are many DNA sequences of various lengths and compositions that do not represent functional genes. Three subdivisions of DNA are recognized: unique DNA, middle repetitive DNA, and highly repetitive DNA. Unique DNA is present as a single copy or as only a few copies. The proportion of the genome taken up by repetitive sequences varies widely among taxa. In mammals, up to 60% of the DNA is repetitive. The highly repetitive fraction is made up of short sequences, from a few to hundreds of nucleotides long, which are repeated on the average of 500,000 times. The middle repetitive fraction consists of hundreds or thousands of base pairs on the average, which appear in the genome up to hundreds of times.

BIOCHEMISTRY

Phosphodiester Bonds and Deoxyribose Molecules

A phosphodiester bond occurs between carbon 5 on one deoxyribose and carbon 3 on an adjacent deoxyribose. The sugar in DNA is deoxyribose: H^+ replaces OH^- at carbon 2. The energy to form this bond is derived from the cleavage of two phosphates from the ribonucleotide triphosphate.

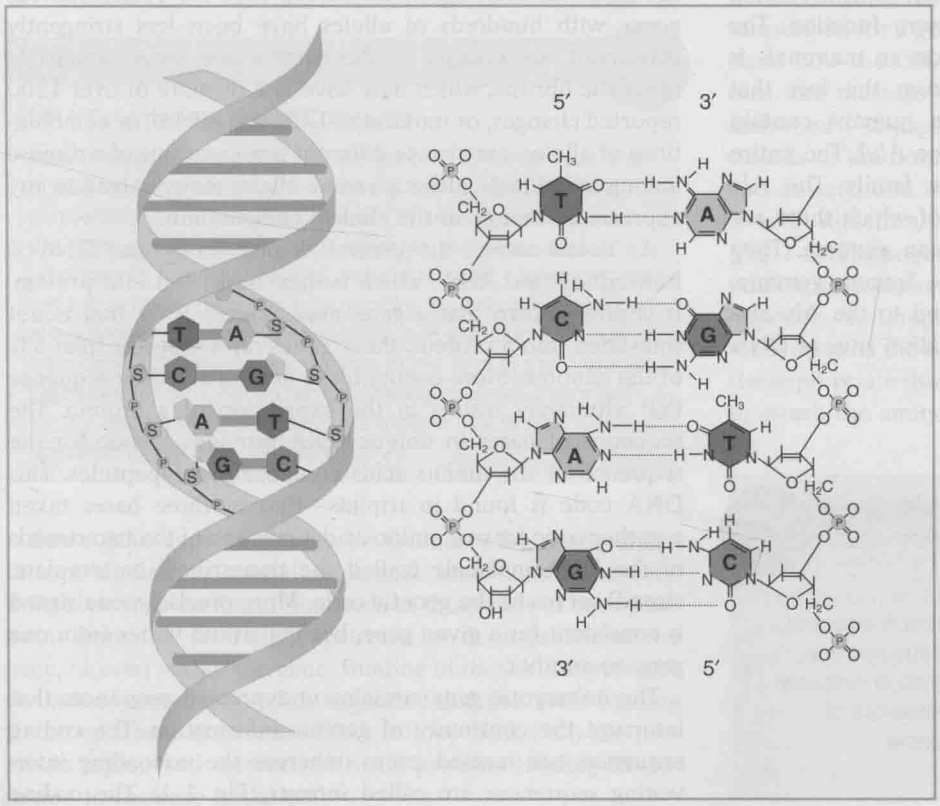

Figure 1-3. DNA is organized in an antiparallel configuration: one strand is 5′ to 3′ in one direction and the other strand is 5′ to 3′ in the opposite direction. A purine is bound to a pyrimidine by hydrogen bonds: A:T and G:C. The helix occurs naturally because of the bonds in the phosphate backbone.

Most unique-sequence genes code for proteins and are essentially structural or enzyme genes. Human DNA encodes 20,000 to 25,000 different gene products. The identification of many genes is known, along with their sequence, but the number of variations that occur within these is harder to predict. A phenotype, or an observable feature of specific gene expression, is associated with a smaller proportion of these variations. (See the *Online Mendelian Inheritance of Man*, available at: http://www.ncbi.nlm.nih.gov/omim.) Much of the time variations in genes are discussed relative to abnormal gene expression and disease; however, many mutations may have either no effect on gene expression or little effect on the function of the protein in the individual. For example, a protein may have less than 100% activity with little or no effect until the activity drops below a certain level.

Middle repetitive sequences represent redundant, tandemly arrayed copies of a given gene and may be transcribed just as unique-sequence genes. Specifically, these sequences refer to genes coding for transfer RNA (tRNA) and ribosomal RNA (rRNA). Because these RNAs are required in such large quantities for the translation process, several hundred copies of RNA-specifying genes are expected. As a striking example, the 18S and 28S fractions of rRNA are coded by about 200 copies of DNA sequences, localized in the tip regions of five acrocentric chromosomes in the human genome. It is estimated that human DNA is about 20% middle repetitive DNA.

Highly repetitive DNA is usually not transcribed, apparently lacking promoter sites on which RNA polymerase can initiate RNA chains. These highly repeated sequences may be clustered together in the vicinity of centromeres, or may be more evenly distributed throughout the genome. Presumably, the clustered sequences are involved in binding particular proteins essential for centromere function. The most common class of dispersed sequences in mammals is the *Alu* elements. The name derives from the fact that many of these repetitious sequences in humans contain recognition sites for the restriction enzyme *Alu*I. The entire group has been referred to as the *Alu* family. The *Alu* sequences are 200 to 300 bp in length, of which there are an estimated million copies in the human genome. They constitute between 5% and 10% of the human genome. Various debatable roles have been ascribed to the *Alu* elements, from "molecular parasites" to initiation sites of DNA synthesis.

MICROBIOLOGY

Restriction Endonucleases

Restriction endonucleases, also called restriction enzymes, are normal enzymes of bacteria that protect the bacteria from viruses by degrading the viruses. Restriction enzymes also recognize and cleave specific short sequences of human DNA, making them highly useful in gene characterization and clinical diagnostics.

BIOCHEMISTRY

Restriction Enzymes

A restriction enzyme cleaves both strands of the DNA helix. Sites of cleavage may produce blunt ends, 3′ overhanging ends, or 5′ overhanging ends. Many sequences recognized are palindromes. *Alu* elements contain an AGCT site that produces a blunt digestion site when exposed to the restriction enzyme *Alu*I, isolated from *Arthrobacter luteus*. The name of the organism from which the enzyme is isolated provides the abbreviation for the enzyme (*Alu*).

GENE ORGANIZATION

Each cell has 23 pairs of chromosomes, or 46 separate DNA double helices, with one chromosome from each pair inherited maternally and the other paternally. Twenty-two pairs are called *autosomes* and one pair is called the *sex chromosomes*. Each pair of autosomes is identical in size and organization of genes. The genes on these *homologous chromosomes* are organized to produce the same proteins. However, slight variations may occur, which changes the organization of the base pairs and can lead to a change in a protein. These changes can be called *polymorphisms* (from Greek "having many forms") and result from mechanisms creating changes, or mutations, within the DNA. Another name for variation in the same gene on homologous chromosomes is *allele*. Stated another way, an allele is an alternative form of a gene. Two alleles in an individual occur at the same place on two homologous chromosomes, and these may be exactly the same or they may be different. The presence of few alleles indicates the gene has been highly conserved over the years, whereas genes with hundreds of alleles have been less stringently conserved. An example of the latter is the gene responsible for cystic fibrosis, which may have one or more of over 1500 reported changes, or mutations. Different alleles, or combinations of alleles, may cause different presentations of a disease among individuals, although some alleles may not lead to any appreciable change in the clinical presentation.

As noted above, the central dogma states that DNA is transcribed into RNA, which is then translated into protein. It is now known that a gene may express RNA that is not translated into a protein; these genes represent less than 5% of the genome. More commonly, a gene is a coding sequence that ultimately results in the expression of a protein. The sequence of bases in unique DNA provides a code for the sequence of the amino acids composing polypeptides. This DNA code is found in triplets—that is, three bases taken together code for one amino acid. Only one of the two strands of the DNA molecule (called the transcribed, or template, strand) serves as the genetic code. More precisely, one strand is consistent for a given gene, but the strand varies from one gene to another.

The eukaryotic gene contains unexpressed sequences that interrupt the continuity of genetic information. The coding sequences are termed *exons*, whereas the noncoding intervening sequences are called *introns* (Fig. 1-4). The coding

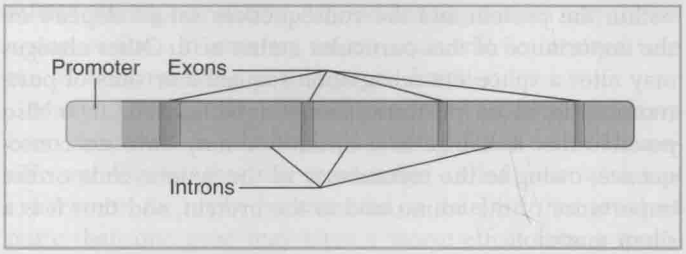

Figure 1-4. Organization of a gene showing the upstream promoter region, exons, and introns. Introns are removed by splicing during the formation of mRNA.

region of the gene begins downstream from the promoter at the initiation codon (ATG). It ends at a termination codon (UAG, UAA, or UGA). Sequences before the first exon and after the last exon are generally transcribed but not translated in protein.

The 5′ region of the gene contains specific sites important for the transcription of the gene. This region, called the *promoter,* has binding sites for transcription factors that regulate transcription initiation. Many cells contain the well-known seven-base-pair sequence TATAAAA, also referred to as the TATA box. The TATA binding protein binds to this site, which assists in the formation of the RNA polymerase transcriptional complex. Other promoter elements include the initiator (inr), CAAT box, and GC box. The latter is very important in regulating expression through methylation. More specific binding sites within the promoter vary from gene to gene. As imagined, this is an extremely complex region. It is the unique combination of different transcription factors binding that regulates differential expression of the gene in different cells and tissues.

BIOCHEMISTRY

DNA Orientation: Basic Concepts

DNA is arranged in a 5′-to-3′ orientation. By convention, the 5′ end is to the left and the 3′ end is to the right. Similarly, sequences to the left of a point are upstream and those to the right are downstream. For example, the promoter is upstream of the initiation site. Although these sequences are not transcribed, they are important for binding proteins to allow proper binding of polymerase and initiation of transcription. Similarly, sequences at the end of the gene are important for termination, and signaling sites are important for the addition of polyadenosine (polyA) that is not specified in the DNA template.

Some gene expression may be facilitated by transcription factors binding to special sequences known as *enhancers.* Enhancers may be found hundreds to thousands of base pairs away from the promoter, upstream or downstream of the gene, or even within the gene. Binding of these sites increases the rate of transcription. It is suggested that the factor binding to the enhancer may cause DNA to loop back onto the promoter region and interact with the proteins binding in this region to increase initiation.

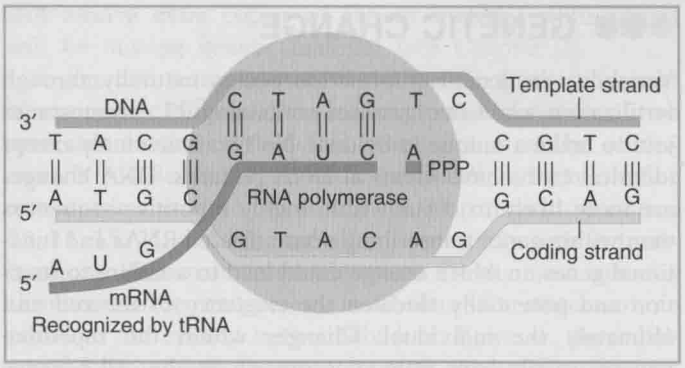

Figure 1-5. RNA is transcribed from the template strand and has a complementary sequence to the coding strand. Therefore, the coding strand sequence more accurately reflects the genetic code.

The entire gene is transcribed as a long RNA precursor, commonly referred to as the primary RNA transcript, or premessenger RNA; this is sometimes called heterogeneous nuclear RNA (hnRNA). Through RNA processing, the introns of the primary RNA transcript are excised and the exons spliced together to yield the shortened, intact coding sequence in the mature messenger RNA (mRNA). Specific enzymes that recognize precise signals at intron-exon junctions in the primary transcript assure accurate "cutting and pasting." There is no rule that governs the number of introns. The gene for the β chain of human hemoglobin contains two introns, whereas the variant gene that causes Duchenne-type muscular dystrophy has more than 60 introns. Nearly all bacteria and viruses have streamlined their structural genes to contain no introns. Among human DNAs, genes with no introns are less common.

The concept, mentioned above, of only one strand being transcribed for a gene can be confusing when trying to understand how the DNA code is transferred to RNA, which is, in turn, the message used to translate the code into a precise amino acid sequence of a protein. As noted, the two DNA strands of the double helix are antiparallel, with a 5′ and 3′ end at each end of the molecule. Transcription occurs in a 5′-to-3′ direction from the transcribed, or template, strand (Fig. 1-5). The sequence of this hnRNA, and subsequently the mRNA, is complementary to the antiparallel strand that is opposite the template strand. The antiparallel strand is also referred to as the coding strand. The anticodons of tRNA find the appropriate three-base-pair complementary mRNA codon to attach the amino acid specified.

BIOCHEMISTRY

Transcription and RNA Processing

Transcription is the synthesis of RNA from a DNA template, requiring RNA polymerase II. RNA is single stranded with an untranslated 5′ cap and 3′ polyA tail.

Small nuclear ribonucleoproteins (snRNPs) stabilize intron loops, in a complex called a spliceosome, for removal of introns. snRNPs are rich in uracil and are identified as U and a number: U1, U2, U3, etc.

GENETIC CHANGE

Variability in genetic information occurs naturally through fertilization when two gametes containing 23 chromosomes join to make a unique individual. No two individuals except identical twins have identical DNA patterns. DNA changes are more likely to occur within highly repetitive sequences than within genes transcribing nontranslated RNAs and functional genes, in which change could lead to a failure to function and potentially threaten the existence of the cell and ultimately the individual. Changes within the repetitive regions usually have little consequence on the cell because of the apparent lack of function. Repetitive sequences are similar but not identical among individuals and represent a great reservoir for mutational changes. These sequences represent the DNA "fingerprint" of an individual, most often referred to in court proceedings, because these regions demonstrate the same heritability observed with expressed regions of the chromosomes.

Aside from fertilization, which brings together chromosomes that have undergone recombination during gamete formation and chromosomes that have assorted randomly into gametes, changes in genetic material are generally observed as numerical or structural. These changes are called *mutations.* Numerical changes generally occur as a result of nondisjunction. This error in the separation of chromosomes may occur in the division of somatic cells, called *mitosis,* or in the formation of gametes, called *meiosis.* In meiosis, nondisjunction may occur in either the first or second stage of meiosis, called meiosis I or meiosis II, respectively. The greatest consequences of nondisjunction are those observed in meiosis because the resulting embryo has too many or too few chromosomes. Humans do not tolerate either excess or insufficient DNA well. Except for a few situations, the absence of an entire chromosome (monosomy) or the addition of an entire chromosome (trisomy) is incompatible with life for more than a few weeks to perhaps as long as a few months (see Chapter 2).

Changes in genetic material, less dramatic than in an entire chromosome, are generally tolerated inversely to the size of the change: the smaller the change, the better the cell may tolerate the change. Changes may occur at a single nucleotide—a *point mutation*—or involve a large portion of a chromosome. At the nucleotide level, a purine may be replaced by another purine, or a pyrimidine by another pyrimidine. This substitution process is known as a *transition.* However, if a purine replaces a pyrimidine, or vice versa, a *transversion* occurs. Consequences of these changes depend on where the change occurs. Obviously, there is a greater opportunity for an effect within an exon rather than within noncoding sequences. Even within an exon, the location of the change is important. If the change results in the creation of a stop codon, known as a *nonsense* mutation, the resulting protein may be truncated and hence either nonfunctional or with reduced function. If the change results in a different codon being presented for translation, the change may cause a different amino acid at a certain position (*missense* mutation) within the protein and the consequences would depend on the importance of that particular amino acid. Other changes may alter a splice site recognition sequence or sites of posttranscriptional or posttranslational modification. It is also possible that a change in a nucleotide may have no consequence, owing to the redundancy of the genetic code or the importance of the amino acid in the protein, and thus it is a silent mutation.

BIOCHEMISTRY

Genetic Code

Three nucleotides code for one amino acid. A change in the third nucleotide may have no effect on the code for a particular amino acid; this is the "wobble effect." For example, arginine is coded for by CGU, CGC, CGA, and CGG. A change in the first or second nucleotide will change the amino acid inserted into the protein. There is one codon for methionine and tryptophan. Other amino acids may be specified by two to six codons (none are specified by five). There are three stop, or "nonsense," codons.

1ST POSITION (5′ END)	2ND POSITION (MIDDLE)				3RD POSITION (3′ END)
U	U	C	A	G	
	Phe F	Ser S	Tyr Y	Cys C	U
	Phe F	Ser S	Tyr Y	Cys C	C
	Leu L	Ser S	STOP	STOP	A
	Leu L	Ser S	STOP	Trp W	G
C	Leu L	Pro P	His H	Arg R	U
	Leu L	Pro P	His H	Arg R	C
	Leu L	Pro P	Gln Q	Arg R	A
	Leu L	Pro P	Gln Q	Arg R	G
A	Ile I	Thr T	Asn N	Ser S	U
	Ile I	Thr T	Asn N	Ser S	C
	Ile I	Thr T	Lys K	Arg R	A
	Met M	Thr T	Lys K	Arg R	G
G	Val V	Ala A	Asp D	Gly G	U
	Val V	Ala A	Asp D	Gly G	C
	Val V	Ala A	Glu E	Gly G	A
	Val V	Ala A	Glu E	Gly G	G

More observable changes can occur when regions of a chromosome are deleted or duplicated. Loss of genetic material may occur from within a chromosome or at the termini and results in what may be called *partial monosomy.* Just as with base changes, a single nucleotide may be added or deleted from a sequence, with the consequences depending on its location. These changes, called *frameshift* mutations, within a coding sequence can alter the reading frame of the

mRNA during translation. Altered reading frames may create a stop codon, or incorrect amino acids will be inserted into the protein, resulting in suboptimal function.

Many deletions of larger regions of chromosomes have been described in which partial monosomies result in specific syndromes that are sometimes called *microdeletion syndromes*. As might be expected, a deletion that involves more than one gene may have a worse effect than a mutation in a single gene. Many of the described disorders involve deletions of millions of base pairs and numerous genes. Most of these are de novo mutations and have such significant presentations that the individuals do not pass the deletion on to another generation (Box 1-1). Duplication of genetic material results from errors in replication. These may occur when a segment of DNA is copied more than once or when unequal exchange of DNA occurs between homologous chromosome pairs. The results may be a direct, or tandem, repeat or an inverted repeat of the DNA. Unequal exchange, or recombination, occurs in meiosis when homologous chromosomes do not align properly. The recombination results in a deletion for one chromosome and a duplication for the other. In either case, DNA that has been gained or lost can result in unbalanced gene expression.

Genetic material may also be moved from one location to another without the loss of any material. Such movements may occur within a chromosome or between chromosomes. Within a chromosome, movements are usually seen as inversions. Inversions either include the centromere (pericentric inversion) or are in one arm of the chromosome (paracentric inversion) (Fig. 1-6). These changes provide significant challenges to the chromosome during meiosis. Proper alignment of homologous chromosomes is impossible. If recombination is attempted, distribution of genetic material to gametes can become unbalanced; some gametes may receive duplicate copies of DNA segments while others lack these DNA segments.

The movement of genetic material between chromosomes is called a *translocation*. Translocations that exchange material between two chromosomes are called *reciprocal* translocations. These translocations generally have little consequence for the individual in whom they arise. However, translocations become important during the formation of gametes and segregation of the chromosomes. Some gametes will receive extra copies of genetic material while others will be missing genetic material (see Chapter 2).

A common rearrangement is the fusion of two long arms of acrocentric chromosomes leading to the formation of two new chromosomes. When this fusion occurs at the centromere, it is called a *robertsonian* translocation. There are five acrocentric chromosomes among the 23 pairs (chromosomes 13, 14, 15, 21, and 22), and all are commonly seen in translocations. Robertsonian translocations are the most common chromosomal rearrangement. In a balanced arrangement, no problems are evident in the individual. However, the unbalanced form presents the same concerns as partial monosomy or partial trisomy.

As noted, a mutation is a heritable change in genetic material. It may be spontaneous, as with some nondisjunctions, insertions, or deletions, or induced by an external factor. This external factor, a *mutagen,* is any physical or chemical agent that increases the rate of mutation above the spontaneous rate; the spontaneous rate of mutation for any gene is 1×10^{-6} per generation. Therefore, determining whether a mutation results from a spontaneous event within the cell or from a mutagen requires evaluation and comparison of the rates of mutation.

Mutagens are generally chemicals and irradiation (Box 1-2). Chemical mutagens can be classified as (1) base analogs that mimic purines and pyrimidines; (2) intercalating agents that alter the structure of DNA, resulting in nucleotide insertions and frameshifts; (3) agents that alter bases, resulting in different base properties; and (4) agents that alter the structure of DNA, resulting in noncoding regions, cross-linking of strands, or strand breaks.

Ionizing radiation damages cells through the production of free radicals of water. The free radicals interact with DNA and protein, leading to cell damage and death. Obviously, those cells most vulnerable to damage are rapidly dividing cells. The extent of the damage is dose dependent. Cells that are not killed have damage—mutations—to the DNA at sublethal doses. Such damage is demonstrated by base mutations, DNA

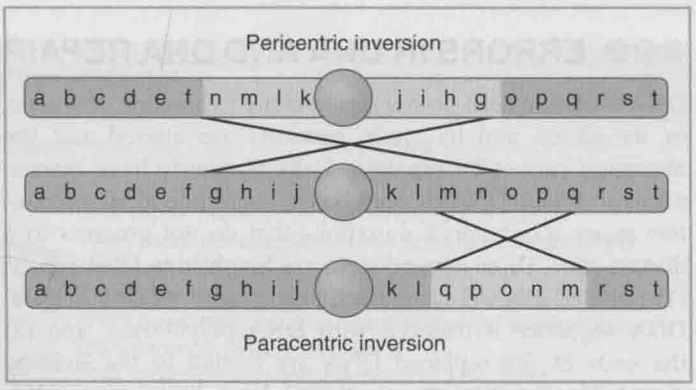

Figure 1-6. Inversions of DNA on a chromosome are distinguished by the involvement of the centromere. Pericentric inversions include the centromere. Paracentric inversions occur in either the p or q arm.

Box 1-1. EXAMPLES OF DELETION SYNDROMES

Cri du chat syndrome (5p15)
Prader-Willi syndrome (15q11-13)
Angelman syndrome (15q11-13)
DiGeorge syndrome (22q11.2)
Smith-Magenis syndrome (17p11.2)
Wolf-Hirschhorn syndrome (4p16.3)

Box 1-2. EXAMPLES OF MUTAGENS

Chemicals

Base analogs
- Aminopurine: resembles adenine and will pair with T or C
- Bromouracil: resembles thymine

Intercalating agents
- Ethidium bromide
- Acridine orange
- Nitrous oxide: causes deamination
- Nitrosoguanidine
- Methyl methanesulfonate: adds methyl or ethyl groups
- Ethyl methanesulfonate
- Psoralens: cause cross-linking
- Peroxides: cause DNA strand breaks

Irradiation

Ionizing radiation
- X-rays
- Gamma rays

Ultraviolet radiation
- UV-A: creates free radicals and some dimers
- UV-B: forms pyrimidine dimers, blocking transcription and replication
- UV-C: forms pyrimidine dimers, blocking transcription and replication

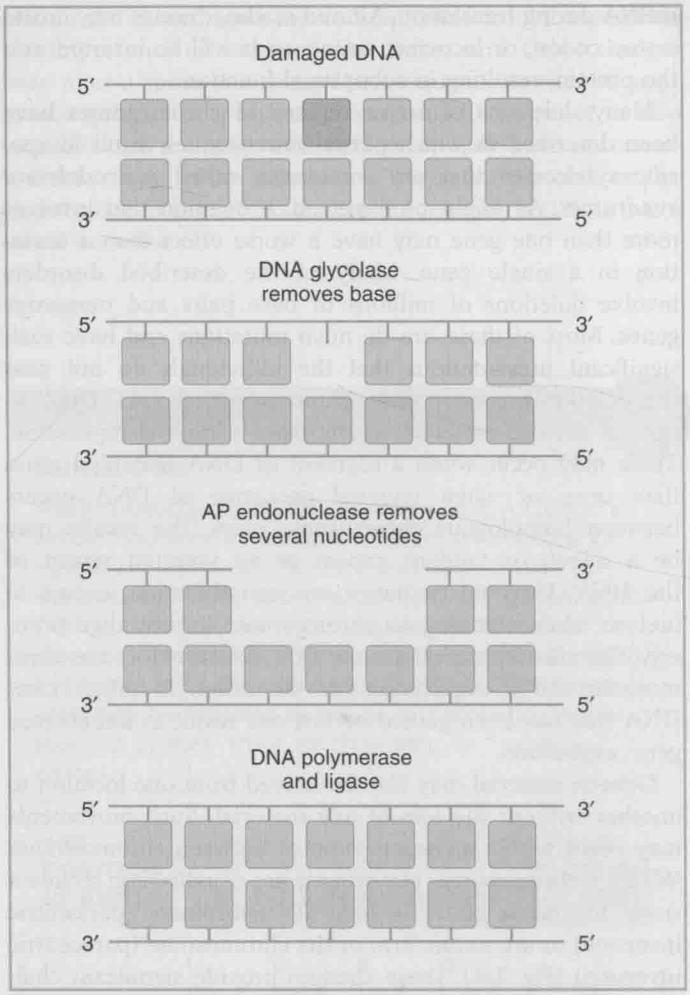

Figure 1-7. Base excision repair is the mechanism most commonly employed for incorrect or damaged bases. Specificity of repair is conferred by specific DNA *N*-glycosylases, such as uracil (or another base) DNA *N*-glycosylase. These glycosylases hydrolyze the *N*-glycosidic bond between the base and the deoxyribose. *AP,* apurinic/apyrimidinic.

cross-linking, and breaks in DNA. Breaks in the DNA of chromosomes may result in deletions, rearrangements, or even loss.

Ultraviolet (UV) radiation is non-ionizing because it produces less energy. UV-A (≥320 nm) is sometimes called "near-UV" because it is closer to visible light wavelength. UV-B (290–320 nm) and UV-C (190–290 nm) cause the greatest damage. The most damaging lesion is the formation of pyrimidine dimers from covalent bonds formed between adjacent pyrimidines. These dimers block transcription and replication.

ERRORS IN DNA AND DNA REPAIR

DNA mutations can be significant if the expression of a gene, or its alleles, and its allelic products are altered and the alteration cannot be repaired. Cells obviously have mechanisms to repair DNA damage, since each individual encounters many spontaneous mutations that do not progress to a disease state. Three general steps are involved in DNA repair: (1) mutated DNA is recognized and excised, (2) the original DNA sequence is restored with DNA polymerase, and (3) the ends of the replaced DNA are ligated to the existing strand. The mechanisms employed by cells to accomplish these steps include base excision, nucleotide excision, and mismatch repair.

Individual bases need replacing because of oxidative damage, alkylation, deamination, or a structural error in which no base is attached to the phosphate-sugar backbone. Unlike other types of mutations, these examples cause little distortion of the DNA and are repaired by *base excision* (Fig. 1-7). DNA glycosylases release the base by cleaving the glycosidic bonds between the deoxyribose and the base. DNA polymerase I replaces the base to restore the appropriate pairing (A:T or G:C), followed by ligation to repair the ends. *Glycosylases are specific for the base being removed,* and if there is a deficiency of a particular glycosylase, repair is compromised.

More extensive damage to DNA than single base pairs may distort the DNA structure. Damage of this type requires the removal of several nucleotides to accomplish repair. *Nucleotide excision repair* (Fig. 1-8) differs from base excision repair, which requires specific enzyme recognition of the base needing repair and of the size of the repair. The general mechanism of nucleotide excision repair is