

Applications of recombinant DNA technology in gastrointestinal medicine and hepatology: Basic paradigms of molecular cell biology.

Part A: Eukaryotic gene structure and DNA replication

Gary E Wild MD CM PhD FRCPC¹, Patrizia Papalia BSc¹, Mark J Ropeleski MD CM FRCPC¹, Julio Faria MD CM FRCSC¹, Alan BR Thomson MD PhD FRCPC²

GE Wild, P Papalia, MJ Ropeleski, J Faria, ABR Thomson. Applications of recombinant DNA technology in gastrointestinal medicine and hepatology: Basic paradigms of molecular cell biology. Part A: Eukaryotic gene structure and DNA replication. *Can J Gastroenterol* 2000;14(2):99-110. Progress in the basic sciences of cell and molecular biology has provided an exciting dimension that has translated into clinically relevant information in every medical subspecialty. Importantly, the application of recombinant DNA technology has played a major role in unravelling the intricacies related to the molecular pathophysiology of disease. This series of review articles constitutes a framework for the integration of the database of new information into the core knowledge base of concepts related to the pathogenesis of gastrointestinal disorders and liver disease. The goal of this series of three articles is to review the basic principles of eukaryotic gene expression. The first article examines the role of DNA in directing the flow of genetic information in eukaryotic cells.

Key Words: DNA; Eukaryotic cells; Nucleic acids; RNA

Applications de la technologie de l'ADN recombinant en gastro-entérologie et en hépatologie : paradigmes fondamentaux de la biologie moléculaire. Partie A : structure et réplication de l'ADN du gène eucaryotique.

RÉSUMÉ : Les progrès de la recherche fondamentale en biologie cellulaire et moléculaire ont ouvert la porte à une nouvelle dimension et à de nouvelles données cliniques pertinentes pour toutes les sous-spécialités médicales. Fait important à noter, l'application de la technologie de l'ADN recombinant a joué un rôle de tout premier plan pour nous aider à comprendre la physiopathologie moléculaire complexe des maladies. Cette série d'articles de synthèse se veut un cadre d'intégration pour l'ensemble des données qui viennent d'enrichir le corpus de nos connaissances relativement aux concepts expliquant la pathogenèse des maladies gastro-intestinales et hépatiques. L'objectif de cette série de trois articles est de passer en revue les principes fondamentaux de l'expression du gène eucaryotique. Le premier article se penche sur le rôle de l'ADN dans la transmission de l'information génétique dans les cellules eucaryotiques.

For most gastroenterologists, the principles of cell and molecular biology have not traditionally played a major role in day to day clinical practice. However, tremendous ad-

vances in the discipline of molecular medicine have provided new insights into the cellular and molecular pathological basis of disease. This ever increasing expansion

¹Department of Medicine, Division of Gastroenterology, McGill University Health Centre, and McGill University Inflammatory Bowel Disease Research Program, Montreal, Quebec; ²Department of Medicine, Division of Gastroenterology, University of Alberta, Edmonton, Alberta
Correspondence: Dr Gary E Wild, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4. Telephone 514-934-8308, fax 514-934-8411, e-mail gwild@is.mgh.mcgill.ca

Received for publication March 23, 1999. Accepted July 15, 1999

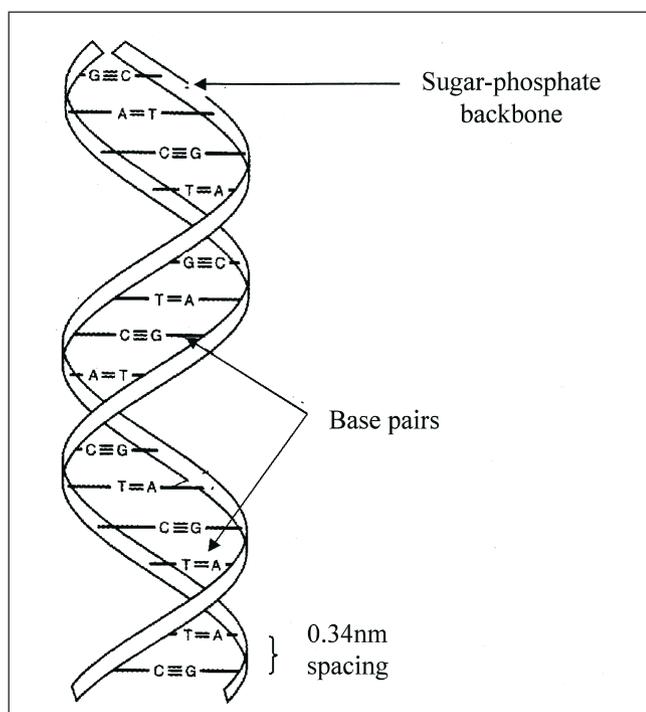


Figure 1) Base pairing and the antiparallel orientation of DNA. The two DNA strands in the helix have opposite polarity, with one strand running in a 5' to 3' direction and the other running in the 3' to 5' direction. Four bases (adenine [A], thymine [T], cytosine [C] and guanine [G]) reside on the inside of the helix to allow hydrogen bonding between purine and pyrimidine residues

of the knowledge base has transformed the understanding and management of a diverse array of diseases. The cumulative research efforts in cell and molecular biology have provided an exciting dimension that has translated into clinically relevant information in every medical subspecialty. For example, hematologists have defined the molecular basis of the hemoglobinopathies. Endocrinologists have defined the cellular and molecular networks that mediate the action of hormones. Neurologists have identified a host of gene mutations that lead to neurodegenerative disorders. Finally, the identification of the cystic fibrosis transmembrane regulator has facilitated the molecular diagnosis of the disease, and, as a result, gene therapy protocols are being conducted at several centres.

Many of the recent advances in molecular medicine have arisen through efforts driven by the Human Genome Project. It is apparent that molecular biology has accounted for a dramatic paradigm shift in both the teaching and the practice of medicine. This series of review articles constitutes a framework for the integration of the database of new information into the core knowledge base of concepts related to the pathogenesis of gastrointestinal disorders and liver disease. We hope to provide the reader with a set of tools to facilitate the understanding of some of the basic concepts of recombinant DNA technology and the role it has played in unravelling the intricacies related to the molecular pathophysiology of disease. As well, we wish to provide the reader with a flavour for the pervasive impact of molecular medi-

cine in the areas of gastroenterology and hepatology. The goal of this first series of three articles is to review the basic principles of eukaryotic gene expression.

NUCLEIC ACIDS AND INFORMATION TRANSFER IN THE CELLS

DNA is the storage form of genetic information in cells. The structure of DNA was determined by Watson and Crick in 1953, and this discovery has revolutionized the thinking in modern cell biology. All DNA molecules consist of four types of nucleotides joined together by phosphodiester bonds to form polynucleotides. The nitrogenous bases found in DNA consist of purines (ie, adenine [A] and guanine [G]) and pyrimidines (ie, cytosine [C] and thymine [T]) (Figure 1). The nucleotides are linked together by covalent phosphodiester bonds that join the 5' carbon of one deoxyribose to the 3' carbon of the adjacent deoxyribose to form polynucleotide genes. The double-stranded DNA helix with its two polynucleotide strands of DNA run in an antiparallel orientation, and the DNA strands are held together by hydrogen bonding between A and T residues, and G and C residues. The antiparallel orientation in base pairing is an important concept in nucleic acid biochemistry. One strand runs in a 5' to 3' direction, and the complementary strand runs in the 3' to 5' direction (Figure 1). Thus, the two strands of the double helix are complementary. For example, the sequence CTGAAGCGCTTA on one strand of DNA has the complementary sequence GACTTCGCGAAT on the opposite strand of DNA in an antiparallel orientation. The variation of the sequence of nucleotides along the DNA strand determines the function of each section of the DNA molecule as well as its ability to transmit information to RNA and protein.

RNA molecules consist of nucleotides linked together by phosphodiester bonds. RNA generally occurs as single-stranded polynucleotides and contains ribose in place of the deoxyribose found in DNA. RNA is made up of four bases, including A, G and C, but contains uracil (U) in place of T. Because U has the ability to bind with A in the same way that T binds with A, the four bases found in RNA – A, U, G and C – can form complementary pairs with other bases found in RNA as well as with the bases found in DNA. These biochemical properties highlight the major function of the RNA molecule in the transfer of information from DNA to protein in eukaryotic cells. RNA often contains intramolecular hydrogen bonding, which gives rise to secondary structures. Intrastrand base pairing creates structures known as stem loop structures, with the base pairing sections forming the stem and noncomplementary bases forming the loop.

Eukaryotic cells contain five classes of RNA: messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), heterogeneous nuclear RNA (hnRNA) and small nuclear RNA (snRNA). mRNA makes up a small percentage of the total RNA (1% to 5%) in eukaryotic cells, has a short half-life and demonstrates a large variation in base sequence from one mRNA molecule to another. mRNA is the chemical messenger that carries information from the DNA helix to the protein synthesizing machinery in the cytoplasm.

tRNA molecules are polynucleotides made up of 75 to 95 nucleotides that carry specific amino acids to the ribosomes during protein synthesis. There is a unique tRNA that specifically recognizes each of the 20 amino acids. In some instances, there is more than one tRNA species for a single amino acid. rRNA is the most abundant of the RNA species in eukaryotic cells and is found associated with proteins in structures called ribosomes. These specific rRNAs of eukaryotic cells are designated by their sedimentation coefficients (S values). Human ribosomes contain 28S, 18S, 5.8S and 5S rRNA species.

hnRNA and snRNA species are located in the nucleus of eukaryotic cells. hnRNA is the immediate product of transcription and is complementary to one strand of the DNA helix. hnRNA is the precursor to mRNA before it undergoes further processing. snRNA is found associated with specific proteins that are involved in the processing of the hnRNA to mRNA before exit of the mRNA from the nucleus to the cytoplasm. The role of these RNA molecules in transcription and translation is discussed in detail in subsequent sections. These topics have been reviewed in detail (1-8).

MOLECULAR ANATOMY OF EUKARYOTIC GENES

Eukaryotic genomes are larger and more complex than those of primitive prokaryotes (ie, bacteria). For example, the human genome contains approximately 100,000 genes, and much of its complexity arises from the abundance of several different types of noncoding DNA sequences.

A gene can be defined as a segment of DNA that is expressed to yield a functional product that may be either an RNA or a peptide. The structural features that are common to all eukaryotic genes are illustrated in Figure 2. The sequence of base pairs confers gene specificity and determines the specificity of the product that it encodes. However, not all of the nucleotides present in the gene are expressed in the final product. Eukaryotic genes are often split into exons – sequences that remain in the final mature mRNA – and introns – sequences that are removed from the primary mRNA transcript early during processing, most which have no known function. In addition to encoding sequence information that ultimately defines the protein product, exons contain other sequences that are essential for the organized function of mRNA. Thus, an exon is defined as a sequence in the primary RNA transcript that is conserved during the processing of the transcript into a mature mRNA molecule.

Unique sequences that signal the start of transcription are present in each gene. These sequences are promoter sequences, and they determine the site at which transcription is initiated on the DNA molecule. Transcription is initiated when RNA polymerase along with transcription factors bind to the promoter site and catalyze the synthesis of RNA. RNA polymerase transcribes RNA by using the sequence of bases from one strand of the DNA double helix that serves as a template. RNA is synthesized as a single-stranded molecule in the 5' to 3' direction.

Further processing of mRNA transcripts to yield a mature

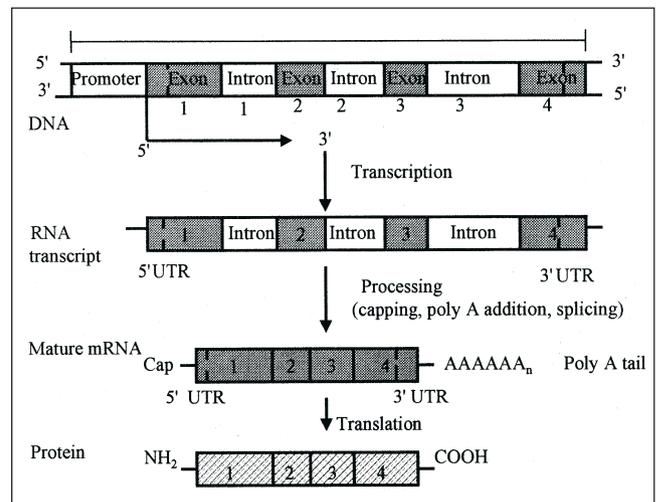


Figure 2) Molecular anatomy of human genes. A typical human gene contains exon and intron sequences that are transcribed by RNA polymerase into the primary transcript. This primary transcript is subsequently processed by the addition of a cap structure at the 5' end and the addition of a poly A tail to the 3' end. The intron sequences are removed, and the exonic RNA sequences are spliced together. The mature mRNA contains only exonic RNA sequences that have information for protein sequences as well as signals for the initiation and termination of protein synthesis. UTR Untranslated regions

RNA product involves a series of steps, including the addition of a cap structure at the 5' end of the mRNA and the addition of a poly A tail at the 3' end. Untranslated regions (UTRs) are situated at both the 3' and 5' ends of the mRNA and are sequences in the exons that remain in the mRNA but are not translated into proteins. These regions contain signals required for mRNA processing and its subsequent translation into protein. For further details, see references 4 and 8.

ORGANIZATION OF EUKARYOTIC GENOMES

The average polypeptide is approximately 400 amino acids long; thus, the average size of the coding sequence of a gene is 1200 base pairs. Each amino acid is determined by a set of three nucleotides called a codon. In contrast to *Escherichia coli* and yeasts, the human genome contains large amounts of noncoding DNA. Thus, only a small proportion of the total 3×10^9 base pairs of the human genome is expected to correspond to protein coding sequences. The average gene spans 10,000 to 20,000 base pairs (including introns) such that the human genome consists of approximately 100,000 genes that correspond to 3% of the total human DNA. This topic is covered extensively in references 4 and 8.

Several types of highly repeated sequences exist in eukaryotic genomes. One class, called simple-sequence DNA, contains tandem arrays of thousands of copies of short sequences ranging from five to 200 nucleotides. Such repeat sequence DNA account for approximately 10% to 20% of the DNA in higher eukaryotes and is called satellite DNA. Other repetitive DNA sequences are scattered throughout the genome rather than being clustered as tandem repeats. These sequences are classified as either short (SINES) or long

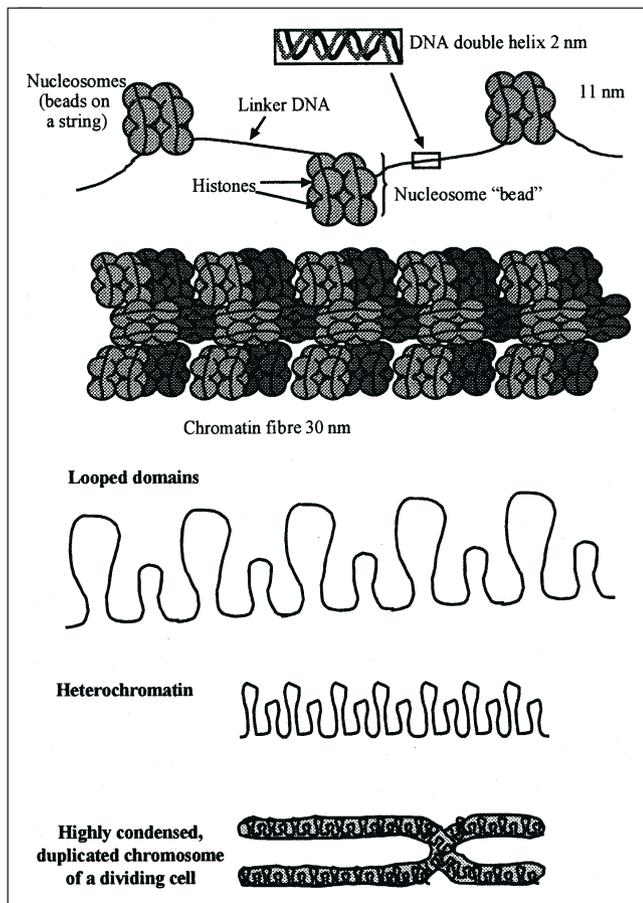


Figure 3) The packaging of DNA in the nucleus. A model is depicted for the progressive stages of DNA coiling and folding in the nucleus. The hierarchy of structure features arising from the DNA double-helix include nucleosomes, chromatin fibres and their looped domains, and heterochromatin, which makes up the arms of the chromosomes

(LINEs) interspersed elements. The major SINEs in the mammalian genome are Alu sequences, which contain a signal site for the restriction endonuclease AluI. These Alu sequences (300 base pairs long) are dispersed throughout the genome and account for nearly 10% of the total cellular DNA. The major human LINEs are about 6000 base pairs in length and repeat approximately 50,000 times in the human genome. In contrast to Alu sequences, LINE sequences are transcribed, and some encode proteins of unknown function.

Eukaryotic DNA is tightly associated with small basic proteins (ie, rich in arginine and lysine) called histones. The complexes between eukaryotic DNA and proteins are called chromatin, which contain about twice as much protein as DNA. The basic amino acids contained in histones have been identified: H1, H2A, H2B, H3 and H4. In addition, chromatin contains a variety of nonhistone chromosomal proteins, which are involved in DNA replication and gene expression. The association of DNA and protein to form chromatin is illustrated in Figure 3.

The basic structural unit of chromatin is called the nucleosome, which is composed of repeating 200 base pair units. Nucleosomes contain a core particle that contains 146 base pairs of DNA wrapped 1.75 times around a histone core

consisting of two molecules each of H2A, H2B, H3 and H4. The other structural feature of the nucleosome is the chromosome, which contains two full turns of DNA (166 base pairs) held in place by one molecule of H1. The structure (ie, degree of condensation) of chromatin is intimately linked to the control of gene expression in eukaryotes. The extent of chromatin condensation varies during the life cycle of the cell. In nondividing cells, most of the chromatin, called euchromatin, is decondensed and distributed throughout the nucleus. Genes are transcribed during this period of the cell cycle, and the DNA is replicated in preparation for mitosis. By contrast, about 10% of interphase chromatin is in a very highly condensed state called heterochromatin. Heterochromatin is transcriptionally inactive and contains highly repeated DNA sequences.

The human genome is distributed among 24 chromosomes (22 autosomes and the two sex chromosomes), each containing between 5×10^4 and 26×10^4 kilobases of DNA. The chromosomes have three well defined structures that are essential for their replication: DNA replication origins, centromeres and telomeres. The DNA replication origins are considered in detail in the section on DNA synthesis. Centromeres consist of highly repetitive DNA sequences and are the site where the two sister chromatids are attached. The function of the centromere is to ensure the equal distribution of each chromosome to the daughter cells at cell division. The telomere is an important structure associated with the ends of all human chromosomes. Telomeric DNA consists of multiple tandem repeats of the sequence TTAGGG located at both ends of each chromatid. Telomeres perform a variety of functions in human cells, including the following.

- Telomeres maintain chromosomal stability and prevent the formation of end-to-end fusions. The presence of telomeric sequences protects chromosomal ends from nuclease degradation.
- Telomeres ensure the proper replication of the ends of chromosomes. DNA ends are not completely replicated during DNA replication and require the presence of the enzyme telomerase to add nucleotides to the extreme ends of the DNA molecule. The presence of noncoding telomeric sequences at the chromosomal ends protects the coding sequences of the DNA that might be located near the terminal ends of a chromosome from being lost during each cycle of replication.
- Telomeres serve as markers of chromosomal integrity. In the event that a chromosome is damaged, the cell cycle stops temporarily such that DNA repair mechanisms can repair the damage.

FLOW OF GENETIC INFORMATION IN EUKARYOTIC CELLS

The expression of genetic information in all eukaryotic cells is largely a one way system of traffic. DNA directs the synthesis of RNA, and RNA specifies the synthesis of polypeptides that subsequently form proteins. Because of its universality, the DNA to RNA to protein flow of genetic in-

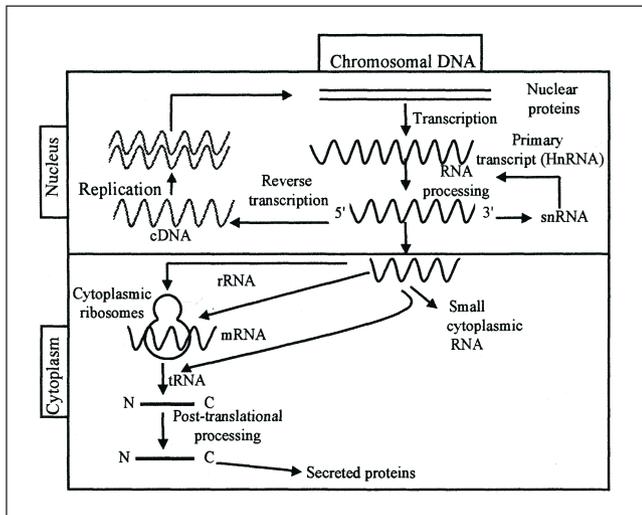


Figure 4) Gene expression in the eukaryotic cell. The expression of genetic information in eukaryotic cells is largely a one way system. DNA specifies the synthesis of RNA, and RNA specifies the synthesis of polypeptides, which subsequently form proteins. A small proportion of nuclear RNA molecules can be converted to cDNA by reverse transcriptases and subsequently integrate into chromosomal DNA

formation is called the 'central dogma of molecular biology'. The synthesis of RNA using DNA as a template and RNA polymerase is called 'transcription'. Transcription occurs in the nucleus of eukaryotic cells and to a limited extent in mitochondria. The second step involves polypeptide synthesis and is called 'translation'. Translation occurs on ribosomes, which are large RNA protein complexes found in the cytoplasm. The RNA molecules that specify polypeptides are known as mRNAs. Gene expression has traditionally followed a colinearity principle where the linear sequence of the nucleotides in DNA is decoded to give a linear sequence of nucleotides in RNA. In turn, this linear sequence can be decoded to give rise to a linear sequence of amino acids in the polypeptide product. A challenge to this concept has been made by recent findings that eukaryotic cells, including mammalian cells, contain nonviral chromosomal DNA sequences that encode cellular reverse transcriptases. Many different classes of viruses have a genome that consists of RNA. Retroviruses such as human immunodeficiency virus are a subclass of RNA viruses in which the RNA replicates via a DNA intermediate by using an RNA-dependent DNA polymerase called reverse transcriptase. Because some nonviral RNA sequences in eukaryotic cells are known to act as templates for cellular DNA synthesis, the principle of unidirectional flow of genetic information is no longer strictly valid. The overall flow of genetic information and gene expression in eukaryotic cells is illustrated in Figure 4, and is reviewed in references 1-8.

THE CELL CYCLE

The cellular processes that determine DNA replication and mitosis are the keys to normal cell growth and development. These processes occur during a well regulated and orderly progression through the mammalian cell cycle (Figure 5).

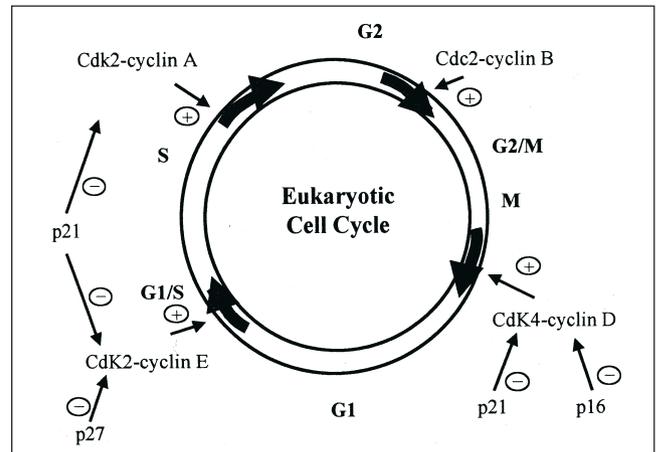


Figure 5) Eukaryotic cell cycle. Cyclin-dependent kinases (Cdk), cyclins and Cdk inhibitors (CKIs) interact during the cell cycle. Progression during the cell cycle is regulated by interaction of positive and negative regulatory factors. The positive progression is directed by multiple cyclin-cyclin-dependent kinase complexes, which act by phosphorylating various proteins at the different stages in the cycle. Negative regulatory factors include CKIs such as p16, p21 and p27, which inhibit phosphorylation of proteins by kinase and stop the cell cycle

The regulation of the cell cycle ultimately determines how a cell will cycle among growth, differentiation and division phases. Cell cycle control is a key determinant of differentiation or the decision to stop cycling. The loss of control of the cell cycle leads to abnormal cell growth, which results in tumourigenesis, developmental defects or premature programmed cell death (ie, apoptosis). The topic of cell cycle control is covered in detail in references 9-13.

The mammalian cell cycle comprises four distinct phases – gap (G) 1 phase, synthetic (S) phase, G2 phase and mitotic (M) phase. The period between one M phase and the next is called interphase. Interphase is divided into the remaining three phases of the cell cycle (ie, G1, S, G2). The G1 phase is the interval between the completion of the M phase and the onset of the S phase. The G2 phase is the interval between the end of the S phase and the beginning of the M phase. DNA is replicated during the S phase and is distributed equally to two daughter cells during the M phase. The cells prepare for either the S phase or the M phase during the G1 and G2 (interval when proteins are synthesized in preparation for mitosis) phases, respectively. Cells that do not undergo division, such as neurons, exit the cell cycle and enter a phase called G0. If cells in G0 are stimulated to grow, they move from G0 into the G1 phase. Progression through the cell cycle is mediated by multiple cyclin-dependent kinases (Cdk) that are sequentially activated by the binding of cyclins. The activated Cdk-cyclin protein complex phosphorylates specific proteins that are required for the reactions unique to each distinct phase of the cell cycle. Cyclins vary dramatically during the cell cycle. For example, cyclin B levels increase during interphase and subsequently decline during the M phase. The changes in the level of cyclin B are correlated with the activity of a specific Cdk called Cdc2, which is active when cyclin B levels peak and becomes inactive.

TABLE 1
Cyclin-dependent kinases (Cdk), cyclins and cyclin-dependent kinase inhibitors (CKIs) at different stages of the cell cycle

Cell cycle phase	Cdk	Cyclin	CKI	
			KIP*	INK [†]
G1	Cdk4	Cyclin D	p21, p27	p15, p16
G1/S	Cdk2	Cyclin E	p21, p27	
S	Cdk2	Cyclin A	p21	
G2/M	Cdc2	Cyclin B	p21	
M	Cdc2	Cyclin B, cyclin A		

*Kinase inhibitory proteins (KIP) (p21 and p27) bind multiple cyclin-Cdk complexes that prevent activation or inhibit kinase activity; [†]Inhibitor of Cdk (INK) proteins (p15 and p16) are specific for Cdk4/6 and cyclin D, and bind Cdk and inhibit the binding of cyclin D

tive as cyclin B declines. Thus, the phosphorylating activity of Cdc2 is modulated during the cell cycle by the availability of cyclin B. The activation of Cdc2 also depends on phosphorylation of a specific threonine residue, thus adding a second layer to the control of the kinase activity. A variety of cell cycle 'checkpoints' monitor progression through the cell cycle. Deviation from the normal cell cycle impedes progression beyond the checkpoint, and the cell cycle is halted until the defect is corrected. Thus, the orderly progression through the cell cycle depends on both positive factors, which drive the cell cycle forward, and negative factors, which halt the cycle at a particular stage. Cdk and specific cyclins are the main positive factors, which function at each stage of the cell cycle. Negative factors block the activity of the specific Cdk and are called cyclin-dependent kinase inhibitors (CKI) (Table 1).

The following mechanisms are responsible for the inactivation of an active Cdk-cyclin complex.

- The cyclin molecule can be degraded through the ubiquitin protein-degrading system.
- The critical phosphate required for activation of the kinase activity can be removed from the protein by a specific phosphatase.
- CKI molecules interact with Cdk or Cdk-cyclin complexes and inhibit the kinase activity. Two classes of CKIs have been described, the inhibitor of Cdk (INK) class and the kinase inhibitory protein (KIP) class (Table 1).

Thus, the interplay between the activation and deactivation of the Cdk activities at various stages of the cell cycle is the key determinant of the normal progression and regulation of the cell cycle.

The G1 phase: The G1 phase heralds the onset of the cell cycle. Resting cells that are stimulated to divide enter the G1 phase, and once the cell passes this point it is committed to entering the S phase and subsequently divides. The key positive regulators of the G1 phase are Cdk4 and cyclins of the D family, which form a complex capable of phosphory-

lating a host of proteins required for cell function in the G1 phase. The retinoblastoma (pRb) protein is a key protein phosphorylated by the Cdk4-cyclin D in G1. pRb exists in a nonphosphorylated form during the first two-thirds of the G1 phase and becomes phosphorylated just before the transition from the G1 to the S phase. Nonphosphorylated pRb functions by restricting cell growth, whereas phosphorylated pRb is associated with a loss of growth inhibitory function and allows the cell to proceed through the cell cycle. Thus, pRb functions as a regulator that represses or activates specific promoters through interaction with and modification of the activities of transcription factors that bind to DNA and regulate the expression of cell cycle genes. The phosphorylation of pRb by the Cdk4-cyclin D complex allows previously repressed genes to be transcribed and allows the cell to progress from the G1 to the S phase.

The Cdk inhibitor p27 is a second important control that regulates the progression of a cell from the G1 to the S phase. This protein binds to the Cdk2-cyclin E complex and inactivates it. The cells are unable to proceed onto the S phase and remain arrested in G1. Growth-promoting factors result in the degradation of p27, activation of the Cdk2-cyclin E complex and transition into the S phase. The ubiquitin protein-degrading system is responsible for the degradation of p27.

The S phase: Entry into the S phase is determined by a putative cytoplasmic signal that is most likely an active Cdk-cyclin complex. Entrance into the S phase from the G1 phase and progression through the S phase to the G2 phase depend on the function of specific Cdk-cyclin complexes. Cdk2 initially binds cyclin E as the cells proceed into the S phase. Cyclin A activates Cdk2 and phosphorylation of proteins required for DNA replication.

The G2/M phase: The G2/M phase is a critical checkpoint where cells decide whether to enter mitosis. The critical proteins involved in the G2/M checkpoint include Cdc2 and cyclin B, which form a complex, and the Cdc2-cyclin B complex is essential for entrance into and exit from the M phase. This involves activation and deactivation of the Cdc2-cyclin B complex through a series of phosphorylation and dephosphorylation steps.

The M phase: The sudden activation of the Cdc2-cyclin B complex by dephosphorylation, which occurs at the G2/M border, results in the phosphorylation of a variety of proteins required for mitosis. Three checkpoints are key to the orderly entrance into and exit from mitosis, with each daughter cell receiving an exact copy of the parental genome. These three checkpoints are the transition from G2 to M concurrent with the activation of the Cdc2-cyclin B complex; the M phase checkpoint that occurs during metaphase (the point that regulates the timing of the separation of the chromatids and the initiation of anaphase); and the immediate proteolytic destruction of cyclin B at the onset of anaphase, with the concomitant inactivation of Cdc2 (which allows the cell to exit the M phase and enter a new G1 phase). These checkpoints are regulated by the ubiquitin pathway.

The role of p53 and p21 in the control of cell damage: The orderly progression within the cell cycle as well as the ability

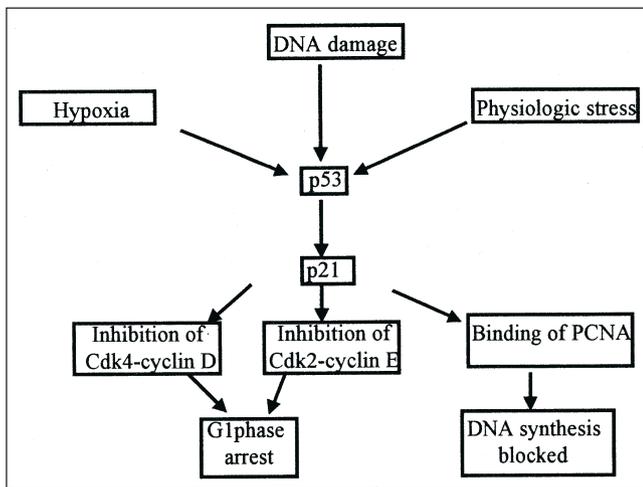


Figure 6) Control of damage by p53 and p21. Cellular damage results in increased p53 activity. p53 functions as a transcription factor and induces the transcription of p21, a cyclin-dependent kinase inhibitor. The p21 interacts with multiple cyclin-dependent kinase (Cdk)-cyclin complexes, inhibits the kinase activity and halts the cells in G1 phase. p21 also binds proliferating cell nuclear antigen (PCNA), inhibiting DNA synthesis

of the cell to sense any perturbation from its normal state is crucial to normal cell growth and development. Cells have evolved negative regulatory mechanisms that sense physiological disturbances, DNA damage, hypoxia, nutrient depletion and viral infection. The cell can arrest at a particular stage of the cell cycle, or in some instances the cell undergoes programmed cell death called apoptosis.

The DNA binding protein, p53, orchestrates the negative regulatory mechanisms that occur when the cell is damaged. The p53 protein is a tumour-suppressor protein and activates transcription of the gene encoding the Cdk inhibitor, p21. The p21 protein binds to multiple cyclin-Cdk complexes and blocks the kinase activity. This inhibits the phosphorylation of proteins required for the various stages of the cell cycle. The binding of p21 to the G1 cyclin-Cdk complexes is central to the cessation of the G1 phase that follows DNA damage by radiation. This allows time for the DNA repair mechanisms to correct the damage. Another function of p21 is to bind proliferating cell nuclear antigen (PCNA). PCNA is a cofactor required for full expression of DNA polymerase-delta. DNA replication is inhibited when p21 is bound to PCNA. The roles that p53 and p21 play in damage control in cells are illustrated in Figure 6.

Mutations that result in the loss or alteration of p53 activity result in cancer development. Abnormal p53 levels are associated with the loss of the cell's ability to halt the progression of the cell cycle under the aforementioned adverse conditions. Therefore, the cell continues to proliferate, resulting in a defective phenotype.

DNA REPLICATION

As described earlier, the replication of DNA occurs during the S phase of the cell. The S phase occupies approximately 30% of the cell cycle time. The replication of DNA is a semi-

TABLE 2
Structural and functional properties of human DNA polymerases

DNA polymerase	Size, catalytic subunit (kD)	Location	Function in the cell
Alpha	160–185	Nucleus	Lagging strand replication
Beta	40	Nucleus	DNA repair
Gamma	125	Mitochondria	Replication of mitochondrial DNA
Delta	125	Nucleus	Leading and lagging strand replication
Epsilon	210–230	Nucleus	DNA repair

conservative process, wherein each parental strand of the DNA helix serves as a template for the synthesis of a new and complementary daughter strand. In human diploid cells, this involves the replication of six billion base pairs of DNA. The reader may consult references 14-19 for further details about DNA replication.

A diverse array of enzymes and proteins are important in the process of DNA replication. The key enzyme involved is DNA polymerase, which catalyzes the ligation of the deoxyribonucleoside 5'-triphosphates (dNTPs) to generate the growing DNA chain. Eukaryotic cells contain five types of DNA polymerases: alpha, beta, gamma, delta and epsilon. The properties of the various human DNA polymerases are described in Table 2. The DNA polymerase-gamma is restricted to the mitochondria where it is responsible for mitochondrial DNA replication. The other four DNA polymerases are localized in the nucleus. DNA polymerase-delta is the major replicating enzyme in human cells.

The process of DNA replication on each chromosome is initiated at designated positions, referred to as origins of replication (ori). Each human chromosome has multiple ori placed at every 150 to 200 kilobase pairs. There are approximately 30,000 initiation sites in the entire human genome. Thus, multiple sections of the genome are replicated simultaneously. Each small replicating unit is called a replicon and has its own ori site where DNA synthesis is initiated. The process of DNA replication proceeds bidirectionally on the chromosome until each replicon comes in contact with the next one. Thus, an entire chromosome can be replicated completely during the S phase of the cell cycle.

As the two parent DNA strands unwind and separate, DNA replication begins at ori and proceeds down the two DNA strands (Figure 7). Because of the inherent properties of DNA polymerase, daughter strand synthesis can only proceed from the ori in the 5' to 3' direction. Thus, one strand is synthesized in a 5' to 3' direction and the opposite strand is also synthesized in the 5' to 3' direction. Because there is no DNA polymerase that can synthesize DNA in a 3' to 5' direction, a DNA strand cannot be used as a template in the 5' to 3' direction. Thus, short fragments of DNA, called Okazaki fragments, use the 3' to 5' strand as a template. The Okazaki

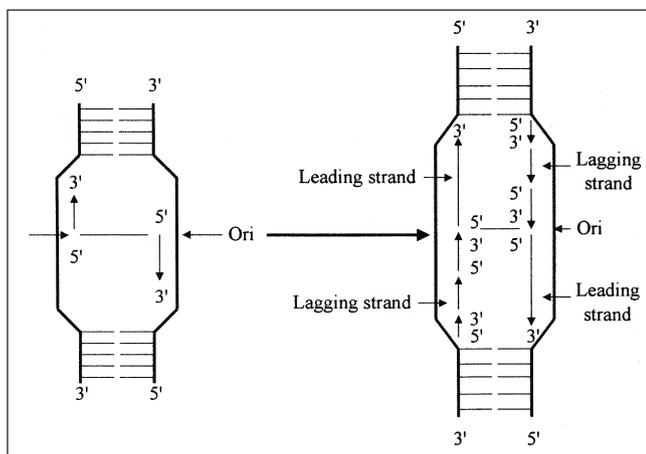


Figure 7) Replicon. DNA polymerase can only synthesize DNA in a 5' to 3' direction. For both strands of the DNA helix to serve as templates, one strand (ie, the leading strand) is synthesized continuously in a 5' to 3' direction, while the other strand (ie, the lagging strand) is synthesized discontinuously in short fragments but still in the 5' to 3' direction. The short DNA fragments (ie, Okazaki fragments), are subsequently joined together by DNA ligase. Ori Origins of replication

TABLE 3
Proteins involved in DNA replication

Protein	Function
DNA helicase	Unwinds DNA and breaks hydrogen bonds
Single-stranded DNA-binding protein (RPA)	Binds single-stranded DNA to prevent hydrogen bonding
Proliferating cell nuclear antigen	Stimulates DNA polymerase-delta activity
DNA polymerase-delta	Stimulates leading and lagging strand DNA replication and 3' to 5' exonuclease proofreading
DNA polymerase-alpha/DNA primase complex	Promotes synthesis of RNA primers and lagging strand synthesis
DNA ligase	Seals 3' terminal hydroxyl and 5' terminal phosphate groups of adjacent nucleotides in DNA
Ribonuclease H1	Removes RNA from RNA-DNA hybrid
DNA topoisomerase	Relaxes DNA by breaking and resealing phosphodiester bonds

RPA Replication protein A

zaki fragments are approximately 200 nucleotides in length and are synthesized in a 5' to 3' direction. The resulting fragments are then joined by an enzyme called DNA ligase to give one continuous DNA strand. The DNA strand that is synthesized continuously in the 5' to 3' direction is called the 'leading strand of DNA synthesis' because it starts at a fixed point and dictates DNA synthesis. The strand of DNA that is synthesized in the 5' to 3' direction in short pieces (ie, discontinuously) is called the 'lagging strand of DNA synthesis'.

The replication fork is the part of the DNA molecule that is being replicated at a given time and is the region between the unreplicated segment of the DNA molecule and a newly replicated portion of DNA. Because DNA is synthesized

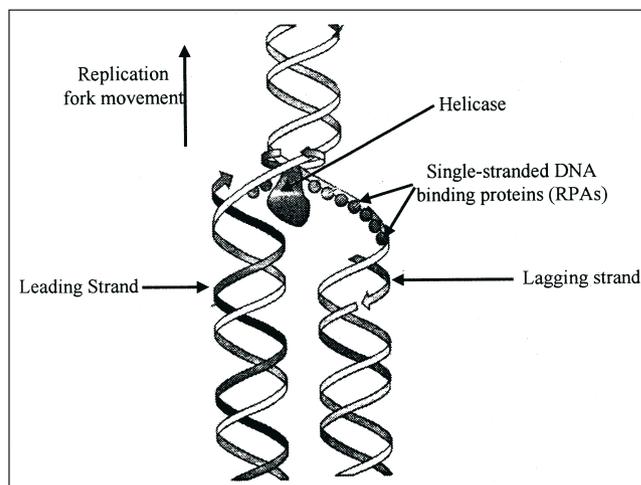


Figure 8) Replication of a DNA molecule illustrating the interaction of the helicase and DNA binding proteins at the replication fork. RPA Replication protein A

bidirectionally, each replicon contains two replication forks. A specific initiator protein has the ability to recognize the origin sequence and signals the initiation of DNA synthesis. It has been hypothesized that this initiator protein binds the ori sequence and attracts the DNA replicating complex to this particular site on the DNA molecule.

All DNA polymerases must have a primer (ie, a free 3' hydroxyl end of a polynucleotide). The primer in DNA replication is not DNA, but rather is a small segment of RNA measuring five to 10 nucleotides in length that is synthesized by the enzyme DNA primase. DNA primase initiates the synthesis of an RNA molecule at the ori, and DNA polymerase uses this RNA primer to add deoxyribonucleotides to the 3' hydroxyl group of the RNA, and synthesizes a new DNA strand that is complementary to the template strand. After completion of DNA synthesis, the RNA molecule is removed from the DNA helix, and the resulting gap in the DNA is filled by a DNA polymerase.

The various proteins that play an important role in the process of DNA replication are listed in Table 3. The separation of the two strands of DNA is catalyzed by an enzyme called DNA helicase, which breaks the hydrogen bonds holding the DNA strands together. The DNA helix is subsequently unwound, and strands remain separated through the action of a protein called replication protein A (RPA). RPA is a single-stranded DNA binding protein (Figure 8). The DNA helicase acts at the edge of the replication fork, opening and unwinding the DNA as replication proceeds along the DNA molecule. As the helicase unwinds the DNA at the replication fork, the DNA helix downstream becomes tightly wound and supercoiled. The tension on the DNA molecule is released by the action of DNA topoisomerase, which breaks phosphodiester bonds, unwinds the downstream DNA helix and then reseals it by forming new phosphodiester bonds. Both DNA helicases and DNA topoisomerases play a pivotal role in the process of DNA replication and transcription.

The DNA polymerases catalyze the formation of phosphodiester bonds between the adjacent deoxyribonucleotides in the DNA molecule. All DNA polymerases catalyze the synthesis of DNA only in the 5' to 3' direction. DNA polymerase-delta is the major replicating protein in human cells, and is involved in both leading and lagging strand replication. DNA polymerase-alpha is complexed with another protein, the DNA primase. Together, these proteins are involved in the replication of the lagging strand. DNA primase makes the small RNA primers with DNA polymerase-alpha. Deoxyribonucleotides are added to the 3' terminal of the primer for a short distance of about 30 nucleotides. The DNA polymerase-alpha/DNA primase complex subsequently falls off the DNA molecule and is replaced with DNA polymerase-delta, which continues the synthesis of the growing DNA chain. The RNA primers used by DNA polymerases must be removed from the DNA molecule. This is accomplished by the action of the enzyme RNase H1, which specifically degrades RNA present in a DNA/RNA hybrid. DNA polymerase later completes the DNA synthesis of the lagging strand by filling in the gap. Then the ligation of the 3' hydroxyl terminus of the DNA of one Okazaki fragment with the 5' terminal phosphate of DNA of the adjacent fragment occurs through the formation of a phosphodiester bond. This reaction is catalyzed by DNA ligase.

DNA polymerase-beta and -epsilon serve in the process of DNA repair and are not directly involved in replicating the entire genome. Finally, DNA polymerase-gamma is responsible for replicating the circular double-stranded DNA found in mitochondria.

An additional protein involved in the replication of DNA in human cells is PCNA. PCNA forms part of the DNA polymerase-delta complex and stimulates the activity in the DNA polymerase. The interactions of various proteins involved in DNA synthesis in the lagging strand are depicted in the model shown in Figure 9.

Some DNA polymerases (eg, DNA polymerase-delta) have intrinsic 3' to 5' exonuclease activity that removes bases sequentially from the end of the DNA molecule (ie, the 3' end). This nuclease activity plays a critical role in preventing mistakes in base pairing during DNA replication. For example, if a C on the new DNA strands binds to an A on the template strand, subsequent replications of this mistake result in a G-C base pair molecule instead of an A-T base pair. Substitution of one base pair with another leads to a mutation in the DNA molecule that may affect cellular function. The 3' to 5' exonuclease recognizes these mispairs as soon as they occur and removes the newly inserted albeit incorrect base. The DNA polymerase then inserts the proper base into the growing DNA chain. This exonuclease component of DNA polymerase is termed the 'proofreading function'.

As mentioned above, the ends (ie, telomeres) of all chromosomes maintain the overall integrity of chromosomes. Telomeres consist of randomly repeated base sequences, TTAGGG, which are repeated 100 to 1000 times. Because DNA polymerases function only in the 5' to 3' direction,

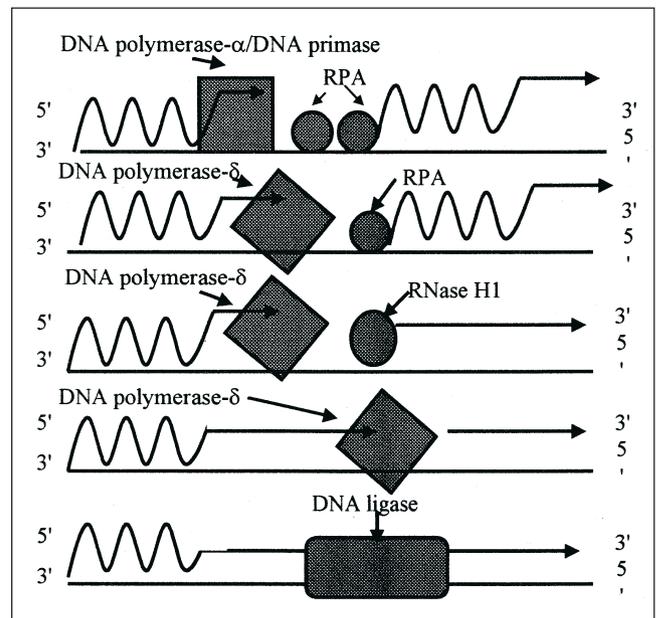


Figure 9) Model for DNA replication in human cells. Replication protein A (RPA), a single-stranded DNA-binding protein, separates the DNA strands to allow the DNA polymerase-alpha/DNA primase complex to bind to the DNA and initiate the synthesis of an RNA primer (indicated by the wavy line). DNA polymerase-alpha adds approximately 30 deoxyribonucleotides to the 3' end of the RNA primer. The DNA polymerase-delta displaces the RNA polymerase-alpha/DNA primase complex and extends the DNA strand by adding deoxyribonucleotides to the 3' end of the newly synthesized DNA strand. Upon completion of the DNA synthesis, RNase H1 removes the RNA primer. The DNA polymerase-delta fills in the gap using opposite DNA strands as the template. Finally, the two Okazaki fragments are joined together. This reaction is catalyzed by DNA ligase

they are unable to copy the extreme 5' ends of linear DNA molecules. These sequences (ie, telomeres) are replicated by the action of the enzyme telomerase, which is a reverse transcriptase. Reverse transcriptases synthesize DNA from an RNA template. Telomerases carry their own template RNA complementary to the telomere repeat sequences. The RNA template allows telomerase to generate multiple copies of the telomeric repeat sequences, thus maintaining telomeres in the absence of a conventional DNA template to direct their synthesis.

Despite the accuracy of DNA replication, cellular genomes are far from static. Gene rearrangements and mutations are required to maintain genetic diversity among individuals. To this end, recombination between homologous chromosomes occurs during meiosis and allows parental genes to be rearranged in new combinations in the next generation of cells. The rearrangements of DNA sequences within the genome create novel combinations of genetic information. In some instances, DNA rearrangements are programmed to regulate gene expression during the cellular processes of differentiation and development. A striking example of this is the rearrangement of antibody genes during the development of the immune system. A key feature of both immunoglobulins and T cell receptors is their enormous diversity. This diversity allows different antibody or T

TABLE 4
Classes of mutations found in human DNA

Class	Result
Single base pair substitutions (point mutations)	
Altered structure of gene product	
Missense mutation	Single amino acid replacement in the protein
Nonsense mutation	Termination codon in the middle of the gene results in premature termination of protein synthesis
RNA splicing mutation	Protein may be missing part or all of an exon sequence
Altered quantity of gene product	
Mutations in regulatory sequences	Transcription of the gene is altered, which can reduce or eliminate the gene product
Mutations in RNA processing and translation	Stability of mRNA is altered, which may reduce the amount of gene product
Insertions or deletions	
One or two base pairs (frameshift mutations)	Addition or deletion of one or two base pairs can affect the reading frame of the gene, resulting in a grossly altered or absent gene product
Large number of base pairs	Large pieces of the DNA may be lost or large segment of DNA may insert into the middle of a gene, resulting in loss of function
Expansion of trinucleotide repeat sequences	Unstable trinucleotide repeats can suddenly expand in number, resulting in the alteration of production or structure of a particular gene product
Chromosomal alteration	Inversions, translocations, duplications or gene amplification may result

cell receptor molecules to recognize a variable array of foreign antigens. These diverse antibodies and T cell receptors are encoded by unique lymphocyte genes that are formed during the development of the immune system as a result of site-specific recombination between distinct segments of immunoglobulin and T cell receptor genes.

MUTATIONS AND DNA REPAIR MECHANISMS

Mutations are the result of permanent changes in the base sequence of the DNA molecule and are central to the pathogenesis of all human genetic diseases. The various classes of mutations that occur in DNA molecules are listed in Table 4. The reader may wish to consult references 20-26 for further details. Many of the concepts concerning the different types of mutations that occur in DNA, and the potential mechanisms associated with the production of these mutations, were originally developed in bacterial cell model systems. Recently, the knowledge base has expanded in the area of the molecular basis of mutations in eukaryotic cells. Studies of diseased human cells have established common

mechanisms by which DNA undergoes mutation. More importantly, DNA repair mechanisms have been defined.

Many of the mutations that occur in DNA are the result of single base pair substitutions in which one base pair (eg, an A-T pair) is replaced with a second base pair (eg, a G-C pair). The substitution of one base pair with a second base pair elicits a change of codon that can lead either to a missense mutation (where one amino acid replaces another amino acid in a protein) or to a nonsense mutation (where one of the terminator codons appears in the middle of a gene). With a nonsense mutation, there is no transfer of an RNA molecule to recognize these codons, and protein synthesis terminates at the site of the nonsense codon. This leads to the production of a truncated polypeptide.

A mutation that alters the splice acceptor or splice donor sequences can result in apparent splicing of an RNA transcript. This leads to the production of an mRNA that may be missing a substantial part of a particular exon and thus codes for a mutant protein. Other base pair substitutions can occur in regulatory sequences required for the binding of transcription factors or RNA polymerase. In this instance, the quantity of the product produced by the gene that is controlled by these sequences is dramatically altered. In the extreme case, base pair substitutions can lead to a complete absence of the gene product or to a dramatic increase in the amount of a particular gene product.

Frameshift mutations are caused by the addition or deletion of one or two base pairs within the coding sequence of a gene. This alters the reading frame of the mRNA. Thus, the mRNA is translated out of frame from the site of the insertion or deletion of the base pair. This results in the production of a protein that is altered in its amino acid sequence, starting from the point of the insertion or deletion of the base pair and continuing to the end of the protein. Often, the altered reading frame also leads to the production of a termination codon in the middle of the gene. This results in premature cessation of protein synthesis.

The insertion and deletion of many base pairs can also occur with DNA molecules. Deletion mutations can occur in a chromosome with the loss of hundreds to thousands of base pairs from the DNA, and the deleted genetic material is permanently lost. Large insertions of DNA sequences have been described and are caused by transposon-like elements, often repetitive DNA sequences such as LINE repeats.

In summary, the possible changes in DNA that give rise to mutations may be illustrated by considering the following literary masterpiece.

Wild type: The cat sat on the mat.

Substitution: The rat sat on the mat.

Insertion (single): The cat shat on the mat.

Insertion (multiple): The cattle sat on the mat.

Deletion (single): The c.t sat on the mat.

Deletion (multiple): The cat the mat.

Inversion (small): The tac sat on the mat.

Inversion (large): Tam eht no tas tac eht.

DNA polymerases catalyze the proper pairing of A to T and G to C with very high accuracy. However, mispairing occurs at a frequency of approximately 10^{-5} bases. For example, an A-C pair forms instead of an A-T pair. If such a mispair remains in the DNA molecule, the initial A-T pair that has become an A-C pair gives rise to a G-C pair during the next replication cycle. To keep the mutation rate at a low level, eukaryotic cells have devised mechanisms for correcting base mispairs before they become a permanent feature of the DNA.

Bases that are present in DNA molecules can undergo spontaneous damage or modification. One frequent form of modification occurs with the purine bases A and G. Purine residues may be lost from the DNA molecules by a process called depurination. The glycosidic bond between the deoxyribose and the base is hydrolyzed, which leads to a gap in one of the DNA strands. This damage must be corrected before the DNA is replicated, otherwise a mutation ensues. The bases C, A and G are capable of undergoing spontaneous deamination, wherein the base loses an amino group and its structure is changed. For example, when C is deaminated it becomes U. This leads to the presence of U in DNA instead of C. U appears with an A residue during the next replication cycle. The original G-C pair, which after deamination is now a G-U pair, subsequently becomes an A-T pair. Finally, ultraviolet rays from sunlight are common mutagenic agents that cause bond formation between adjacent pyrimidines on the same DNA strand. The most frequent type of pyrimidine dimer is the T-T dimer. The presence of a T-T dimer in the DNA molecule blocks DNA replication and leads to death of the cell if it is not removed. The 3' to 5' exonuclease activity associated with DNA polymerase- δ and - ϵ is responsible for cleaving mispaired nucleotides from the 3' end of newly replicated DNA strands. This allows the polymerase a second opportunity to add the correct base. The entire process is known as the proofreading function.

If base mispairing remains in the DNA, it leads to a mutation at the next DNA replication cycle. However, eukaryotic cells have evolved a mechanism to deal specifically with persistent base mispairing immediately after replication. Human cells have a methyl-directed mismatch repair system that appears to be similar to that of the bacterial strains. The methyl-directed mismatch repair systems scan the DNA molecule, and when base mispairs as well as insertions and deletions are detected, correction of the error occurs on the nonmethylated, newly synthesized DNA strand. This allows the repair system to correct the nascent strand that has a normal base in the wrong location and prevents the mispaired bases from giving rise to a permanent mutation.

DNA molecules are methylated at specific sites, either on an A or a C residue. In human cells, C residues located in CpG islands are methylated. Methylation is a postreplication event. During the initial period of DNA replication, one strand (ie, the template strand) is methylated, while the newly synthesized DNA strand is not methylated.

Mutator (Mut) proteins are involved in methyl-directed mismatch repair. Human homologues have been identified for MutS (hMSH2 and GTBP) and MutL (hMLH1 and hPMS2), but there are no known homologues for MutH. Methyl-directed mismatch repair appears to be similar in bacteria and humans. In human cells, mismatches are recognized by the protein hMSH2 or a dimer composed of hMSH2 and GTBP. Base mispairing creates a bulge in the DNA that is recognized and bound by the MutS protein. The MutS protein that is bound to the mismatch recruits the MutL homologue to the site. MutH cleaves the nonmethylated DNA strand. This is followed by the stepwise removal of nucleotides by an exonuclease, and the resulting gap in the DNA molecule is repaired by DNA polymerase using the base sequence in the template strand. The final phosphodiester bond is sealed by DNA ligase.

One of the most common hereditary cancers, hereditary nonpolyposis colon cancer (HNPCC), arises from mutations in the methyl-directed mismatch repair system. HNPCC affects one in 200 people in North America and accounts for approximately 15% of all colon cancers. There are at least five genetic loci involved in the human mismatch repair process. These include hMSH2, hMLH1, hPMS1 and hPMS2, and the GTBP gene. Cells with HNPCC are characterized by microsatellite instability. Microsatellites are repetitive nucleotide sequences (di-, tri- or tetranucleotides) located throughout the human genome. The presence of these repeats in the DNA presents a 'road block' to the DNA polymerase molecule during DNA replication. When DNA polymerase is confronted with a long repetitive sequence of DNA, it produces a strand of DNA with extra bases that are not base paired with the template and that loop away from the DNA helix. The mismatch repair system recognizes these loops as defective and removes them. The loops remain if the repair system is defective. Microsatellite instability signals that the cell has developed a Mut phenotype and has an increased rate of overall mutation. These cells also develop mutations in such genes as the *p53* gene or other tumour suppressor genes at a much higher rate than do normal cells.

Another type of DNA mutation is incurred through damage of bases of a DNA molecule that is not undergoing replication. Cells have evolved two major repair systems to deal with this type of DNA damage. The first system is called base excision repair. When a U residue occurs in a DNA molecule, it is recognized by U-DNA glycosylase and is removed from the DNA, leaving behind a gap. The lack of a base in the DNA helix is recognized by specific endonucleases known as apurinic/apyrimidinic (AP) endonucleases. The AP endonuclease cleaves the DNA at the site of the missing base. The resulting gap is repaired by DNA polymerase using the base present in the complementary strand as a template. This is followed by ligation via DNA ligase. If the U residue is not removed, it eventually results in a G-U mismatch, and the original G-C pair becomes an A-T pair or a mutation. A more general repair mechanism is known as nucleotide excision repair, which repairs bulky distortions in the DNA molecule. The overall scheme for nucleotide excision repair

resembles that of base excision repair and methyl-directed mismatch repair. All systems have specific proteins that recognize the damaged area of DNA, as well as specific proteins involved in the removal of the damage from the DNA. Following removal of the damage, the gap is filled by repair synthesis, catalyzed by DNA polymerase and sealed by DNA ligase.

Xeroderma pigmentosum (XP) is a rare autosomal recessive disorder characterized by skin neoplasms. Skin cells from XP patients are unable to repair DNA damage caused by exposure to ultraviolet light. Ultraviolet light damages DNA and results in the formation of dimers between adjacent pyrimidines on the same DNA strand (eg, T-T dimer). These T-T dimers distort the DNA helix, and result in the cessation of replication and transcription at that point until the dimer is removed. The nucleotide excision repair system removes these T-T dimers. The initial step is the recognition of the damage by the XPA protein, which binds along with XPF-ERCC1 protein and the single-stranded DNA binding protein, RPA. Helicase activity unwinds the helix and stimulates the excision activity of two endonucleases, XPF and XPG, which cut the DNA. This creates a large gap in the DNA molecule, and the 3' hydroxyl is recognized by DNA polymerase-delta or -epsilon, which carries out repair synthesis using the undamaged DNA strand as a template. The final nick is sealed by DNA ligase.

A new type of mutation that results in a number of human genetic diseases has been recently described. These mutations are the result of the expansion of trinucleotide repeats (CAG, CTG, CGG or GAA) found throughout the human genome. Long runs of these repeat triplets are found in exons at the 5' or 3' end of genes. Individuals affected with one of the expansion disorder diseases have an increase in the number of copies of the trinucleotide repeats. Expansion of the repeat sequences can alter either the structure or func-

tion of a particular protein. One of the best characterized examples of this is the trinucleotide CAG, which codes for the amino acid glutamine. In Huntington's disease the CAG repeat is located in the coding region of the first exon at the 5' end of the gene. These repeats are translated and appear as a long stretch of glutamines within the structure of the protein such that the mutant protein has a range of 40 to 100 glutamines at that particular site. All of the CAG repeat diseases are autosomal dominant disorders that are characterized by late onset neuronal loss.

CONCLUSIONS

The fundamental similarities among different types of cells constitute a unifying theme in cell and molecular biology. The basic principles derived from experiments with prokaryotic cells, coupled with the availability of a variety of experimental tools, provided the framework to define the molecular processes that determine the flow of genetic information in eukaryotic cells. The importance of DNA in providing a blueprint that coordinates all cellular activities is underscored in the present review. The subsequent reviews will examine the intricate array of cellular processes responsible for the orderly transcription and translation of genetic information into proteins – the major determinants of eukaryotic gene expression.

ACKNOWLEDGEMENTS: This work was supported by operating grants from the Medical Research Council of Canada and the Crohn's and Colitis Foundation of Canada. Dr Gary E Wild is a chercheur boursier clinicien of Les Fonds de la Recherche en Sante du Québec. Dr Wild extends his appreciation to Drs David Fromson, John Southin, Howard Bussey and Bruce Brandhorst of the McGill Biology Department. Their tireless efforts in the area of undergraduate science education fostered a sense of inquiry and collegiality that guided a cohort of students through the early recombinant DNA era.

REFERENCES

- Lodish H, Baltimore D, Berk A, Zipursky SL, Matsudaira P, Darnell J. *Molecular Cell Biology*, 3rd edn. New York: WH Freeman, 1995.
- Alberts B, Bray D, Johnson A, et al. *Essential Cell Biology – An Introduction to the Molecular Biology of the Cell*. New York: Garland Publishing, 1998.
- Cooper GM. *The Cell – A Molecular Approach*. Washington: ASM Press, 1997.
- Lewin B. *Genes VI*. New York: Oxford University Press, 1997.
- Glick BR, Pasternak JL. *Molecular Biotechnology – Principles and Applications of Recombinant DNA*, 2nd edn. Washington: ASM Press, 1998.
- Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. *Molecular Biology of the Cell*, 3rd edn. New York: Garland Publishing, 1994.
- Jameson JL. *Principles of Molecular Medicine*. New Jersey: Humana Press, 1998.
- Strachan T, Read AP. *Human Molecular Genetics*. New York: Wiley-Liss, 1996.
- Hartwell LH, Kastan MB. Checkpoints: Controls that ensure the order of cell cycle events. *Science* 1989;246:629-34.
- Murray AW. Creative blocks: Cell-cycle checkpoints and feedback controls. *Nature* 1992;359:599-604.
- Norbury C, Nurse P. Animal cell cycles and their control. *Ann Rev Biochem* 1992;61:441-70.
- Morgan DO. Principles of CDK regulation. *Nature* 1995;374:131-4.
- Levine AJ. The tumor suppressor genes. *Ann Rev Biochem* 1993;62:623-51.
- Blackburn EH. Telomeres. *Ann Rev Biochem* 1992;61:113-29.
- Diller DJ, Raghuraman MK. Eukaryotic replication origins: Control in space and time. *Trends Biochem Sci* 1994;19:320-5.
- Heintz NH, Dailey L, Held P, Heintz N. Eukaryotic replication origins and promoters of bidirectional DNA synthesis. *Trends Genet* 1992;8:376-81.
- Kelman Z, O'Donnell M. DNA polymerase III holoenzyme: Structure and function of a chromosomal replicating machine. *Ann Rev Biochem* 1995;64:171-200.
- Roca J. The mechanism of DNA topoisomerases. *Trends Biochem Sci* 1995;20:156-60.
- Zakian VA. Telomeres. Beginning to understand the end. *Science* 1995;270:1601-7.
- Kolodner RD. Mismatch repair: Mechanisms and relationships to cancer susceptibility. *Trends Biochem Sci* 1995;20:397-401.
- Leach FS. Mutations of *mutS* homolog in hereditary nonpolyposis colorectal cancer. *Cell* 1993;75:1215-25.
- Modrich P. Mismatch repair, genetic stability, and cancer. *Science* 1994;266:1959-60.
- Sancar A. Mechanisms of DNA excision repair. *Science* 1994;266:1954-6.
- Seeberg E, Eide L, Bjoras M. The base excision repair pathway. *Trends Biochem Sci* 1995;20:391-7.
- Tanaka K, Wood RD. Xeroderma pigmentosum and nucleotide excision repair of DNA. *Trends Biochem Sci* 1994;19:83-6.
- Davis MM. T cell receptor gene diversity and selection. *Ann Rev Biochem* 1990;59:475-96.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

