2

BASICS OF MOLECULAR BIOLOGY, GENETIC ENGINEERING, AND METABOLIC ENGINEERING

Michael X. Wang[1] and Huidong Shi[1,2]

[1]Department of Pathology and Anatomical Sciences, Ellis Fischel Cancer Center, University of Missouri School of Medicine, M263 Medical Science Building, One Hospital Drive, Columbia, Missouri 65203

[2]Department of Biochemistry and Molecular Biology, Medical College of Georgia, MCG Cancer Center CN4130, 1120 15th Street, Augusta, Georgia 30912

2.1 BIOMOLECULES IN LIVING CELLS

Carbon, hydrogen, oxygen, and nitrogen are the most abundant elements in living organisms. Carbon can covalently bond to hydrogen, oxygen, and nitrogen to form biomolecules. Small biomolecules can combine to form more complex macromolecules such as nucleic acids, proteins, and carbohydrates. All living cells are built with these biomolecules.

2.1.1 Nucleic Acids

Nucleic acids carry the genetic information in the cell. The major types of nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Both of them are polymers of nucleotides.

Systems Biology and Synthetic Biology Edited by Pengcheng Fu and Sven Panke
Copyright © 2009 John Wiley & Sons, Inc.

2.1.1.1 Nucleotides Nucleotides are building blocks of nucleic acids. Nucleotides have three characteristic structural components: (1) base, (2) pentose, and (3) phosphate. The bases are derivatives of two parental compounds, purine and pyrimidine. The two major purine bases are adenine (A) and guanine (G), and three major pyrimidines are cytosine (C), thymine (T), and uracil (U). Two types of pentose are 2′-deoxy-D-ribose and D-ribose. Deoxyribonucleotides (deoxyribonucleoside 5′-monophosphate), the structural units of DNA, contain 2′-deoxy-D-ribose. Ribonucleotides (ribonucleoside 5′-monophosphates), the structural units of RNA, contain D-ribose. The phosphate group gives the nucleic acid a negative charge property.

2.1.1.2 DNA In 1953, Watson and Crick postulated a three-dimensional model of the DNA molecule based on the available data at the time. It consists of two helical polynucleotide strands twisted around the same axis to form a right-handed double helix structure. The hydrophilic backbones of deoxyribose and phosphate groups are outside the double helix, whereas purine and pyrimidine bases are stacked inside the double helix. Each purine base of one strand is paired in the same plane with a pyrimidine base of the other strand by hydrogen bonds. There are three hydrogen bonds between G and C and only two between A and T. As a result, the two antiparallel strands are not identical but complementary to each other (Fig. 2-1a and b).

The double helix strands of DNA can be separated from each other (denatured or melting) by heating or at extremes of pH *in vitro*. The temperature at which 50 percent of the double-stranded DNA molecules separate into single strand is the melting temperature (T_m). DNAs rich in G/C pairs have higher melting points than DNAs rich in A/T pairs. The T_m can be calculated according to G/C content of a given DNA fragment. On the contrary, denatured single-stranded DNAs can anneal to form a double helix (renaturation or hybridization). High G/C content, decreasing temperature, increasing the ion concentration, or neutralizing the pH are favorable to DNA renaturation. The nucleotide sequences of DNA can be determined. The human genome and many other genomes of organisms have been successfully sequenced. These sequences are available in the public database (http://www.ncbi.nlm.nih.gov/). DNA can also be synthesized with simple, automated protocols involving chemical and enzymatic methods such as polymerase chain reaction (PCR).

The biological significances of the double-stranded helical structure of DNA are threefold. It stores genetic information in a form of linear nucleotide sequence with chemically stable features; it allows the genetic information to be passed on to the next generation of cells by semiconservative DNA replication with very high fidelity during cell division; and it acts as the template to transfer genetic information into messenger RNAs (mRNAs) and then amino acid sequences of proteins.

2.1.1.3 RNA RNA is usually single-stranded polynucleotides. The chemical compositions of RNA differ from DNA in two ways: (1) sugar-phosphate backbone contains D-ribose rather than 2′-deoxyribose in DNA, and (2) nucleotide base thymine (T) in DNA is replaced in RNA with uracil (U), which is paired with

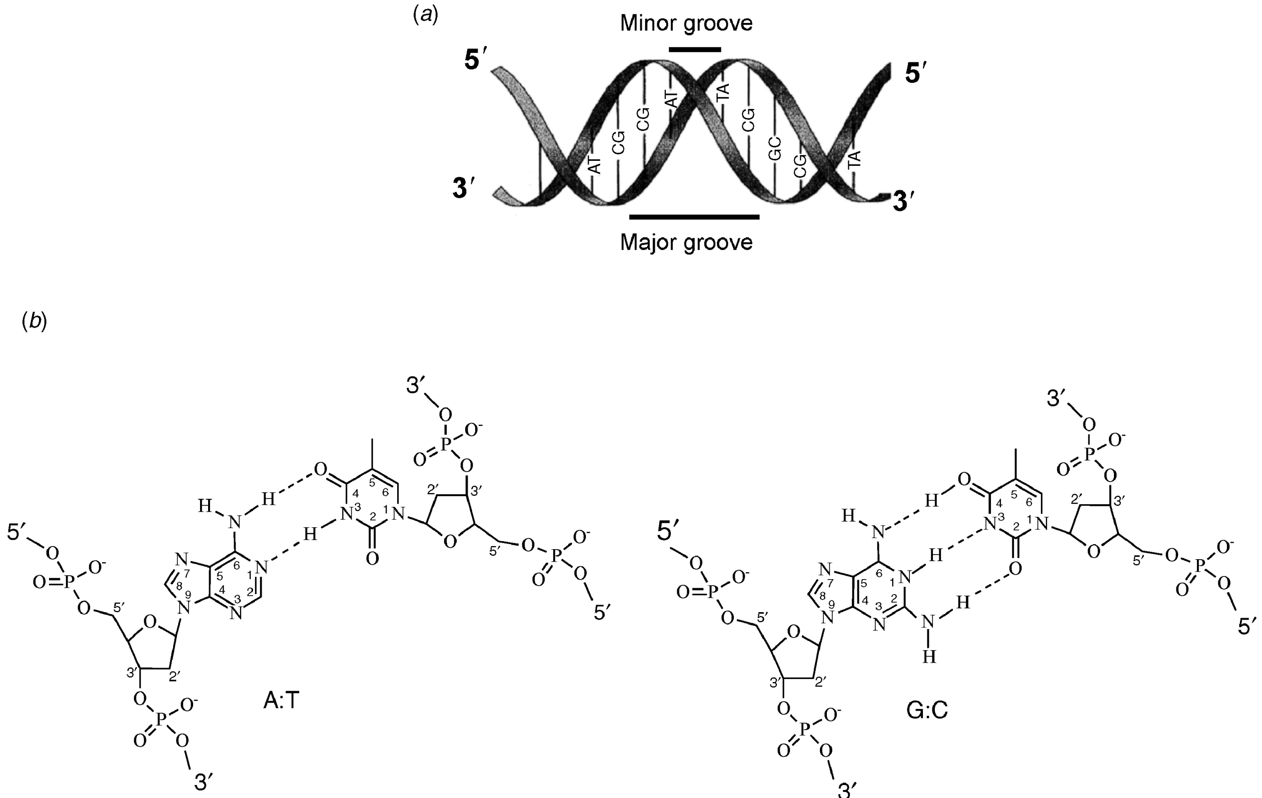

Figure 2-1 (a) Chemical structure of DNA double helix and base pairs defined by Watson and Crick. Major and minor grooves on the surface of double helix are indicated. (b) Formation of hydrogen bonds between A and T and C and G bases. The antiparallel complementary strands of double helix structure of DNA are held by hydrogen bonds.

Table 2-1 Types and functions of RNAs

Type of RNA	Function
mRNA (messenger RNA)	RNA that functions as the intermediary (transcript) between DNA in the nucleus and protein production in the cytoplasm
tRNA (transfer RNA)	RNA that transfers an amino acid to a growing polypeptide chain during translation
rRNA (ribosomal RNA)	RNA that is a component of ribosomes for mRNA processing
Ribozyme RNA	Functions as an enzyme by catalyzing chemical reactions in the cell
Small RNAs	Include miRNA and siRNA. Functions as translational repression and RNA degradation (RNAi)

adenine (A). RNA can form complex three-dimensional structures by intramolecular base pairing.

There are five types of RNAs with distinct biological functions (Table 2-1). They are mRNA, transfer RNA (tRNA), ribosomal RNA (rRNA), ribozyme RNA, and small RNAs (micro RNA (miRNA) and short interfering RNA (siRNA)). RNAs are cell specific. All cells have identical DNA content in a given organism; however, RNA levels and types differ in different cell types. RNA can be synthesized from a DNA template (transcription). In contrast, RNA can also function as a template to synthesize DNA (complementary or cDNA) by a reverse transcriptase. RNA is chemically unstable compared with DNA and is degraded easily *in vitro*.

2.1.2 Proteins

There are thousands of different proteins that perform the bulk of cellular activities. Proteins are the polymers of amino acids.

2.1.2.1 Amino Acids Amino acids are the building blocks of proteins. There are a total of 20 amino acids in cells. Amino acids share a common chemical structure. They have a carboxyl group and an amino group bonded to the same carbon atom (α-carbon). But they differ from each other in their side chains or R-groups, which vary in structure, size, and electric charge (Table 2-2). Each amino acid has its own chemical features determined by the R-group. For example, cysteine ($R-CH_2-SH$) is readily oxidized to form a covalently linked dimeric amino acid called cystine by forming a disulfide bond. Disulfide bonds play a special role in the structures of many proteins either in the same polypeptide (intra) or between two different polypeptide chains (inter). Both tryptophan and tyrosine have similar light absorption spectra with the maximal wavelength of 280 nm, which gives the spectroscopic properties of proteins. Amino acids have been assigned three-letter and one-letter abbreviations (Table 2-2).

2.1.2.2 Protein Structure Protein is a polypeptide in which amino acids are linked by a peptide bond. The linkage is formed by removing a water molecule (dehydration) from the α-carboxyl group of one amino acid and the α-amino group of

Table 2-2 Amino acids, abbreviations, and the R-groups

Amino Acids	Three Letter	Single Letter	R-Group, Basic Structure:
Alanine	Ala	A	$CH_3-CH(NH_2)-COOH$
Arginine	Arg	R	$HN{=}C(NH_2)-NH-(CH_2)_3-CH(NH_2)-COOH$
Asparagine	Asn	N	$H_2N-CO-CH_2-CH(NH_2)-COOH$
Aspartic acid	Asp	D	$HOOC-CH_2-CH(NH_2)-COOH$
Cysteine	Cys	C	$HS-CH_2-CH(NH_2)-COOH$
Glutamine	Gln	Q	$H_2N-CO-(CH_2)_2-CH(NH_2)-COOH$
Glutamic acid	Glu	E	$HOOC-(CH_2)_2-CH(NH_2)-COOH$
Glycine	Gly	G	NH_2-CH_2-COOH
Histidine	His	H	$NH-CH{=}N-CH{=}C-CH_2-CH(NH_2)-COOH$
Isoleucine	Ile	I	$CH_3-CH_2-CH(CH_3)-CH(NH_2)-COOH$
Leucine	Leu	L	$(CH_3)_2-CH-CH_2-CH(NH_2)-COOH$
Lysine	Lys	K	$H_2N-(CH_2)_4-CH(NH_2)-COOH$
Methionine	Met	M	$CH_3-S-(CH_2)_2-CH(NH_2)-COOH$
Phenylalanine	Phe	F	$Ph-CH_2-CH(NH_2)-COOH$
Proline	Pro	P	$NH-(CH_2)_3-CH-COOH$
Serine	Ser	S	$HO-CH_2-CH(NH_2)-COOH$
Threonine	Thr	T	$CH_3-CH(OH)-CH(NH_2)-COOH$
Tryptophan	Trp	W	$Ph-NH-CH{=}C-CH_2-CH(NH_2)-COOH$
Tyrosine	Tyr	Y	$HO-p-Ph-CH_2-CH(NH_2)-COOH$
Valine	Val	V	$(CH_3)_2-CH-CH(NH_2)-COOH$

Note: All 20 amino acids share a common chemical structure. They have a carboxyl group and an amino group bonded to the same carbon atom (α-carbon), but differ from each other in their side chains or R-groups. The R-groups vary in structure, size, and electric charge. The chemical features of each amino acid are determined by its R-group.

another amino acid. The polypeptide terminates in an amino group at one end (N-terminal) and a carboxyl group at the other end (C-terminal). The length of polypeptide chain varies considerably. For example, human cytochrome *c* has 104 amino acid residues with the molecular weight of 13 kDa, whereas human titin protein has 26,926 amino acid residues with molecular weight of 2993 kDa. The average size of human proteins is about 50 kDa.

The structure of a protein is categorized into four levels: primary, secondary, tertiary, and quaternary structures (Fig. 2-2). *Primary structure* is the linear order of amino acid sequence in the polypeptide chain. *Secondary structure* is the local steric interaction resulting from the hydrogen bonding between O and N of the C=O and the N=H of peptide backbone. The most prevalent elements of the secondary structure are *α-helix*, *β-sheet*, *and turn*. Particular combinations of these secondary structure elements form *motifs* such as helix–loop–helix. *The tertiary structure* is a folded three-dimensional shape of a polypeptide resulting from long-distance interactions between different regions of the protein molecule. The three-dimensional globular region is known as a *domain*, the functional unit of proteins. The large protein contains multiple domains. The final folded protein structure or *conformation* is largely stabilized by weak interactions such as hydrogen bonds and ionic interactions by

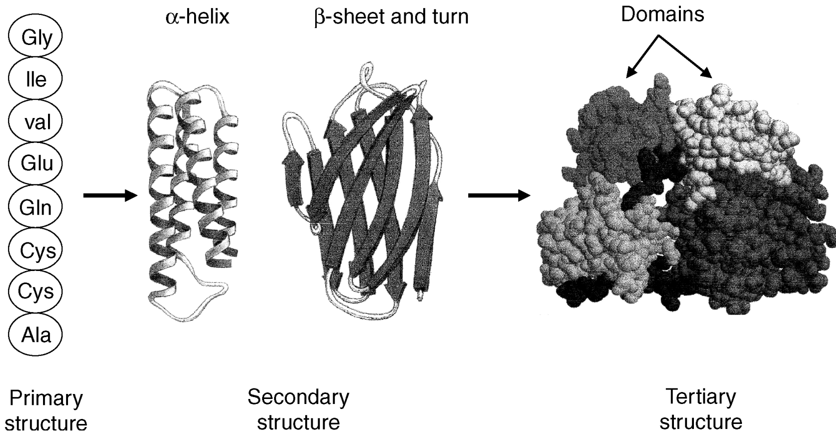

Figure 2-2 Levels of the hierarchical structure in protein. The *primary structure* is a sequence of amino acids linked by peptide bonds; the resulting peptide is coiled into *secondary structure*, α-helix and β-sheet; they are folded to form the *tertiary structure* that contains several functional domains (illustrated are α-helix, β-sheet, and spacing–filling model of Src protein kinase (Copyright 2002 from Molecular Biology of the Cell by Albert et al, reproduced by permission of Garland Science).

which free energy is minimal. Interruption of the hydrogen bonds results in loss of the secondary and tertiary structure or denaturation. Heat and extremes of pH or high concentration of salt induce the protein denaturation. The two general classes of proteins are fibrous and globular, based on their tertiary structures. *The quaternary structure* is the arrangement of two or more polypeptide *subunits* that fit together in space to form a single functional complex. For example, hemoglobin contains four subunits, two α- and two β-subunits, with a symmetrical arrangement to form a functional multisubunit protein to carry oxygen. Furthermore, multiple proteins can assemble spontaneously into complex structures as functional cellular machinery, such as replisome, ribosome, and proteasome.

The protein can be purified from the tissues by various techniques such as chromatography and electrophoresis. The amino acid composition and sequence of proteins can be determined chemically or by mass spectrometry. Three-dimensional structures of proteins are determined by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. The protein can also be synthesized with recombinant DNA technology in significant quantity. Many useful protein databases are also available. SwissProt, PDB, and SCOP are examples of such databases.

2.1.2.3 Protein Function Proteins are the most abundant biomolecules in mammalian cells and consist of 18 percent of total cell weight compared with 0.25 percent of DNA and 1 percent of RNA. Proteins are the basic structural and functional molecules of cells, whereas DNA and RNA simply serve as vehicles to store and express genetic information. The proteins carry out almost all biological activities in the cells. There are about 200,000 different proteins in human body. According to its particular structure, each protein has its specific function. However, it usually takes

more than one protein to accomplish a biologic task. All biological activities are archived by interaction of multiple proteins and other biomolecules such as RNA or small molecules. Often these interactions are reversible binding a *ligand* through the binding site of protein; simple examples include enzyme binding to its substrate and the receptor binding to a hormone ligand.

The largest group of proteins with a related function is the *enzymes*. These proteins specialize in catalyzing chemical reactions within cellular compartments. Enzymes increase the rate at which a chemical reaction reaches equilibrium, but they do not alter the end point of the chemical equilibrium. Enzymes can enhance reaction rates by a factor of 10^5–10^{17} in a very heterogeneous biochemical mixture. Their highly effective and specific catalytic properties largely determine the metabolic capacity of any given cell type. The catalytic properties and specificity of an enzyme are determined by the specific chemical configuration on the protein surface, *active site*. These sites are associated with a pocket, a cleft, or a pit on the surface of the enzyme, which binds the reactants or substrates facilitating chemical change by reducing activation energy. Enzymatically catalyzed reactions control metabolic activities in all cells.

There are many proteins other than enzymes that are basic structural components of cells or critical functional molecules in the organisms. These include such diverse examples as collagen, the connective tissue molecule; actin and myosin, the contractile proteins; insulin, the pancreatic hormone for glucose metabolism; and immunoglobulins (Igs), the antibody molecules of the immune system; histones, the proteins integral to chromosome structure in eukaryotes, and so on.

The potential for such diverse functions rests with the enormous variation of three-dimensional conformation that may be achieved by proteins. The final conformation of a protein is the direct result of the unique linear sequence of amino acids. To come full circle, the amino acid sequence of protein is determined by DNA sequence.

Any given biological function is the sum of work involving hundreds of different related proteins. Life depends on thousands of proteins with specific properties and functions. The activity of each protein component as well as the whole network is highly regulated to meet physiological needs in any given time and condition. These dynamic biochemical processes define the forms of the life. For example, the budding yeast *Saccharomyces cerevisiae* contain 6000 genes. By using the two-hybrid screening or the double mutation scoring method, a large scale of protein–protein interactions has been mapped. Elucidating the whole set of proteins and the interaction of the proteins in living cells is an important task of systems biology.

2.1.3 Polysaccharides

Polysaccharides (glycans) are carbohydrate polymers made up of many monosaccharides joined together by glycosidic linkages. The most abundant monosaccharide is D-glucose (dextrose), a six-carbon sugar containing an aldose group and five hydroxyl groups. Glucose and other hexose derivatives usually form a ring structure in aqueous solution with either α- or β-anomer. Thousands of monomers of the same type, such as glucose, link together to form homopolysaccharides. Examples include storage polysaccharides such as starch and glycogen. Heteropolysaccharides contains two or more

different monomers. Polysaccharides are very large, often branched, molecules. They tend to be amorphous, insoluble in water, and have no sweet taste. Polysaccharides have a general formula of $(CH_2O)_n$; therefore, they are sometimes called carbohydrate.

Carbohydrates are not only the primary source of fuel and structural components of cells but also important informational molecules. Monosaccharides can be assembled into an almost unlimited variety of oligosaccharides, which differ in the stereochemistry and position of glycosidic bonds, the type and orientation of substituent groups, and the number and type of branches. These oligosaccharides are covalently linked with proteins or lipids to form glycoproteins or glycolipids on cell surface. The specific configuration of these oligosaccharides provides recognition sides for cell–cell interaction, bacterial toxin, or viral adhesion onto the cells. For example, blood types are determined by different oligosaccharides on the red blood cells.

2.1.4 Lipids

Biological lipids comprise a diverse group of molecules that are relatively water insoluble or nonpolar. Lipids commonly found in animals or plants include fatty acids and fatty acid-derived phospholipids, sphingolipids, glycolipids, sterols, and waxes. Some lipids are linear aliphatic molecules, whereas others have ring structures. Some are flexible, whereas others are rigid. The biological functions are as diverse as their chemistry. Fats and oils are the principal storage forms of fuel. Too much storage of fats results in obesity, which is becoming a severe problem in public health in the modern society. Phospholipids and sterols are major structural components of biomembrane. Steroid hormones, eicosanoids, and phosphorylated derivatives are important molecules in cell signaling.

In addition to being largely nonpolar molecules, most lipids have some degree of polar property. Generally, the bulk of their structure is nonpolar or hydrophobic acyl chains consisting of an even number of 10–22 hydrocarbon units (CH_2). They are either saturated or unsaturated. Another part of their structure is polar or hydrophilic containing carboxyl, hydroxyl, or phosphorated group. The bipolar feature of lipids (polar head and nonpolar tail) makes them amphophilic molecules. In the case of cholesterol, the polar group is a mere hydroxyl group. In the case of phospholipids, the polar groups are considerably larger.

Phospholipids, or, more precisely, glycerophospholipids or phosphoglycerides, have a glycerol core where two fatty acid-derived ‘‘tails’’ are linked to the first two carbons by ester bonds and one ‘‘head’’ group is linked to the third carbon by a phosphodiester bond. The phospholipids found in biological membranes are phosphatidylcholine (lecithin), phosphatidylethanolamine, phosphatidylserine, and phosphatidylinositol.

2.2 MICROSTRUCTURE AND FUNCTION OF CELLS

Biomolecules can be defined as any molecules found in living organisms. They can be either macromolecules or small molecules. Macromolecules include proteins, nucleic

acids, polysaccharides, and lipids as already described. Small molecules are water, inorganic ions, and hundreds of organic metabolites. A simple mixture of these molecules cannot make life. All living organisms are highly organized structural and functional systems that are characterized by their ability to metabolize and self-replicate.

Since Leeuwenhoek first observed cells with his simple microscope in 1674, it has been confirmed that all living organisms are composed of cells. Many animals have trillions of cells, whereas bacteria are single-celled organisms. Cells are basic structural and functional units of life. The defined function of a living cell is determined by its particular structure.

Cells are small and complex. Under light microscope, cells can be divided into three parts: cell membrane, cytoplasm, and nucleus. There are many distinct functional structures or organelles in cells observed through electron microscope.

2.2.1 Prokaryotic Versus Eukaryotic Cells

There are 10–100 million living species in the biological universe on Earth. They consist of two basic cell types, prokaryotic and eukaryotic. Prokaryotic cells are less complex with a single compartment surrounded by cell membrane and with no defined nucleus. In contrast, eukaryotic cells have defined membrane-bounded nucleus and extensive internal membrane compartments or organelles. Bacteria are single-celled prokaryotes, whereas numerous animals, plants, and fungi are eukaryotes. Since the genetic code in DNA is same in all living organisms, prokaryotes and eukaryotes probably evolved from a common single-celled progenitor. Prokaryotic cells may represent the primitive cell type on Earth and eukaryotic cell types evolved from them. Table 2-3 shows the comparison between the two cell types. Figure 2-3 shows the microstructure of a typical animal cell.

2.2.2 Cell Membrane

A eukaryotic cell is classically divided into three compartments: cell membrane, cytoplasm, and nucleus. The basic architecture of cell membranes consists of a lipid bilayer associated with peripheral extrinsic proteins and intrinsic integral proteins (fluid mosaic model). Peripheral proteins are loosely associated with membrane through electrostatic and hydrogen bonds or by covalently attached lipid anchors. Integral proteins associate firmly with the membrane by hydrophobic interactions between the interior of the lipid bilayer and nonpolar amino acid side chains. The transmembrane sequences consist of about 20 or more amino acid residues in either α-helix or β-barrel structure. The composition of both lipids and proteins in the inner and outer lefts of the membrane is asymmetric. Many extrinsic proteins on the outside surface of the cell membrane are attached by oligosaccharides that are molecules for cell–cell interaction. The plasma membrane defines the external boundaries of cells and regulates the molecular traffic across the boundary. In eukaryotic cells, the biomembrane also divides cytoplasmic space into many distinct functional compartments (organelles). Biomembranes are crucial for life;

Table 2-3 Comparison of prokaryotic and eukaryotic cells

Characteristics	Prokaryotic Cell	Eukaryotic Cell
Prototype	Bacteria	Animal and plant cells
Structure	Simple	Complex
Cell size	1–10 μm	1–500 μm
Genome makeup	DNA with nonhistone proteins	DNA with histone and nonhistone proteins forming chromosomes
DNA size	$1–4 \times 10^6$ bp	$1–3 \times 10^9$ bp
Gene number	∼4300 (*E. coli*)	∼25,000 (human)
Nucleus envelope	Absent	Present
Cell division	Fission or budding	Mitosis
Membrane-bounded organelles	Absent	Present such as mitochondria or chloroplasts (plants), endoplasmic reticulum, Golgi apparatus, and lysosomes
Energy metabolism	Variable metabolic patterns, no mitochondria	More unified oxidative metabolism in mitochondria
Cytoskeleton	None	Complex with microtubules, intermediate filaments, and actins

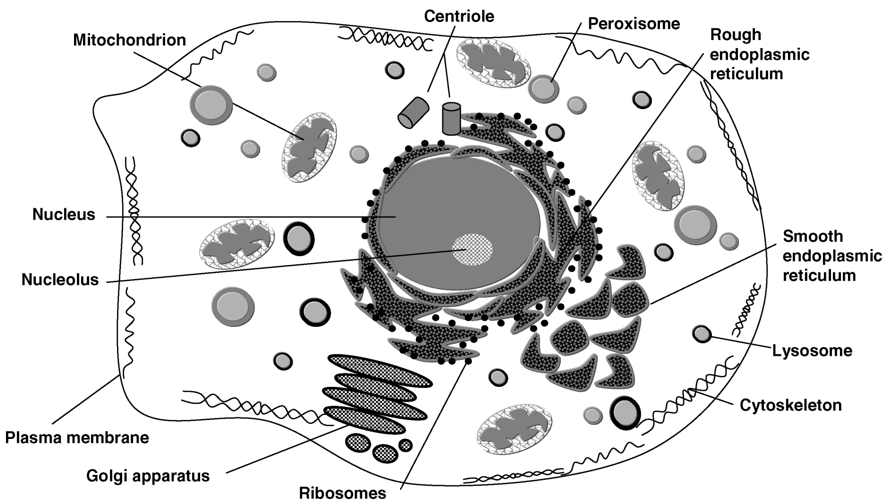

Figure 2-3 Substructures of a typical animal cell.(a) Plasma membrane with bilayer fluid mosaic structure. (b) Nucleus is filled with chromatin composed of DNA and nuclear proteins. (c) Nucleolus is a nuclear subcompartment area where rRNA is synthesized. (d) Mitochondria are surrounded by a double membrane, where ATP is generated. (e) Centriole for mitosis. (f) Peroxisomes. (g) Rough endoplasmic reticulum, attached with many ribosomes. (h) Smooth endoplasmic reticulum. (i) Lysosomes, biomembrane structure with an acidic lumen. (j) Cytoskeletal fibers form networks and bundles. (k) Ribosome. (l) Golgi apparatus. (m) Secretory vesicles store secreted proteins. (n) Nuclear envelope, a double membrane, outer membrane, is continuous with the rough ER .

there are two major reasons why the cell must separate itself from the outside environment. First, it must keep its biomolecules (DNA, RNA, proteins, and metabolites) inside the cell and keep foreign material outside the cell. Second, it must communicate with the environment to continuously monitor the external conditions and adapt to them and exchange the materials and energy with its surroundings. For example, when an *Escherichia coli* bacterium detects a high concentration of lactose in medium, it begins synthesizing proteins for metabolism of lactose. It needs to pump lactose in through a lactose transporter and release toxic metabolic products (see Fig. 2-8a). Thus, the cell membrane uses the lipid bilayer to function as a physical barrier and the integral proteins to function as selective biochemical transporters. Some types of transporters have ATPase activity and pump ions (channels) or small molecules (transporters) against electrochemical or concentration gradient.

In addition, some membrane proteins on the cell surface have specific functions in signal reception or receptors (see Fig. 2-12). The ability of cells to receive and act on signals in their surrounding environment is important for survival and cell–cell interaction in multicellular organisms. Each of the cells in all tissues communicates with dozens, if not hundreds, of other types of cells about a variety of important issues, such as when it should grow or differentiate or die, when it should release certain protein products such as growth factors or hormones needed by other cells at distant sites in the body, and what other cells it should associate with to build complex tissue architectures. These crucial decisions are made at tissue or whole body level to maintain a dynamic living system by cell receptors and signaling.

2.2.3 Cytoplasm

The homogeneous region of the cell between the plasma membrane and the nucleus is defined as the cytoplasm. In fact, the cytoplasm is not "homogeneous," but when viewed under an electron microscope, it is a highly compartmentalized structure. After centrifugation at high speed, the cytoplasm is separated into two fractions, supernatant aqueous phase (cytosol) and pellet phase (organelles). The cytosol is composed of water, ions, nutrients, and soluble macromolecules such as enzymes, carbohydrates, RNA, and a vast variety of metabolites. The cytosol makes up some 50 percent of the cell volume and functions as a perfect biochemical matrix in which hundreds of metabolic reactions occur in any given moment. All protein synthesis and glycolysis are carried out within the cytosol. On the contrary, pellet phase is rich in organelles with diverse functions (see subsequent sections).

2.2.4 Nucleus

The nucleus is the central compartment formed with two layers of concentric continuous biomembrane that is punctured with *nuclear pores*. This is where DNA is stored and RNA is synthesized. DNA is the inherited genetic material containing all information for the cell to live and to function. Synthesized RNAs are transported out of the nucleus through the nuclear pores. Proteins needed inside the nucleus are

imported through the nuclear pores. The *nucleolus* is usually visible as a dark or red region in the nucleus where much of rRNA is synthesized and the ribosome is assembled.

2.2.5 Organelles

Organelles are membrane-bound small structures within eukaryotic cells that perform dedicated functions. Structural compartmentalization creates a stable environment and increases local concentration of reactive molecules, thus improving the biochemical efficiency. There are a dozen different types of organelles commonly found in eukaryotic cells, including mitochondria, chloroplast (in plants), lysosomes, peroxisomes, ribosomes, endoplasmic reticulum (ER), Golgi apparatus, and some vacuoles. Here, we will focus on only a handful of organelles and their roles at a molecular level in the cell with a brief description of the structure.

2.2.5.1 Cytoskeleton Network The cytoplasm contains numerous filaments that form an interlocking three-dimensional network or the cytoskeleton. There are three types of cytoplasmic filaments: microfilaments (actins), intermediate filaments, and microtubules. They differ in diameter (from about 6 to 22 nm), protein subunits, and specific function. The cytoskeleton provides a framework for the traffic of intracellular organelles, the organization of enzyme pathways, mitosis, cell shape, and cell movement.

2.2.5.2 Mitochondria Mitochondria are oval-shaped organelles formed with two layers of biomembrane, an inner and an outer layer resulting in two internal compartments, central matrix and intermembrane space. Aerobic respiration occurs in the mitochondria. Many enzymes for oxidative reactions including oxidation of pyruvate, fatty acid, and citric acid cycle are enriched in the central matrix. The inner membrane contains 80 percent of proteins where adenosine triphosphate (ATP) is generated by linking oxidative phosphorylation. The outer membrane contains 50 percent of proteins similar to cytoplasm membrane. The mitochondria contain its own circular DNA (16,569 bp) that encodes some proteins used in oxidative phosphorylation. The existence of the double membrane and complete genetic system have led many biologists believe that mitochondria are the descendants of some bacteria that has been endocytosed by a larger cell a billion years ago and coexist in an endosymbiotic relationship.

2.2.5.3 Chloroplasts These organelles are the site of photosynthesis in plants and other photosynthesizing organisms. Similar to mitochondria, they also have a double membrane.

2.2.5.4 Endoplasmic Reticulum The ER is an extensive network of biomembrane-bounded sacs, the site for many biosynthesis. There are two types of ER, rough and smooth. Rough ER is attached with many *ribosomes* where RNA is translated into protein. The smooth ER is the site for fatty acid, steroid, and phospholipid synthesis.

2.2.5.5 Golgi Apparatus (Complex) Like ER, Golgi apparatus is a stack of flattened membrane vesicles. This organelle modifies (such as glycosylation) and packages newly synthesized proteins from rough ER into small membrane-bound vesicles. These vesicles can be targeted to various locations in the cell and secreted out of the cell.

2.2.5.6 Lysosome This organelle contains a group of enzymes to digest large biomolecules into small monomeric subunits. All lysosomal enzymes have high activity at acid pH in lumen and collectively termed acid hydrolases. When the enzymes are released into cytoplasm, their activity is diminished.

2.2.5.7 Peroxisomes Unlike mitochondria, peroxisomes contain several oxidases for fatty acid oxidation, but do not produce ATP. Instead, the energy is released into heat. In addition, H_2O_2 can be formed and degraded by catalase, important reactions for detoxification, as in the cases of ethanol and other toxic molecules.

2.3 SYNTHESIS OF BIOMOLECULES IN LIVING CELLS

2.3.1 Bioenergetics: The Law of Order

Each cell can be viewed as a tiny chemical factory. Cells require an ongoing supply of energy to carry out various kinds of work, including synthesis, movement, concentration, charge separation, the generation of heat, and bioluminescence. The energy required for these processes comes either from the sun or from the organic molecules such as carbohydrates, fats, and proteins. Organisms including plants, algae, and certain groups of bacteria are capable of capturing light energy by means of photosynthetic reaction. This group of organisms is called phototrophs. Another group of organisms is called chemotrophs because they require the intake of chemical compounds such as carbohydrates, fats, and proteins. All animals, fungi, and most bacteria are chemotrophs.

The flow of energy through cells is followed by the laws of thermodynamics. *Bioenergetics* is the application of thermodynamic principle in the biological system. The first law of thermodynamics (conservation) states that energy is always conserved, it cannot be created or destroyed, but energy can be converted from one form into another. The second law of thermodynamics states that "in all energy exchanges, if no energy enters or leaves the system, the potential energy of the state will always be less than that of the initial state." The second law provides a measure of thermodynamic spontaneity, although this only means that a reaction can go and says nothing about whether it will actually go or at what rate. Free energy change, ΔG, is a measure of thermodynamic spontaneity and is defined so that negative values correspond to favorable reaction and positive values represent unfavorable reaction. A negative ΔG is a necessary prerequisite for a reaction to proceed, but it does not guarantee that the reaction will actually occur at a reasonable rate. The presence or absence of an appropriate catalyst such as an enzyme will determine the rate at which a reaction can occur.

2.3.1.1 Energy Carrier Adenosine triphosphate is a universal "energy currency" in the cell and it can store and release the energy efficiently. Adenosine may occur in the cell in the unphosphorylated form or with one, two, or three phosphates attached, forming adenosine monophosphate (AMP), adenosine diphosphate (ADP), and adenosine triphosphate (ATP), respectively. ATP is extremely rich in chemical energy, in particular between the second and third phosphate groups. The net change in energy of the decomposition of ATP into ADP and an inorganic phosphate is -12 kcal/mol *in vivo* and -7.3 kcal/mol *in vitro*. This massive release in energy makes the decomposition of ATP extremely exergonic, and hence useful as a means for chemically storing energy. Many biochemical reactions that occur inside a cell are coupled with the formation or decomposition of ATP.

2.3.1.2 Electron Carriers Most of the energy of eukaryotic cells is generated from oxidizing fuel molecules, which involves the transfer of electrons from fuel molecules to oxygen. The fuel molecules are oxidized, while the oxygen is reduced. The electron carrier molecules of choice are nicotinamide adenine dinucleotide (NAD) and its relative nicotinamide adenine dinucleotide phosphate (NADP), two of the most important coenzymes in the cell. The oxidized forms, NAD^+ and $NADP^+$, serve as electron acceptors by acquiring two electrons, thereby generating the reduced form NADH and NADPH. However, only one proton accompanies the reduction. The other proton, produced as two hydrogen atoms are removed from the molecule being oxidized, is liberated into the surrounding medium.

2.3.2 Enzymes as Catalysts of Life

Enzymes allow many chemical reactions to occur within the homeostasis constraints of a living system. Enzymes function as biological catalysts. The use of enzyme can decrease the free energy of activation of chemical reactions. The first step in catalysis is the formation of an enzyme–substrate complex. By bringing the reactants closer together, chemical bonds may be weakened and reactions will proceed faster than without the enzyme. An enzyme-catalyzed reaction proceeds via an enzyme–substrate intermediate and follows Michaelis–Menten kinetics, which is characterized by a hyperbolic relationship between the initial reaction rate (velocity, v) and the substrate concentration [s].

Enzymes are regulated by many ways to adjust their intracellular concentrations and activity levels to meet the cellular needs. First, all protein enzymes are sensitive to temperature and pH. Changes in temperature or pH may denature the enzyme and most enzymes are adapted to operate at a specific pH or pH range. Second, enzyme activity is influenced not only by substrate availability but also by products, alternative substrates, substrate analogues, cofactors, and coenzymes. The binding of the substrate to the active site of an enzyme alters the structure of the enzyme, placing some strain on the substrate and further facilitating the reaction. Cofactors are nonproteins essential for enzyme activity. Ions such as Zn^{2+} and Cu^{2+} are cofactors. Coenzymes are nonprotein organic molecules such as NAD^+ or $NADP^+$ bound to enzymes near the active site. Additional control mechanisms include allosteric

regulation. Most allosterically regulated enzymes catalyze the first step in a reaction sequence and are multisubunit proteins with multiple catalytic subunits and multiple regulatory subunits. Each of the catalytic subunits has an active site that recognizes substrates and products, whereas each regulatory subunit has one or more allosteric sites that recognize specific effector molecules. A given effector may either inhibit or activate the enzyme, depending on which form of the enzymes is favored by effector binding. Such a mechanism is commonly employed in feedback inhibition. Often one of the products, either an end or near-end product, acts as an allosteric effector, blocking or shunting the pathway. The biosynthesis of enzymes in living cells is , also subjected to various regulations at transcriptional and translational levels described subsequently.

2.3.3 Metabolism and Metabolic Pathways

Metabolism is all of the biochemical reactions that occur within a cell. This includes the biosynthesis of complex organic molecules such as nucleic acids, proteins, lipids, and carbohydrates (anabolism) and the degradation of these large molecules into smaller, simpler ones with the release of chemical energy (catabolism) in the form of ATP. Catabolism can be carried out either in the presence of oxygen (aerobic conditions) or in the absence of oxygen (anaerobic conditions). The energy yield is much greater in the presence of oxygen, which probably explains the preponderance of aerobic organisms in the world. However, anaerobic catabolism is also important, both for organisms in environments that are always devoid of oxygen and for organisms and cells that are temporarily deprived of oxygen. Photosynthesis, a phototrophic energy metabolism, is an important biochemical process in which plants, algae, and some bacteria acquire the energy of sunlight to produce food. Ultimately, nearly all living things depend on energy produced from photosynthesis for their nourishment, making it vital to life on Earth.

Metabolism usually consists of sequences of enzymatic steps, also called metabolic pathways. Metabolic pathways are of two general types: anabolic pathways are connected with the synthesis of cellular components and are usually involved in a substantial increase in molecular order and require energy, whereas catabolic pathways are involved in the breakdown of cellular constituents and release energy. Catabolic pathways play two roles in cells: they give rise to the small organic molecules or metabolites that are the building blocks for biosynthesis and the production of energy that is used to synthesize the macromolecules and other cellular function.

A very large number of metabolic pathways including both catabolic and anabolic pathways have been discovered and they can be found in various databases on the Internet. The KEGG collection (http://www.genome.jp/kegg/) of metabolic and regulatory databases currently has a record of 54,622 metabolic pathways. As an example, Figure 2-4 shows the most common and well-known carbohydrate catabolic pathway, including *glycolysis*, *pentose–phosphate* (*PP*) *pathway*, *fermentation*, and *aerobic respiration* in recombinant *E. coli*. Using glucose as a prototype substrate, catabolism under both anaerobic and aerobic conditions begins with the glycolytic

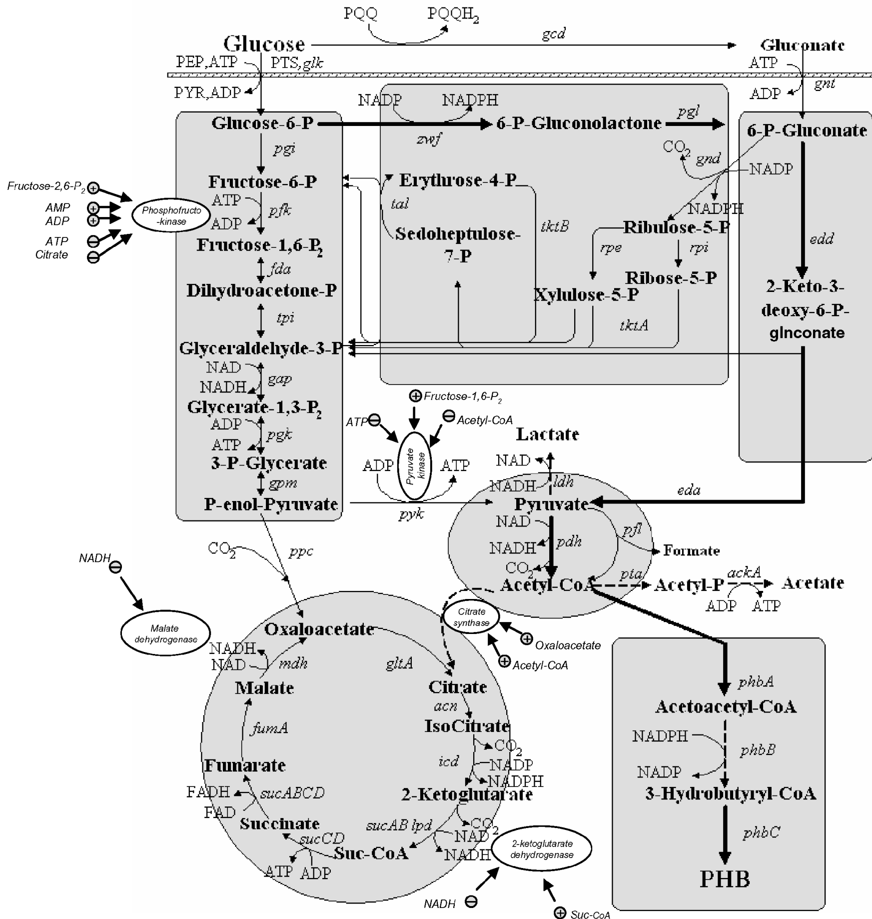

Figure 2-4 Central metabolic pathways in a recombinant *E. coli* expressing three genes for poly-3-hydroxybutyrate (PHB) synthesis. Glycolysis, fermentation, TCA cycle, and PHB synthesis pathways are shown. Genes encoding the important metabolic enzymes are also indicated (italic letters). Major regulatory effects are indicated as either activation (+) or inhibition (−). Kinetic regulators are highlighted: fructose-2, 6-P2, citrate, and acetyl-CoA for glycolysis; acetyl-CoA, oxaloacetate, and succinyl-CoA for TCA cycles. Key allosteric effectors are NAD^+, ADP, and AMP as activator and acetyl-CoA, NADH, and ATP as inhibitors.

pathway, a 10-step sequence of reactions in which glucose is converted to pyruvate with the net production of two molecules of ATP per molecules of glucose. In the absence of oxygen, reduced coenzyme generated during glycolysis must be reoxidized at the expense of pyruvate, leading to fermentation end products such as acetate, lactate or ethanol, and carbon dioxide. The most common and well-known type of glycolysis is the *Embden–Meyerhof* (*EMP*) pathway. Compared to fermentative process, *aerobic respiration* gives the cell access to much more of the free energy that is released by the oxidation of organic substrates. The completion of catabolism of carbohydrates begins with the glycolytic pathways mentioned above, but the pyruvate

is converted to acetyl CoA. The acetyl CoA is then oxidized fully by enzymes of the *tricarboxylic acid* (*TCA*) cycle. The reduced coenzymes (NADH, $FADH_2$) are further reoxidized by the *electron transport system* and generated additional ATP. A total of 38 molecules of ATP are generated per molecule of glucose in most prokaryotic and some eukaryotic cells. This is a factor of 19 times more energy per sugar molecule than what the typical anaerobic reaction generates. The small molecules such as acetyl-CoA and energy such as ATP, NADH, and NADPH generated by these catabolic reactions can be used for biosynthesis of other cellular products and cell mass. ATP, NADH, and NADPH are continually generated and consumed. NADPH, which carries two electrons at a high potential, provides reducing power in the biosynthesis of cell components from more oxidized precursors. Figure 2-4 also illustrates a biopolymer PHB synthetic pathway that is introduced into *E. coli* by recombinant DNA techniques. In recombinant *E. coli*, PHB can be synthesized from acetyl-CoA by a sequence of three enzymatic reactions catalyzed by β-ketothiolase, acetoacetyl-CoA reductase, and PHB synthase. Moreover, key reaction types are used repeatedly in metabolic pathways.

2.3.4 Regulation of Metabolism

As already discussed, metabolic pathways form as a result of the common occurrence of a series of dependent chemical reactions. These reactions are carefully regulated to ensure that the rate of product formation is tuned to actual cellular need. The end product of the pathway depends on the successful completion of all sequential reactions, each mediated by a specific enzyme. Also, intermediate products tend not to accumulate, making the process more efficient. The metabolism is regulated at two levels to achieve the overall cellular "fitness" and balance. First, catalytic activities of many enzymes are directly regulated by allosteric interactions (as in feedback inhibition) and by covalent modification. These processes regulate the activity of preexisting enzymes in both catabolic and anabolic pathways. Most of the constitutive enzymes such as the enzymes operating in glycolysis and TCA cycles are regulated at this level. Several key regulatory steps in glycolysis and TCA cycles are highlighted in Figure 2-4. Second, the amounts of many enzymes are controlled by regulation of the rate of protein synthesis and degradation. The processes of end product repression, enzyme induction, and catabolite repression are involved in the control of synthesis of enzymes. End product repression and enzyme induction are mechanisms of negative control that lead to a decrease in the transcription of proteins. Catabolite repression is considered a form of positive control because it leads to an increase in transcription of proteins. Many inducible or repressible enzymes are regulated at this level. Such examples of enzyme induction and catabolite repression (i. e., repression and induction of *lac* operon, see Fig. 2-8a) will be discussed subsequently. In addition, the movement of many substrates into cells and subcellular compartments is also controlled. Distinct pathways for biosynthesis and degradation contribute to metabolic regulation. The energy charge, which depends on the relative amounts of ATP, ADP, and AMP, plays an important role in metabolic regulation. A high-energy charge inhibits ATP-generating (catabolic) pathways and stimulates ATP-utilizing (anabolic) pathways.

2.4 THE INFORMATION FLOW IN LIVING CELLS

As we discussed, the cell is the basic structural and functional unit for all living species. In cells the genetic information is carried by DNA molecules and its specificity is determined by the sequence of nucleotides. The proteins are functional forms of life and perform most cellular activities. The largest group of proteins with a related function is enzymes. In this section, we will discuss how genetic information flow transmits from one generation to the next; how genetic information is expressed as functional proteins; and how expression is controlled in a living cell by cell signaling. On the other hand, genetic information can also be inherited epigenetically. In addition, genetic, epigenetic, and biochemical information flow are integrated for final cellular controls.

2.4.1 Genetic Information Flow

The so-called central dogma of molecular biology comprises the three major biological processes: replication, transcription, and translation. Replication is the copying of parental DNA to form daughter DNA molecules with identical nucleotide sequences as well as identical epigenetic modifications. Transcription is the process by which the genetic information stored in DNA sequence is copied precisely into messenger RNA. Translation is the genetic information encoded in messenger RNA being translated into a polypeptide. With the completion of human genome project and with genome sequences of many other species available, these processes can be studied at genome levels (Fig. 2-5).

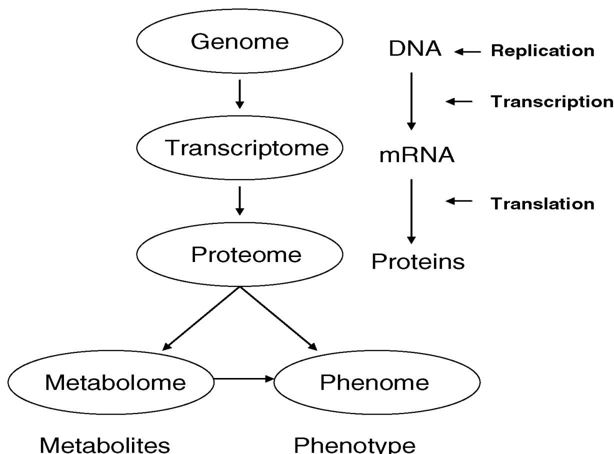

Figure 2-5 New central dogma of molecular biology. The classical central dogma (DNA → RNA → protein in single gene) has been redefined at whole genome level. Ultimately, the total phenotypic characteristic of an organism is determined by the proteome and the whole set of functional proteins and their interactions.

2.4.1.1 DNA Replication Genome is defined as the complete set of genetic information carried by a cell in the organism. DNA is the primary material that carries genetic information. For many viruses and prokaryotes, the genome consists of one linear or circular DNA molecule. In the eukaryotic cell, DNA is packed in multiple chromosomes and confined by nuclear envelope in the nucleus. The size of genome is quite different among species. The human genome contains 3.2×10^9 nucleotide pairs, divided into 22 different autosomal chromosomes and 2 sex chromosomes, while prokaryote *E. coli* contains 4.6×10^6 nucleotide pairs in a single circular DNA. Eukaryotic chromosomes consist of numerous highly coiled DNA/histone complex or nucleosome connected by linker DNA. Each nucleosome contains the protein core of eight histones (two copies each of H2A, H2B, H3, and H4) and a 200 bp segment of DNA. The histone core is encircled by DNA fiber (Fig. 2-11).

With these basic concepts in mind, we will discuss the chemical mechanism of heredity in which the genetic information is passed from a cell to its daughter cells at cell division and from one generation to the next through the reproduction of organisms.

DNA replication is semiconservative; each strand of double helix of DNA can be used as the template to synthesize a new complementary strand of DNA (Fig. 2-6a). It

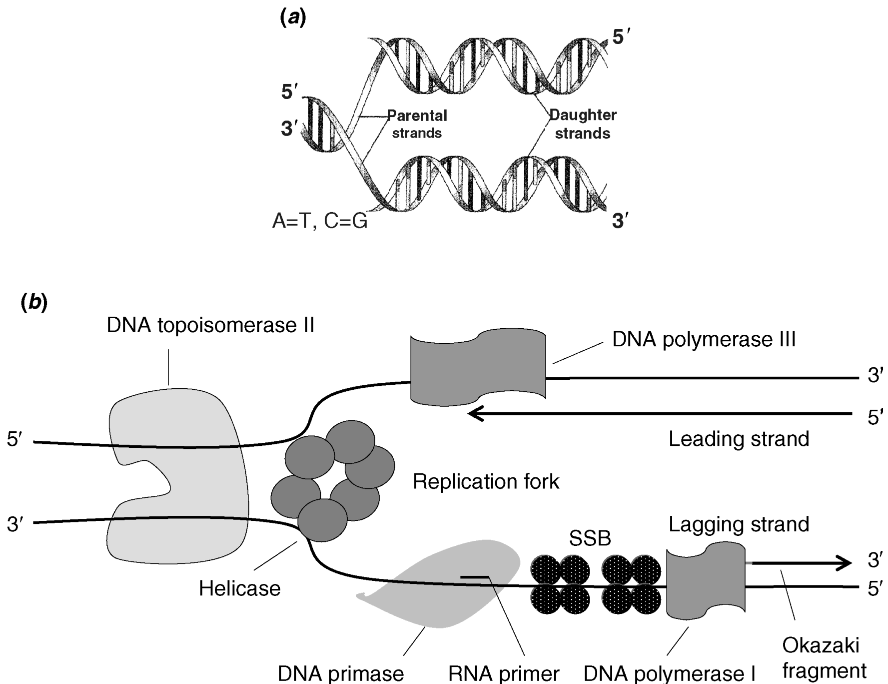

Figure 2-6 DNA replication. (a) Two strands of parental DNA helix are unwound and used as templates to produce new daughter strands. The outcome is two copies of identical DNA, each containing one of the original strands and one new complementary strand (semiconservative replication). (b) The main biochemical steps (see text) and the proteins involved in DNA replication are illustrated.

is carried out in three identifiable phases: initiation, elongation, and termination. The replication starts at the origin and usually precedes bidirectionally. Parental double helix DNA must be separated from one another (denaturation) to be the templates. This process is accomplished by many enzymes such as helicases and topoisomerases under normal physiological condition. DNA is then synthesized in the $5' \rightarrow 3'$ direction by DNA polymerases. Since the two strands of a DNA are antiparallel, this $5' \rightarrow 3'$ DNA synthesis can take place continuously on only one of the strands at a replication fork (the leading strand). On the other strand (lagging strand), the short DNA fragments are synthesized discontinuously as Okazaki fragments, which are subsequently ligated and the gaps are filled by ligase. The complex of many proteins and enzymes at the replication fork is called replisome. The key component in replisome is DNA polymerase. Most cells have several DNA polymerases. In *E. coli*, DNA polymerase III is the primary replication enzyme (Fig. 2-6b). Eukaryotic chromosomes have many replication origins and proceed at multiple sites by utilizing DNA polymerase α.

The fidelity of DNA replication is maintained by several mechanisms: (1) base selection by DNA polymerase according to template nucleotide following the role of Watson–Crick base pairing (A/T; C/G); (2) $3' \rightarrow 5'$ proofreading exonuclease activity that is part of most DNA polymerases, and (3) specific DNA repair systems for mismatch correction.

Epigenetic information such as DNA methylation is also inherited during DNA replication. As a result, the two replicated DNA molecules in parental cell are exactly the same sequences and methylcytosine content are equally divided into two daughter cells. Once the two daughter cells receive the same genetic material and epigenetic information, heredity is pursued.

2.4.1.2 From DNA to RNA—Transcription As discussed above, genomic DNA contains all information to build a cell or organism. Although the genome in all somatic cells (except lymphocytes) in a given multicellular organism is same, the structure and function of the different types of cells are totally different. Hepatocytes are different from neurons. Myocytes are different from epithelial cells. The difference is not due to DNA, but mRNAs and proteins.

The central dogma is an early attempt to understand how the amino acid sequence of the protein is determined by nucleic acid sequence of DNA based on following observations. First, the DNA is confined in the nucleus, while protein synthesis occurs in association with ribosomes in the cytoplasm. Second, RNA is synthesized in the nucleus and then transported to the cytoplasm. Third, RNA is chemically similar to DNA. Collectively, these observations suggest that genetic information, stored in DNA, is transferred to an RNA intermediate (mRNA), which directs the synthesis of proteins.

The process by which RNA molecules are synthesized on a DNA template is called transcription. It results in an mRNA molecule complementary to the DNA sequence of one strand (template strand) of the double helix DNA.

Like DNA replication, transcription is also a complicated biochemical process and many protein factors and enzymes are involved. However, transcription occurs only in

particular DNA regions. The process of transcription can be divided into four phases: initiation, elongation, termination, and processing. During initiation, RNA polymerase binds to a specific site in DNA (the promoter), locally melts the double-stranded DNA to reveal the unpaired template strand, and polymerizes the first two nucleotides. There are as many as 30 polypeptides (general transcription factors) assembled as an initiation complex facilitating the initiation. During strand elongation, RNA polymerase (II in eukaryotic DNA) moves along the DNA, melting sequential segments of the DNA and adding nucleotides to the growing RNA strand. When RNA polymerase reaches a termination sequence in the DNA (terminator), the enzyme stops transcription, leading to the release of the completed RNA and dissociation of the enzyme from the template DNA. RNA polymerase and transcription factors can be reused for the next round of transcription.

In eukaryotic DNA, the initial primary transcript (pre-mRNA) very often contains noncoding regions (introns) interspersed along coding regions (exons). Transcripts from genes containing introns undergo splicing, the removal of the introns and joining of the exons, which is catalyzed by small nuclear RNAs in the spliceosome. During processing, the ends of nearly all primary transcripts are also modified by addition of a 5′-cap and 3′-poly (A). Then the mature mRNAs are exported through the nuclear pores to cytoplasm for protein synthesis (Fig. 2-7).

Not all DNA sequences in genome are used for coding RNAs or proteins. In fact, only a small fraction of genome is coding sequence in eukaryotes. For instance, only

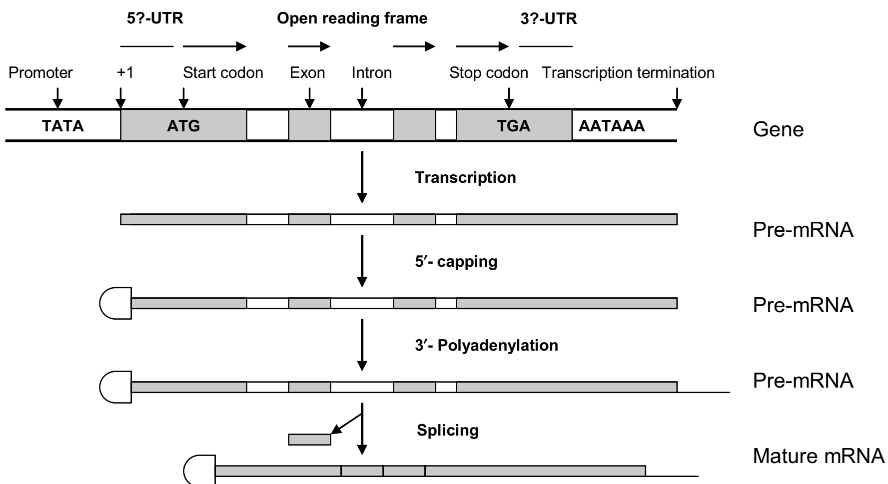

Figure 2-7 Gene transcription. The regulatory region and the coding region of a gene are illustrated in upper panel. The steps of transcription and mRNA processing are illustrated. In both prokaryotes and eukaryotes, a promoter such as TATA box, located upstream of transcription start site (+1), is required for RNA polymerase binding and transcription initiation. The mRNA processing occurs in eukaryotes only. The newly synthesized pre-mRNA needs to be processed, which includes capping 5′-end with 7-methylguanylate, adding poly (A) tail at 3′-end, and removal of introns and connection of exons.

1.2 percent of human genome DNA codes RNAs then proteins. In genome, the region that directs the synthesis of a single polypeptide or functional RNA (such as tRNA) is called a gene. A gene is the physical and functional genetic unit composed of a segment of DNA. A gene consists of a regulatory region and a coding region. A human has proximally 25,000 genes, whereas *E. coli* contains about 4288 genes. However, these genes are not always expressed in the cells. Hepatocytes express only liver function-related genes, whereas neurons express only brain function-required genes. However, some housekeeping genes needed for basic cellar activities express in all cells.

Gene expression is tightly controlled in a given cell type at any given moment in the organisms. Transcription is the most important control point, which can be activated or repressed. In prokaryotes, the classical example of transcriptional control is *lac* operon in *E. coli*. The *lac* operon encodes three enzymes for the metabolism of lactose. For transcription of the *lac* operon to begin, the σ^{70} subunit of the RNA polymerase must bind to the *lac* promoter, which lies just upstream of the transcription start site. When no lactose is present, binding of the *lac* repressor protein to a sequence called the *lac* operator, which overlaps the transcription start site, blocks transcription initiation by the polymerase. When lactose is present, lactose molecules bind to specific binding sites in each subunit of the tetrameric *lac* repressor, causing a conformational change in the protein that makes it dissociate from the *lac* operator. As a result, the polymerase binds to promoter to initiate transcription of the *lac* operon (Fig. 2-8a).

Transcriptional control in eukaryotes is much more complicated than that in prokaryotes. There are at least three types of promoter proximal elements: silencer element, upstream activator sequence (UAS), and core promoter (TATA box, initiator (INR), and downstream promoter elements (DPE)). These proximal elements, located within 200 bp of transcription start sites, direct the basal level of transcription. In contrast, enhancers and suppressors may be located up to 10 kb either upstream or downstream from a promoter. The transcription activator or repressor protein binds to enhancers or suppressors and interacts with basal transcriptional machinery to enhance or repress the transcription. In most cases, the action is mediated by a set of other proteins, the so-called coactivators or corepressor (Fig. 2-8b). Furthermore, transcription activity can also be affected by changing the chromatin configuration. Histone acetylase and deacetylase complex, as well as DNA methylation in CpG islands, can regulate transcription initiation. The epigenetic regulation is important in normal development and pathological processes (Section 2.4.2).

The discovery of miRNA and siRNA has elucidated additional mechanisms of the posttranscriptional control. Both miRNAs and siRNAs contain 21–23 nucleotides that are generated from longer hairpin-like double-stranded precursor RNA ($\sim$70 bp) by DICER ribonuclease. One strand of the shorter duplex intermediate RNA is assembled into a multiprotein, RNA-induced silencing complex (RISC). The miRNA in complex forms imperfect hybrids with sequences in 3′-untranslated region (3′-UTR) of specific target mRNAs and represses the translation initiation. The siRNA in complex forms a perfect hybrid with target mRNA and leads to degradation of target mRNA. Hundred types of miRNA and siRNA have been identified in higher eukaryotes. The double-stranded RNA (dsRNA) containing these small RNA sequences can be constructed or synthesized and then introduced into a cell *in vitro* or *in vivo*. It provides a huge

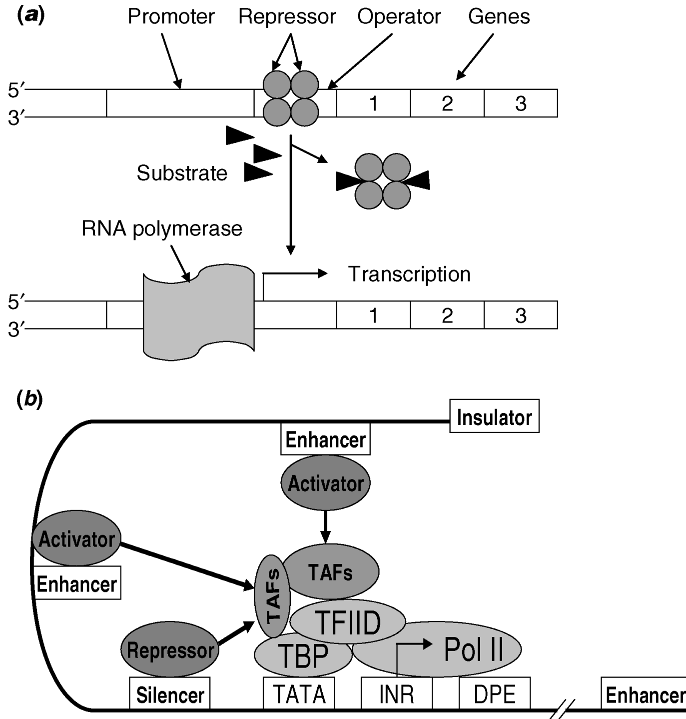

Figure 2-8 Transcription regulation in prokaryotes and eukaryotes. (a) A prokaryotic regulatory unit, operon. Transcription is induced by the substrate that binds to and releases the repressor from the operator, and then RNA polymerase binds to the promoter and initiates the transcription. (b) A eukaryotic regulatory unit. The core promoter contains TATA box (TATA), initiator sequences (INR), and downstream promoter elements. A complex arrangement of multiple enhancers interspersed with silencer and insulator elements that can be located 10–50 kb either upstream or downstream of the core promoter. Transcription is initiated at core promoter regions by interactions of the RNA polymerase II core complex, general transcription factors (TFIID), and multiple subunits of cofactors including TBP-associated factors (TAFs). Transcription factors binding enhancer or silencer regulate the transcription by interaction with TAFs or other cofactors.

potential to specifically knockdown or silence a gene at posttranscriptional level. The target mRNA can be viral (such as HIV) or oncogenic gene transcripts (Fig. 2-9).

2.4.1.3 From RNA to Protein—Translation There are three basic types of RNAs in cells, namely, mRNA, tRNA, and rRNA. Other RNA species include miRNA, siRNA, and snRNA (Table 2-1). All these RNAs somehow play a specific role in mRNA-directed protein synthesis or translation.

mRNA Genetic information stored in DNA in the form of a nucleotide sequence is transcribed into mRNA. The nucleotide sequence in mature mRNA contains

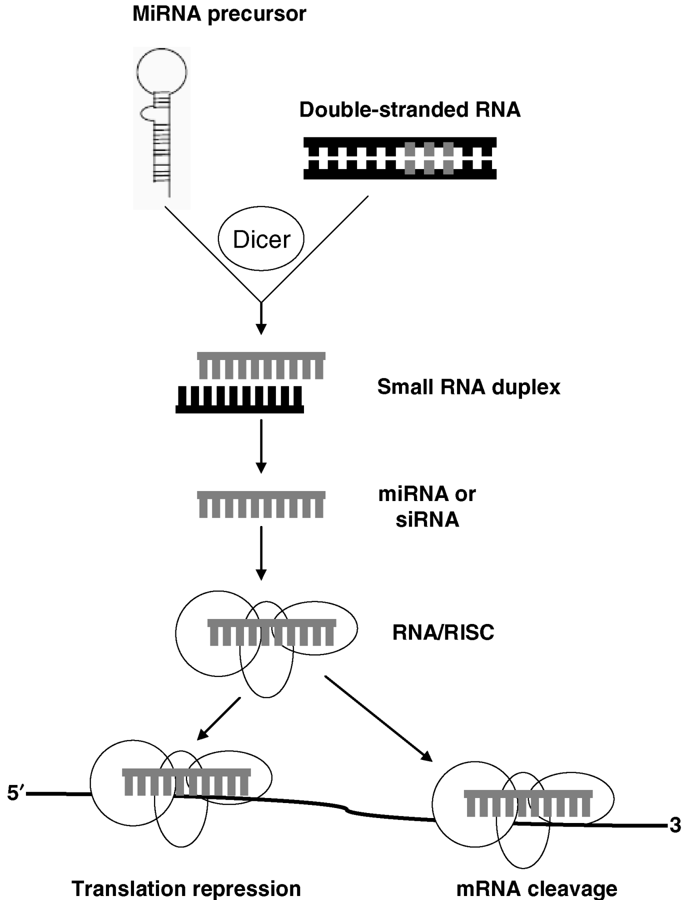

Figure 2-9 RNA interference. Both miRNA and siRNA are cleaved from their precursors, the large double-stranded RNAs, by the Dicer ribonuclease. After binding to multiple proteins and forming RISC, one strand of 21–23 nucleotides hybridizes the target mRNA and degrades mRNA (by siRNA) or inhibits the translation (by miRNA). The gene expression is inhibited by this posttranscriptional regulation.

continuous degenerated triplet nucleotides that determine specific amino acids called codons. Each triplet nucleotide (or codon) in the mRNA is, in turn, complementary to a triplet nucleotide (anticodon) corresponding tRNA. Each tRNA carries a specific amino acid that is correctly inserted into the polypeptide chain during translation. The complete genetic codes have now been elucidated (Table 2-4). Many amino acids are encoded by more than one codon. The AUG codon for methionine is the most common start codon, while three codons (UAA, UGA, UAG) function as stop codons specifying no amino acids. The region of mRNA from the start codon to a stop codon is called the reading frame. The 5′-cap (7-methylguanylate) and poly (A) tail define the 5′- and 3′-ends of mRNA. The region between the 5′-cap and the start codon AUG is known as the 5′-untranslated region (5′-UTR). The region from the stop codon to the start point

Table 2-4 The genetic code in mRNA

First Position	Second Position U	C	A	G	Third Position
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Note: The first base of the codon (5′-end) is shown in the left column and the second base is shown in the third row. The third base in the right column plays lesser specific role; AUG (for Met) is most common initiator codon. Sometimes GUG and UUG are also used. The three termination codons (UAA, UGA, and UAG) match no aminoacyl-tRNA and are recognized by termination factors for translation termination.

of poly (A) tail is known as 3′-untranslation region (3′-UTR). The pre-mRNAs are matured (processing) in nucleus, transported to cytoplasm, and used as templates for protein synthesis during translation. The processes of pre-mRNA maturation include adding 5′-cap, poly (A) tail, and splicing (Fig. 2-7).

tRNA There are 64 types of tRNAs found in eukaryotes. All tRNAs have a similar three-dimensional structure including an acceptor arm for attachment of a specific amino acid and a stem-loop with a three-base anticodon. Each type of amino acid has its own set of tRNAs, which bind the specific amino acid and carry it to the growing end of a polypeptide. The anticodons in each tRNA can base-pair with its complementary codon in the mRNA, by which the nucleotide sequence in the mRNA is translated into amino acid sequence in the peptides.

rRNA The rRNA associates with a set of proteins to form ribosomes. These large ribonucleoprotein complexes move along an mRNA and catalyze the assembly of each amino acid into peptide chain. Both prokaryotic (70S) and eukaryotic (80S) ribosomes consist of a small and a large subunits. Each subunit contains numerous different proteins and one major rRNA molecule. The topological structure of the ribosome has been elucidated in detail. The genes encoding rRNA are in the nucleolus region.

Protein Synthesis Similar to DNA replication and RNA transcription, translation is a highly organized and regulated biochemical process in which many proteins factors are involved. It can be divided into three phases: initiation, elongation, and termination. Before translation initiation, each amino acid needs to be activated by one of the 20 specific aminoacyl-tRNA synthetases. As the result, the amino acid is linked to the acceptor arm of tRNA by a high-energy bond, aminoacyl-tRNA.

During initiation, the small ribosomal subunit binds to mRNA near translation start site with the initiator tRNA carrying the amino-terminal methionine (Met-tRNAi^{Met}). Then the large subunit and multiple initiation factors (eIF2, 3, 4, 5) also bind to form initiation complex. The complex precedes the scan along mRNA ($5' \rightarrow 3'$) until it encounters the start codon AUG. The anticodon (5′CAU3′) of Met-tRNAi^{Met} is base-paired with the start codon AUG in mRNA at P-site of ribosome.

During chain elongation, each incoming aminoacyl-tRNA moves through three ribosome sites, A, P, and E. First, the new aminoacyl-tRNA binds to the A-site that makes Met-tRNAi^{Met} move to the P-site, the large rRNA subunit catalyzed peptide bond formation between Met and incoming amino acid. At the same time, the ribosome undergoes conformational change and moves one codon down along the mRNA, the unacylated tRNAi^{Met} is shifted to the E-site from the P-site and the peptidyl-tRNA from the A-site to the P-site. In the next step, the incoming aminoacyl-tRNA binds to the A-site and unacylated tRNAi^{Met} is ejected from the E-site. The cycle repeats and the chain is elongated. This process is fast and accurate. A peptide of 100 amino acid residues needs only 5 s to synthesize (Fig. 2-10).

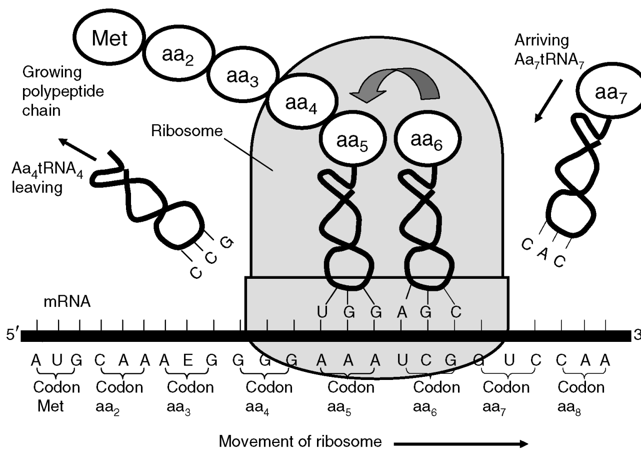

Figure 2-10 Translation. Translation is initiated from formation of initiation complex with fMet-tRNA^{met} binding to start codon AUG in mRNA. The second aminoacyl-tRNA enters the A-site of the ribosome. The peptidyl transferase of rRNA ribozyme in large subunit catalyzes the peptide bond formation between fMet-tRNA^{met} and the second aminoacyl-tRNA, The ribosome moves one codon toward the 3′-end of mRNA. The third incoming aminoacyl-tRNA binds the A-site of ribosome. The cycle is repeated and newly synthesized peptide is elongated until it hits the stop codons. The translation is then terminated. Many protein factors are involved in translation.

At the termination phase, when the peptide chain-bearing ribosome reaches a stop codon (UAA, UGA, UAG), the elongation stops since no tRNA molecules with anticodons matches the stop codons. The release factor eRF1 then enters the ribosomal complex and cleaves the peptide chain from tRNA at P-site. The peptides spontaneously fold into the active three-dimensional forms. The tRNA and two ribosomal subunits are dissociated and released. The ribosome is recycled for the next round of translation. In most reactions, GTP binding proteins hydrolyzing GTP to GDP provide the energy.

2.4.1.4 Posttranslational Modification of Proteins Although proteins are the end products of the genetic information flow, they need additional modifications to become the functional molecules. Posttranslational modification means the chemical modification of a protein after being translated. It is the last step in protein biosynthesis for many functional proteins. Posttranslational modification may involve the formation of disulfide bridges and attachment of any number of biochemical functional groups, such as carbohydrates, acetate, phosphate, and various lipids. Enzymes may also remove one or more amino acids from the ends of the polypeptide chain or cut the polypeptide in the middle of the chain. For instance, proinsulin is cut twice after disulfide bond formation to form the active form of insulin. In other cases, two or more polypeptide chains that are synthesized separately may associate to form the quaternary large protein, such as immunoglobulin and hemoglobulin. The most common posttranslational modification is glycosylation by which the carbohydrate chains are added to the side chains of the peptides to form glycoprotein. The carbohydrate chains are important for cell–cell recognition (sugar code). Notably, glycosylation is absent in bacteria and is somewhat different in each type of eukaryotic cells. Protein phosphorylation is part of common mechanisms for controlling the function of a protein, for instance, activating or inactivating an enzyme by protein kinase in cell signaling pathways (see Fig. 2-12). All proteins are eventually degraded by ubiquitin-dependant proteolysis in a large protein complex, proteasome.

2.4.2 Epigenetic Inheritance

In classic genetic inheritance, traits are passed from one generation to the next via DNA sequences in the genome. Differences in a DNA sequence specify differences in a trait. Epigenetic inheritance involves passing a trait from one generation to the next without the difference in DNA sequence, but DNA structure. Known mechanisms of epigenetic inheritance include changes in molecular structures in the DNA (such as DNA methylation) or histones (such as histone methylation and acetylation), chromosome remolding, and RNA interference (RNAi) so that while the gene (DNA sequence) is the same, the gene expression is different (Fig. 2-11). For example, genes switch on and off in response to hormonal signals. Changes in molecular conformation around the gene can influence the gene transcription. This can change developmental processes and can alter the course of diseases or result in genomic imprinting. The study of epigenetic inheritance is known as epigenetics.

Epigenetic inheritance systems allow cells of different phenotype but of identical genotype to transmit their phenotype to their offspring. Proteins or chemical groups

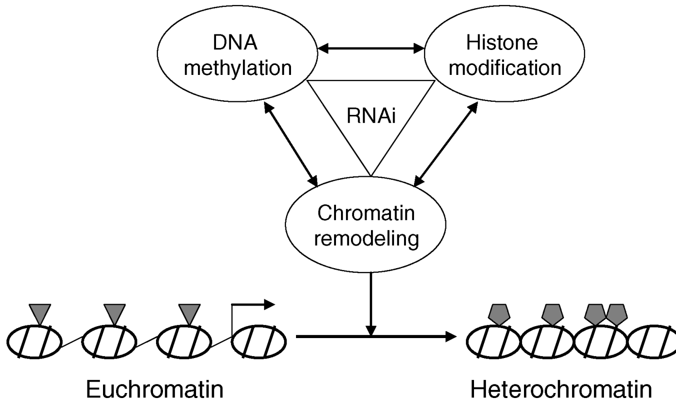

Figure 2-11 Cell epigenetics and gene silencing mechanisms. Cell epigenetic system consists of DNA methylation and histone methylations (pentagon), histone acetylation (triangle), chromatin (nucleosome) remodeling, and RNAi. Interaction between these components results in transition of euchromatin to heterochromatin and the transcription is inactivated (silenced) permanently.

that are attached to DNA and modify its activity are called chromatin marks. These marks are copied when DNA replicates. For example, cytosines in eukaryotic DNA can be methylated (5-methylcytosine). The number and pattern of such methylated cytosines influence the functional state of the gene: low levels of methylation correspond to high level of gene expression, whereas high levels of methylation correspond to low levels of gene expression. Although there are random changes in the DNA methylation pattern, specific changes induced by environmental factors do occur. After DNA replication, maintenance DNA methyltransferases (DNMT 1) at replication fork make sure that the methylation pattern of the parental DNA is copied to the daughter strand of DNA precisely. In such a way, the pattern of DNA methylation is maintained from parental cells to the daughter cells. If the DNA methylation occurs in the germ cells, the pattern can be inherited from one generation to the next. In a few cases, expression of a gene solely depends on whether it is inherited from the mother or father. This phenomenon is called genomic imprinting. The molecular mechanism of genomic imprinting is DNA methylation. The insulin-like growth factor-2 (Igf-2) gene is one example of an imprinting gene. In this case, only the copy of Igf-2 gene from paternal side is transcribed, whereas the maternal copy of Igf-2 gene is silenced by DNA methylation. The loss of Igf-2 imprint is associated with carcinogenesis, especially in colorectal cancer.

During the course of evolution, accidental deamination of unmethylated C gives rise to U, which is recognized by DNA repair system, uracil DNA glycosylase. The U in DNA sequence is excised, then replaced with C, and again restored in the original DNA sequence. However, deamination of a methylated C in the genome tends to be eliminated and replaced by a T by DNA repair system. As a result, most dinucleotide C–G sequences have been lost because of the elimination of methylcytosine. The remaining residual C–G sequences are distributed very unevenly in the genome.

In some regions, dinucleotide C–G sequences are present at 10–20 times more than their average density, called CpG islands. The CpG dinucleotide differs from C–G base pairing. The CpG islands are often located in the promoters of the housekeeping genes. The housekeeping genes encode many proteins that are essential for cell basic metabolism and viability and are therefore expressed in most cells all of the time. In most cases, the CpG islands seem to remain in an unmethylated state in all cell types. However, some tissue-specific genes, which code for proteins needed only in selected types of cells, are also associated with CpG islands. For instance, DNA methylation in the dividing fibroblasts gives rise to only new fibroblasts rather than some other cell types, even though the genome is identical in all cells.

If the CpG islands in promoter regions are abnormally methylated, the transcription will be blocked by the binding of methyl binding protein complex and the gene will be silenced. In the case of cancer cells, many tumor suppressor genes are inactivated by DNA methylation in CpG islands surrounding the promoters. Abnormal epigenetics including DNA methylation pattern is a hallmark of cancer cells (Fig. 2-11).

2.4.3 Cell Signaling and Integrated Controls

As we have discussed so far, there are four basic components in a living cell: functional structure, metabolism, energetic transfer, and information flow. The information flow can be divided into genetic flow and biochemical signaling. The integration of these two pathways results in the final control of the system. The genetic flow has been discussed in detail. This section will focus on the biochemical signaling.

For single-celled organisms to survive, cells must sense the changes in their environment and make adaptive responses constantly. These responses include a movement toward the nutrients or away from the toxin, changes of the metabolic patterns, and induction of certain protein expression. The *lac* operon in bacterial genome is an example of this type of control (Fig. 2-8a).

For a multicellular organism, cells not only need to adapt to their surrounding environment but also need to communicate with their neighboring cells and to adjust their behaviors and function to fit the whole system needs. The human body consists of about 50 trillion cells with 200 different cell types; the whole body control is carried out by a nerve system and an endocrine system conducted by cell signal transduction or cell signaling. In general, a cell signaling pathway consists of seven components: ligand, receptor, transducer, effector, second message, amplifier, and target (Fig. 2-12a). The ligands can be a hormone, neurotransmitter, growth factor, or even a gas (such as nitric oxide or NO). The concentration of the ligands is usually extremely low and it requires very specific binding to its receptor. There are various types of cellular receptors, including nuclear receptors and cell surface receptors. The nuclear receptors are proteins specific for steroid hormone binding. The cell surface receptors usually are integrated cell membrane proteins consisting of extracellular domain, transmembrane domain, and cytoplasmic domain. There are at least seven superfamilies of surface receptors binding different types of ligands to conduct different signaling pathways (Fig. 2-12b). The specificity of binding to the ligand is determined by the specific three dimensions of extracellular domain, while the cytoplasmic

domain usually has kinase activity. The transducer transmits the signal from the membrane into the cells. The various G-proteins functioning as the transducer trigger the effecter enzymes to produce the second message. The second message is usually small molecules, such as cAMP, cGMP, Ca^{2+}, inositol 1,4,5-triphosphate (IP3), or 1,2-diacylglycerol (DAG). An amplifier is a series of protein kinases, such as protein kinases A, B, C, or tyrosine protein kinase. The signal is amplified by a cascade of kinase reactions. The target is either an enzyme, ion channel, or a gene. Signal transduction research is concerned with the mechanisms by which cells receive, interpret, integrate, and act upon information received. There are extensive interactions or cross talks between the pathways. The final output of cell signaling is the result of changes of cell metabolism, function, or gene expression.

There has been an explosion of information in recent years related to the signal transduction mechanisms whereby cell surface receptors transmit external stimuli, delivered in the form of hormonal or other environmental cues, to the intracellular response machinery in the cytoplasm and nucleus. Particularly, it is now widely recognized that signal transduction abnormalities involving the changes of gene

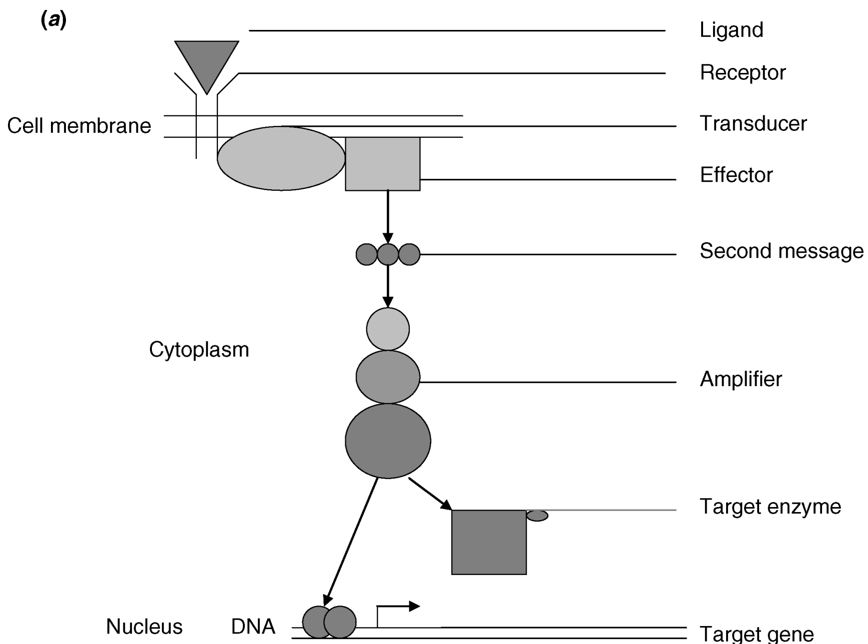

Figure 2-12 Cell signal transduction pathways. (a) A simplified model of a cell signal transduction pathway includes multiple components: ligand, receptor, transducer, effector, second message, amplifier, and targets. A ligand molecule binds to a receptor protein, thereby activating an intracellular signaling pathway that is mediated by a series of signaling proteins. Finally, one or more of these intracellular signaling proteins interacts with the target proteins, either in cytoplasm or nucleus, altering cell metabolism or gene expression. (b) Examples of signal transduction pathways in eukaryote. These pathways are important for cell metabolism and gene regulation. A detailed description can be found in the references.

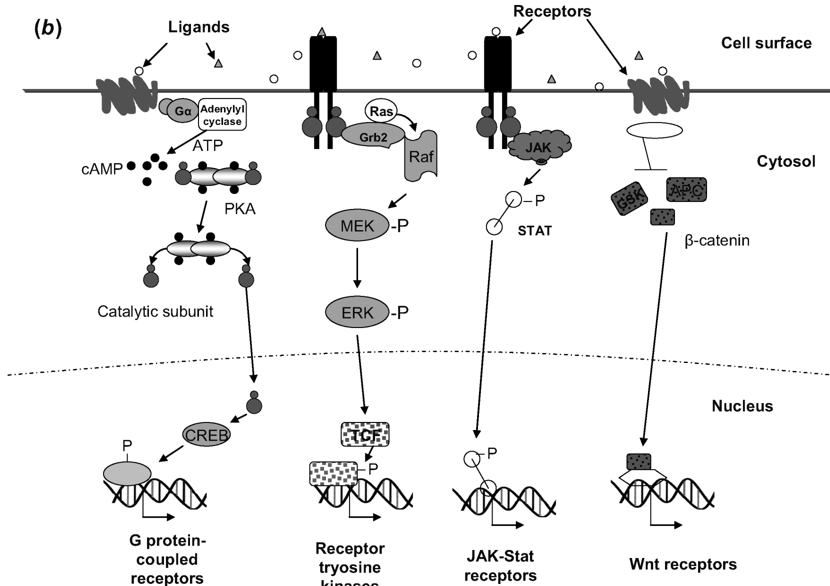

Figure 2-12 (*Continued*).

expression play important roles in several major human diseases. These findings have fueled a massive scientific effort aimed toward the identification and functional dissection of various signaling pathways. Furthermore, a substantial proportion of the current effort in modern drug discovery is founded on the premise that pharmacologic manipulation of signaling proteins will prove beneficial in the prevention and treatment of major human diseases.

2.5 GENETIC ENGINEERING OF LIVING CELLS

The term genetic engineering refers to the process of manipulating genes, usually outside the organism's normal reproductive process. It often involves the use of recombinant DNA technologies for the isolation, manipulation, and reintroduction of DNA into cells or model organisms and ultimately to express a protein. The goal is to introduce new characteristics such as increasing the yield of a crop species, introducing a novel trait, or producing a new protein or enzyme. The completions of the sequencing of the human genome, as well as the genomes of many agriculturally and scientifically important plants and animals, have significantly increased the opportunities for genetic engineering research. Expedient and inexpensive access to comprehensive genetic data has become a reality, with billions of sequenced nucleotides already online and annotated. Genetic engineering has become the gold standard in biotechnology research, and major research progress has been made using a wide variety of techniques.

2.5.1 Recombinant DNA Technology

Recombinant DNA technology is a set of techniques used for cutting apart and splicing together different pieces of DNA. The pieces of foreign DNA are introduced into another cell or organism and continue to produce their own coded proteins or substances within the new host cell. The cell becomes a factory for the production of these foreign proteins. The techniques for the introduction of foreign genes into bacteria were first developed in the early 1970s. In 1978, Herbert Boyer used recombinant DNA technology to produce recombinant human insulin, the first product of biotechnology, as we know it today. Although many methods are available, the basic procedure of making recombinant DNA involves the following steps (see Fig. 2-13 for an overview):

Isolating DNA. The first step in making recombinant DNA is to isolate donor and vector DNA. The procedure used for obtaining donor and vector DNA depends on the nature of the resource. The bulk of DNA extracted from the donor will be nuclear genomic DNA in eukaryotic cells or the main genomic DNA in prokaryotic cells; these types of DNA are generally the ones required for analysis. Bacterial plasmids are commonly used vectors, and these plasmids must be purified from the bacterial genomic DNA. DNA isolation is now simple with various commercial kits allowing a fast and chemically safe technique.

Cutting and Joining DNA. The cornerstone of recombinant DNA technology is a class of bacteria enzymes called restriction endonucleases. Restriction enzymes recognize a specific nucleotide sequence and cut both strands of the DNA within that sequence. To date, over 3000 restriction crosses have been identified and they can be found in the restriction enzyme database (REBASE, http://rebase.neb.com/rebase/). The first restriction enzyme named *EcoRI* was identified in *E. coli*. The DNA fragments produced by *EcoRI* digestion have overhanging single-stranded tails (called sticky ends) that reanneal with complementary single-stranded tails on other DNA fragments. If two pieces of DNA digested with same restriction enzyme are mixed under the proper conditions, DNA fragments from two sources form recombinant molecules by hydrogen bonding of their sticky ends. The enzyme DNA ligase covalently links these fragments to form recombinant DNA molecules.

Vectors or Plasmids. "Vector" is a carrier DNA molecule that can bring a foreign DNA fragment into a host cell. "Cloning vector" is used for reproducing the DNA fragment and "expression vector" is used for expressing certain genes in the DNA fragment. Commonly used vectors include plasmid, Lambda phage, cosmid, and yeast artificial chromosome (YAC). Vector DNA molecules are able to independently replicate themselves and the DNA segment they carry. They also contain a number of restriction enzyme cleavage sites that are present only once in the vector. One site is cleaved with a restriction enzyme and is used to insert a DNA segment cut with the same enzyme. Vectors usually carry a selectable marker such as antibiotic resistance genes or genes for

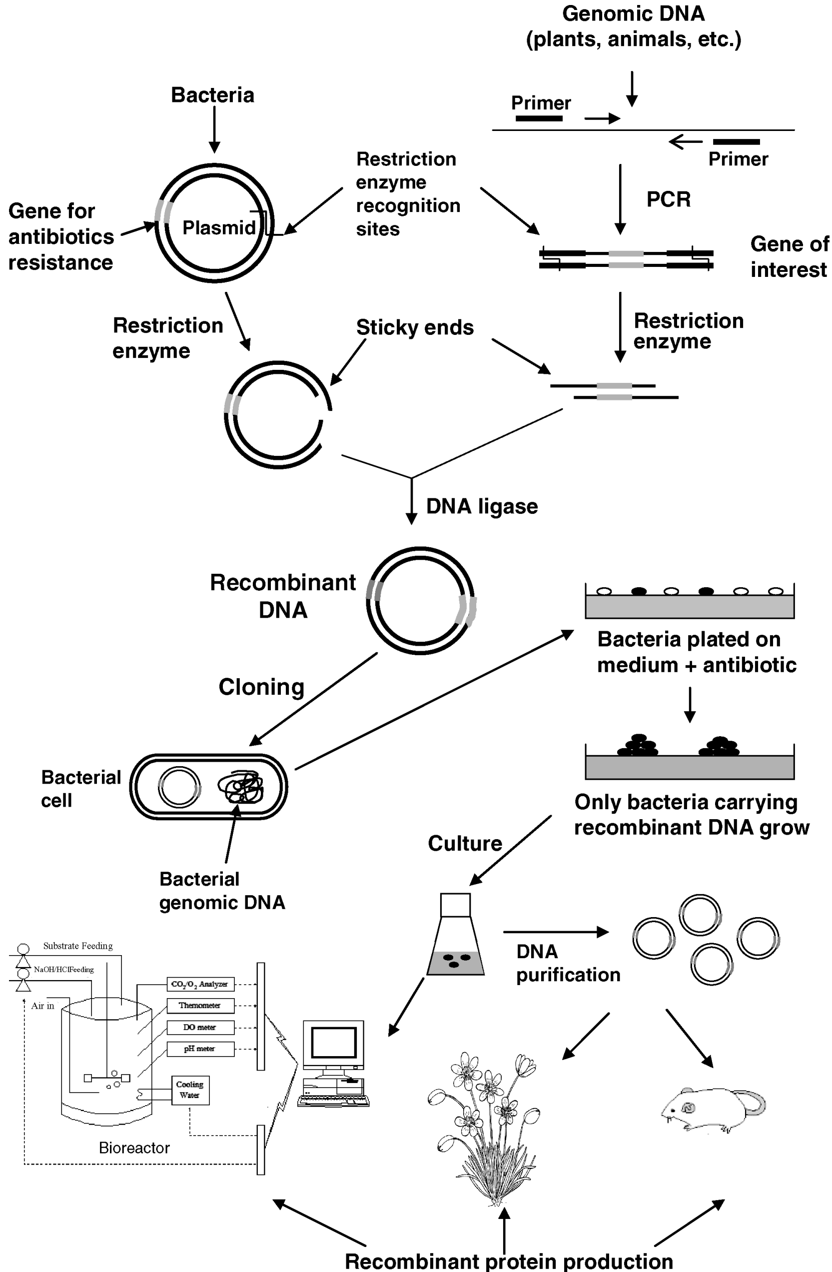

Figure 2-13 The recombinant DNA technology enables individual fragments of DNA from any genome to be inserted into vector DNA molecules such as plasmids and transformed into bacteria. Each of such recombinant DNA molecules can then be used for the production of foreign proteins in microorganisms and for creating transgenic plants and animals.

enzymes missing in the host cell. These markers can distinguish host cells that carry vectors from host cells that do not carry vectors. Many genetically engineered plasmid vectors are now available, and certain features make it easier to identify host cells carrying a plasmid with an inserted DNA fragment. Although only a single plasmid may enter a host cell, many plasmids can replicate themselves in the host cell so that several hundred copies are present. When used as vectors, such plasmids allow more copies of cloned DNA to be produced.

DNA Library. DNA library is a collection of many cloned DNA fragments. There are two types of DNA library. The genomic library is made of DNA fragments representing the entire genome of an organism. The cDNA library is generated from complementary DNA molecules synthesized from mRNA molecules in a cell. Therefore, the cDNA library contains only the coding region of a genome. To prepare a cDNA library, the first step is to isolate the total mRNA from the cell type of interest. Because eukaryotic mRNAs consist of a poly-A tail, they can easily be separated. A DNA strand complementary to each mRNA molecule is then synthesized by the enzyme called reverse transcriptase. After the single-stranded DNA molecules are converted into double-stranded DNA molecules by DNA polymerase, they are inserted into vectors and cloned.

Polymerase Chain Reaction. PCR is an enzyme reaction that targets a segment of DNA and then produces multiple amounts of the same segment; it is based on the ability of a DNA polymerase enzyme that can synthesize a complementary strand to a targeted segment of DNA. The PCR reaction mixture contains appropriate amounts of four deoxyribonucleotides and two short DNA oligonucleotides (each about 20 bases long), called primers, which have sequences complementary to areas adjacent to each side of the target sequence. If chosen well, the primer sequences will be unique in the entire genome and match only the place specifically chosen, thus limiting and defining the area to be copied. Figure 2-14 illustrates step-by-step the process of PCR reaction. Repeated heating and cooling cycles multiply the target DNA exponentially, since each new double strand separates to become two templates for further synthesis. Theoretically, the number of target sequences produced equals 2^n; that is, 20 PCR cycles can amplify the target by a million fold.

DNA Sequencing. The ability to sequence cloned recombinant DNA has greatly enhanced our understanding of gene structure, gene function, and the mechanisms of regulation. The most common method of DNA sequencing (Sanger sequencing) is based on dideoxy chain termination. In this procedure, a single-stranded DNA molecule whose sequence is to be determined is extended by DNA polymerase, similar to elongation during DNA replication. In addition, each tube contains a small amount of one of the four base-specific analogues called dideoxynucleotides (e.g., ddCTP). When ddCTP is incorporated into the extension, termination takes place at each of the C nucleotides in the newly synthesized DNA. A similar stop occurs with the other ddNTPs (Fig. 2-15). As the reaction proceeds, the tubes accumulate a series of DNA molecules that

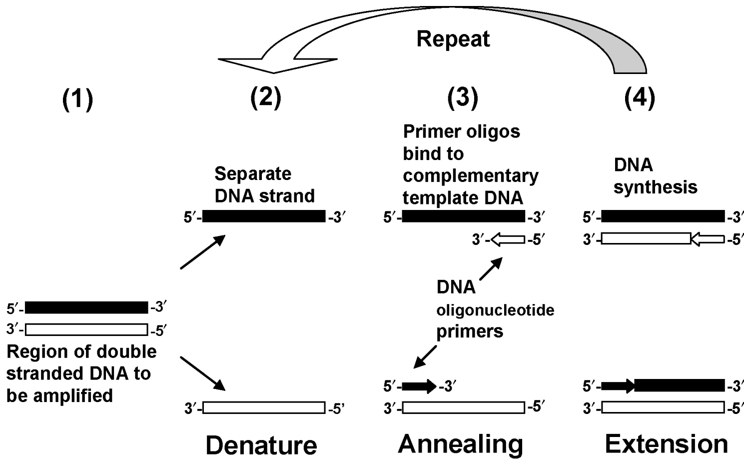

Figure 2-14 The polymerase chain reaction. (1) Double-stranded DNA containing the target sequence. (2) The strands are separated by heating at 94°C. (3) Two primers have sequence complementing primer binding sites at the 3′-ends of the target gene on the two strands. The DNA is allowed to cool to about 55°C, which allows the primers to stick to the single-stranded DNA at either end. (4) Taq polymerase then synthesizes the first set of complementary strands in the reaction. The final product is double-stranded DNA, which comes from the region defined by the primers. Steps 2–4 can be repeated to produce (in theory, if the process is 100 percent efficient) 2^n times the amount of template DNA (where n = number of cycles).

differ in length by one nucleotide at their 3′-ends. The fragments from each reaction tube are separated in four adjacent lanes (one for each tube) by gel electrophoresis (Fig. 2-15). Electrophoresis separates DNA fragments in each lane that differ in size by a single nucleotide. The nucleotide sequence of the DNA can be read directly from bottom to top, corresponding to the 5′–3′ sequence of the DNA strand complementary to the template. The newly developed florescence labeling and multiple channel capillary electrophoresis technology have greatly enhanced the speed and capacity of sequencing. Most recently, massively parallel sequencing technologies developed by several companies such as 454 Life Science, Solxa, and Applied Biosystems have dramatically increased the sequencing throughput. Now it is possible to sequence a bacterial genome in several hours to several days with the automated sequencing machine.

2.5.2 Genetically Modified Microorganisms

Organisms containing introduced foreign DNA in their genome are referred to as being *transgenic*. The introduced foreign gene is called a *transgene*. Hence, bacteria containing eukaryotic gene are transgenic bacteria. A major use for many of these transgenic microorganisms is to produce proteins that have immense commercial value. Numerous studies have focused on finding ways to produce them efficiently and in a functional form. As illustrated in Figure 2-13, the gene encoding a foreign protein

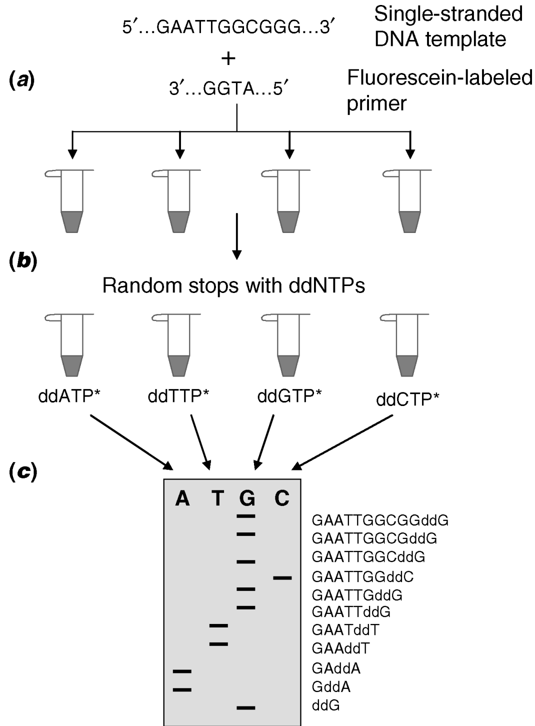

Figure 2-15 DNA sequencing with the dideoxy chain termination method. Single-stranded DNA template and single-stranded primer labeled with fluorescein are mixed and aliquoted into four tubes.(a) In the presence of DNA polymerase and a mixture of the four deoxynucleotides (dGTP, dATP, dTTP, and dCTP), primer extension occurs from the primer/template annealing site. (b) Random stops in extension are then generated by adding to each tube one of the dideoxynucleotides (ddNTP). The reaction results in a mixture containing variable lengths of extended DNA segments. (c) Finally each reaction mixture is separated electrophoretically in the gel track (or a capillary electrophoresis system) corresponding to the dideoxynucleotides added. As illustrated in the first lane, ddATP will produce two random stops where there is an adenine nucleotide and so two double-stranded DNA products are formed at corresponding sites. The remaining three dideoxynucleotides will do likewise in their individual reactions. The DNA sequence is read from bottom to top as illustrated in the figure.

can be cloned into an expression vector and transformed *E. coli* for protein synthesis under the control of a specific promoter, that is, *lac* promoter. Now various expression systems are available for overproducing foreign proteins in *E. coli*. The first human gene product manufactured using recombinant DNA and licensed for therapeutic use was human insulin, which is produced in *E. coli*. There are hundreds of therapeutic recombinant protein products in the market since then.

The use of genetically engineered microorganisms is a cost-effective, scalable technology for the production of recombinant proteins. However, the production of a functional protein is intimately related to the cellular machinery of the organism

producing the protein. Posttranslational modifications usually performed in higher eukaryotes, for example, correct folding, disulphide bond formation, *O*- and *N*-linked glycosylation, and processing of signal sequences are absent in the bacterial expression system. Therefore, eukaryotic expression systems have been developed, such as yeast (*S. cerevisiae* and *Pichia pastoris*), the insect cell/baculovirus system, and the plant cell and mammalian cell culture systems.

2.5.3 Transgenic Plants

Because of their economic significance, plants have long been the subject of genetic engineering aimed at developing improved varieties. Progress is being made on several fronts to introduce new traits into plants using recombinant DNA technology. Since it is able to grow a whole plant from a single cell, researchers can engage in the genetic manipulation of the cell, let the cell develop into a completely mature plant, and examine the whole spectrum of physiological and growth effects of the genetic manipulation within a relatively short period of time. The most commonly used cloning vector for making transgenic plants is the "Ti" plasmid. This plasmid is carried by the bacterium known as *Agrobacterium tumefaciens*, which has the ability to infect plants. When these bacteria infect a plant cell, a 30,000 base pair segment of the Ti plasmid, called T DNA, separates from the plasmid and incorporates into the host cell genome (Fig. 2-16). Therefore, the Ti plasmid can be used to shuttle exogenous genes into host plant cells. Foreign genes such as bacterial, plant, or mammalian DNA engineered with plant regulatory elements can be inserted into the Ti plasmid and then be placed back into the *A. tumefaciens* cell. That cell can be put into plant cells either by the process of infection or by direct insertion. The foreign DNA (T DNA and the inserted gene) can be incorporated into the host plant genome and passed on to future generations of the plant. For the plant cell types that are not susceptible to *A. tumefaciens* transfection, naked DNA molecules can be delivered into the target cells by using other gene delivery methods such as microinjection, electroporation, and particle bombardment, which will be discussed subsequently. These developments, important in the commercial application of plant genetic engineering, render the valuable food crops of corn, rice, and wheat susceptible to a variety of manipulations by the techniques of recombinant DNA and biotechnology. In recent years, progress has been made to improve nutritional quality, increase insect, disease, and herbicide resistance, and salt tolerance.

Moreover, the production of foreign proteins in transgenic plants has become a viable alternative to conventional production systems such as microbial fermentation or mammalian cell culture. Transgenic plants, acting as bioreactors, can efficiently produce recombinant proteins in larger quantities than those produced using mammalian cell systems. Plant-derived proteins are particularly attractive, since they are free of human diseases and mammalian viral vectors. Large quantities of biomass can be easily grown in the field and may permit storage of material prior to processing. Thus, plants offer the potential for efficient, large-scale production of recombinant proteins with increased freedom from contaminating human pathogens. A wide variety of other therapeutic agents have been derived from plants, including antibodies,

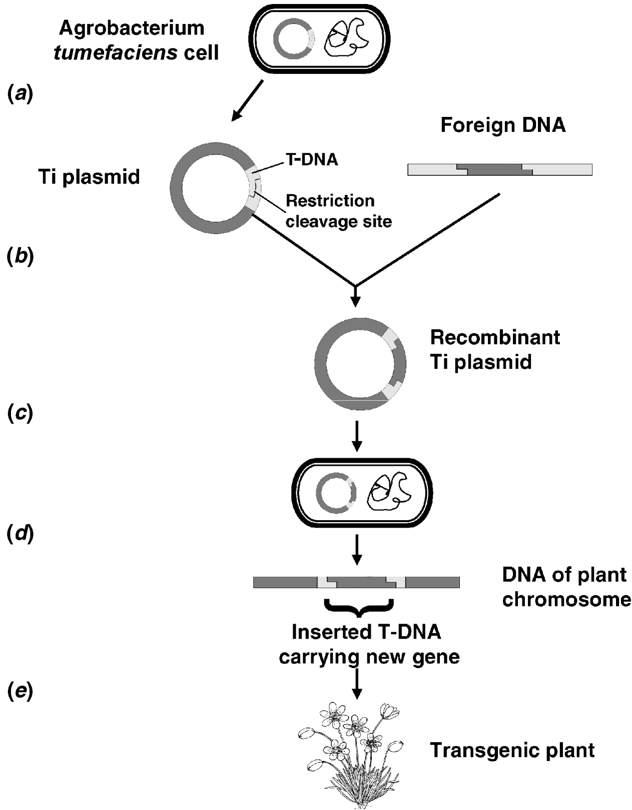

Figure 2-16 Generation of a transgenic plant through the growth of a cell transformed by T-DNA. (a) The Ti plasmid is isolated from *Agrobacterium* cells, (b) subjected to standard recombinant DNA procedures to insert the desired DNA into the T-DNA region of the plasmid, and then (c) put back into *Agrobacterium*. (d) Cultured plant cells are infected with bacteria containing the recombinant plasmid, and (e) these plant cells are then used to regenerate whole plant. The resulting transgenic plants contain the recombinant T-DNA region stably integrated into the genome of every cell. (modified with permission from Becker WM, Reece JB, and Poenie MF. *The World of the Cell*, 3rd ed. Copyright 1996 The Benjamin/Cummings Publishing Company).

vaccines, hormones, enzymes, interleukins, interferons (IFN), and human serum albumin (HSA).

2.5.4 Transgenic Animals

A transgenic animal is one that carries a foreign gene that has been inserted into its genome. The foreign gene is constructed using recombinant DNA methodology. Transgenic sheep and goats that express foreign proteins in their milk have been produced. Transgenic chickens are now able to synthesize human proteins in the egg

whites. These animals should eventually prove to be valuable sources of proteins for therapeutic purpose.

Mice are typically the most important models for mammals. Much of the general technology developed in mice can be applied to humans and other animals. Transgenic mice are produced by microinjection of recombinant DNA into the pronucleus of a fertilized oocytes. Figure 2-17 illustrates step-by-step the generation of a transgenic mouse. First, the DNA molecule containing the gene of interest (e.g., the insulin gene) is constructed by using recombinant DNA methods. The vector DNA also contains promoter and enhancer sequences to enable the gene to be expressed by host cells. The recombinant DNA molecule is then transfected into the cultured embryonic stem (ES) cells and successfully transfected ES cells will be selected. Second, the transformed

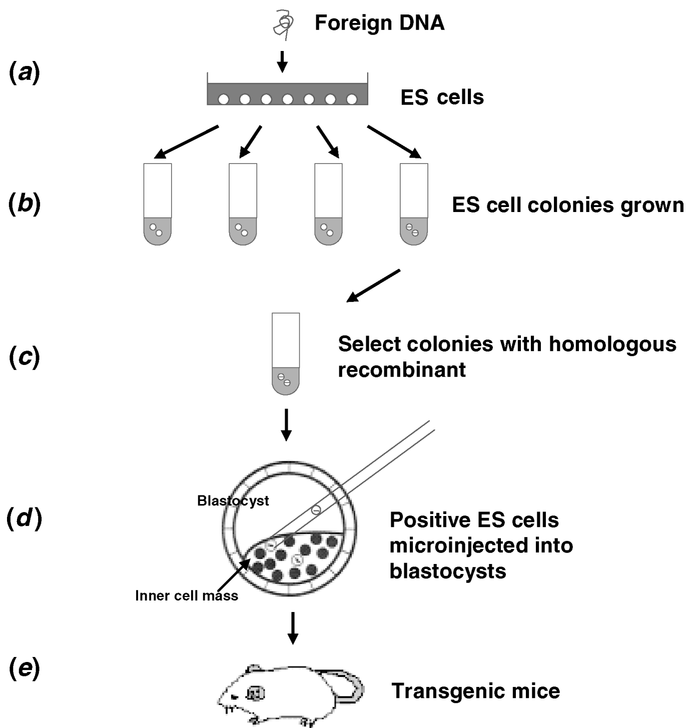

Figure 2-17 Generation of transgenic mouse carrying recombinant DNA.(a) Embryonic stem cells are transfected with foreign DNA. Many ES cells will take up the DNA, but this will involve different sites in the mouse genome because of random integration. In a very rare case, the integration will involve the correct part of the genome by a process of homologous recombination. (b) Colonies of ES cells are grown. (c) DNA is isolated from pools of colonies. The colony that has DNA integrated into the corrected position in the genome by homologous recombination can be identified by PCR. (d) ES cells that have the homologous recombined DNA are injected into mouse blastocysts. (e) If the transgene has also integrated into the germline, then some transgenic eggs or sperm will be produced and the next generation of mice will be fully transgenic—where every cell contains a copy of the foreign DNA (modified with permission from Trent RJ. *Molecular Medicine*, 3rd ed. Copyright 2005 Elsevier Academic Press).

ES cells are microinjected into an early embryo. The resulting progenies are chimeric, having tissue derived from either recipient or transplanted ES cells. The early vectors used for gene insertion could place the gene (from 1 to 200 copies of it) anywhere in the genome. However, if you know some of the DNA sequence flanking a particular gene, it is possible to design vectors that replace the gene. The replacement gene can be one that restores function in a mutant animal or knocks out the function of a particular locus.

All genes are associated with specialized DNA sequences such as promoters and enhancers that support transcription. Once in the nucleus, exogenous DNA requires a strong promoter and enhancer upstream of the transgene for its expression. The human cytomegalovirus (CMV) promoter has been used extensively to drive the expression of transgenes in mammalian cells. However, these viral promoters result in uncontrolled expression of transgenes. The uncontrolled expression of the transgene during embryonic development could also be lethal if it is not normally expressed at this time. Attempts controlling gene expression in transgenics involve the use of inducers, for example, chemical or hormonal signaling molecules. This was possible because the inserted gene had a promoter element that was inducible when exposed to the signal. There are also techniques with which transgenic mice can be made where a particular gene gets knocked out or introduced in only one type of cell.

2.5.5 New Tools for Gene Knockdown

The sequence-specific interaction of short nucleic acids with target RNAs or DNAs can be exploited as a tool for targeted inhibition of gene expression. These methods have been used in a wide variety of applications ranging from understanding the function of a given gene to molecular therapeutics. Although the mechanisms of these two methods differ, the problems for effective application are quite similar since both antisense oligonucleotides (AOs) and RNA interference are subject to artifacts such as off-target effects and CpG motif immune stimulation. It may be necessary to validate findings made using one method through use of a second approach.

2.5.5.1 Antisense Oligonucleotides The evidence that small AOs could specifically inhibit gene expression was discovered in 1978 by Zamecnik and Stevenson, who demonstrated that viral replication could be blocked by treating infected cells with an AO that was complementary to a portion of the viral mRNA. Since then, there have been thousands of published reports describing the application of AOs both as research tools and as medicines. Synthetic 15–25mer nucleotides could enter living cells and might be tailored to target and elicit RNAse H-mediated cleavage of mRNA molecules, and subsequently lead to reduced production of a target protein. Using this technology, researchers could quickly determine specific gene functions in organisms at any stage of their life cycle. Although AOs have became a great research tool, some difficulties had been anticipated, such as finding ways to make AOs resistant to extracellular nucleases that would degrade them before they entered cells and devising efficient means of gene delivery. Other challenges were less anticipated, including finding a systematic way to predict which portion of a selected

mRNA would make the best target. Still other hurdles came in the form of startling surprises. Foremost among them was the discovery that the DNA dinucleotide sequence C–G has immune stimulatory action that has nothing to do with the intended antisense effects.

2.5.5.2 RNA Interference As we discussed earlier, a newly discovered group of small RNA (miRNA and siRNA) regulate gene expression at posttranscriptional level (Fig. 2-9). RNAi refers to the inhibition of gene expression in a sequence-specific fashion by double-stranded RNA. Although this method has only been in routine use in mammalian systems for 7–8 years, it is currently being successfully used in thousands of labs and ambitious whole genome screening projects. The most straightforward approach to RNAi is use of duplex RNAs made by chemical synthesis. Like AOs, RNAi occurs when oligonucleotides form base pairs with target sequences within specific mRNAs to silence them. Although the RNAi processes are complex, the end results are conceptually similar to what occurs when using AOs: The presence of an oligonucleotide inhibits a specific mRNA function, leading to a decrease in the amount of protein translated by that mRNA. In contrast to the antisense process using AOs, however, RNAi uses synthetic RNAs to mimic naturally occurring processes. These natural processes involve, for example, genomically encoded microRNAs for fine-tuning the regulation of gene expression and siRNAs for potentiating certain host–microbe interactions. Researchers have uncovered examples of siRNAs or microRNAs acting at transcriptional and posttranscriptional levels. As molecules that interfere with posttranscriptional (i.e., translational) events, the microRNAs may interact with their mRNA targets either destructively or reversibly, depending upon the biological circumstance. Each of these new RNAi functions inspires novel applications for research and treatment tools. In addition, the fact that applied RNAi mimics real cellular processes has led some researchers to argue that harnessing RNAi for therapeutic uses has a much greater likelihood of succeeding than an unnatural process such as AO-mediated mRNA silencing. The level of enthusiasm for RNAi has been extraordinary. In 2002, *Science* proclaimed RNAi to be the "breakthrough of the year"—and that was just an early milestone.

2.5.6 Gene Delivery

Both viral and nonviral vectors can be used for delivering a gene into the cell nucleus. These vectors carry exogenous DNA or RNA (including siRNA and AOs) across the cell membrane into the nucleus to allow transcription. Since the nucleases in the endolysosomes and cytoplasm actively degrade free nucleic acid, the vectors have to escape from the degradation process before unpackaging their genetic materials and inserting into the host cell genome. The vectors also need to cross the plasma membrane into the nucleus of target cells, either by passive diffusion through the nuclear pore complex or by an energy-dependent translocation that requires a group of proteins called importins. Therefore, for a successful gene delivery, a number of cellular barriers must be overcome.

Viruses use multiple mechanisms to infect their host cells efficiently, either by fusion with the cell membrane or by receptor-mediated endocytosis, followed by nuclear localization of the viral genome. As a result, viral vectors are able to mediate gene transfer with higher efficiency and possibility of long-term gene expression. Most modern viral vectors are unable to replicate freely, owing to the deletion of essential genes, and carry little risk of proliferation or reversion to wild type. Their main drawback is that they tend to be immunogenetic and cause the acute immune response, which limits their *in vivo* potential. In addition, viral vectors can alter cellular function after transduction. The limitation in the size of the transgene that recombinant viruses can carry and issues related to the production of viral vectors present additional practical challenges.

Approaches of nonviral gene delivery have also been explored using physical and chemical approaches. Physical approaches, including needle injection, electroporation, gene gun, ultrasound, and hydrodynamic delivery, employ a physical force that permeates the cell membrane and facilitates intracellular gene transfer. For instance, electroporation uses brief pulses of high-voltage electricity to induce the formation of transient pores in the membrane of the host cell. Such pores appear to act as passage ways through which the naked DNA can enter the host cell; microinjection involves the direct injection of material into a host cell using a finely drawn micropipette needle. In addition, particle bombardment actually shoots DNA-coated microscopic pellets through a plant cell wall and delivers the foreign DNA into the host cell. The chemical approaches use synthetic or naturally occurring compounds as carriers to deliver the transgene into cells. For instance, the cell and nuclear membranes can be made more permeable to DNA following coprecipitation of DNA with calcium phosphate and DNA can also be packaged into cationic liposomes that are able to cross the cell membrane. Nonviral vectors are less able than their viral counterparts to overcome the problems of binding, escape from endosomes, uncoating, and transport into the nucleus. Thus, they possess lower transfection efficiency. As a result, new strategies to improve internalization and endosomal escape of nonviral vectors are being developed.

2.6 METABOLIC ENGINEERING OF LIVING CELLS

Metabolic engineering is a powerful approach to the understanding and utilization of metabolic processes and has become a new paradigm for the improvement of cellular properties or metabolite production. As the name implies, metabolic engineering emphasizes the targeted and purposeful modification of metabolic pathways found in an organism. Built largely on the theoretical and computational analysis of a biosystem, the field has embraced a growing number of genome-scale experimental tools. The rapid expansion of genomics information across various species and the integration of system biology approach have transformed our ability to carry out metabolic engineering approaches. As such, we are on the cusp of a new age of metabolic engineering involving many applications that address the new challenges in the 21th century including energy, pollution, global warming, food and human health.

2.6.1 Principle of Metabolic Engineering

Metabolic engineering can be defined as directed modification of cellular metabolism and properties through the introduction, deletion, and modification of metabolic pathways by using recombinant DNA technologies. Much of this effort has focused on microbial organisms, but important work has been done in plants, insects, and animals. In a global sense, this is not different from what genetic engineers have been doing for years with phenotypic improvements resulting from the manipulation of genes directly involved in creating the product of interest. However, with metabolic engineering, the focus is placed on understanding the larger metabolic network inside the cell in a systematic fashion. Thousands of chemical and biological reactions occur in a typical cell, which serve a multitude of purpose critical for maintaining cellular physiology and fitness within its environment; this reinforces the need for a systematic approach to understand the cellular activities as a whole. As we discussed earlier, various known and unknown regulatory (i.e., transcriptional, translational, enzymatic, signal transduction) mechanisms exist in a cell to manage and direct the resource to process that optimize cellular fitness. Thus, changing pathways that do not improve fitness or even detract from the fitness within a population often lead to relatively small improvements in product formation despite large increase in specific enzymatic activities. Metabolic engineering approaches embrace techniques that fill the gaps between genetic engineering and classical strain improvement. Metabolic engineers place the emphasis on understanding the mechanistic features that genetic modifications confer, thereby adding knowledge that can be used for relational approaches while searching the metabolic landscape.

Metabolic engineering also employs concepts from reaction engineering and thermodynamics for the analysis of biochemical reaction pathways. Although it shares common fundamentals with traditional biochemical engineering, the focus has shifted away from equipment to analysis of cells as integral units. Another novel aspect of metabolic engineering is the emphasis it places on integrated metabolic pathways as opposed to individual reactions. As such, metabolic engineering is concerned with complete bioreaction networks and issues of pathway synthesis, thermodynamic feasibility, and pathway flux and flux control. An enhanced perspective of metabolism and cellular function can be obtained by considering reactions in their entirety rather than in isolation from one another; this is of central importance in understanding the metabolic network. The main issue is amplification and/or redistribution of pathway flux. This is very different from chemical plant scale-up issues. Instead of increasing the capacity of a processing plant by increasing the capacity of its units, one now attempts at increasing the capacity of a single cell by amplifying the activity of some key enzymes.

Various metabolic engineering strategies have been widely applied for the more efficient production of desired metabolites and biomolecules. Metabolic engineering is an iterative process: cycle of genetic modification–analysis of metabolic consequences of change (identifying limitations)–choice of next genetic modification (see Fig. 2-18 for a hypothetical example). Measurement requires the ability to assay a large part of the network and extract as much information about the effect of an imposed network perturbation as possible. Another important part of metabolic

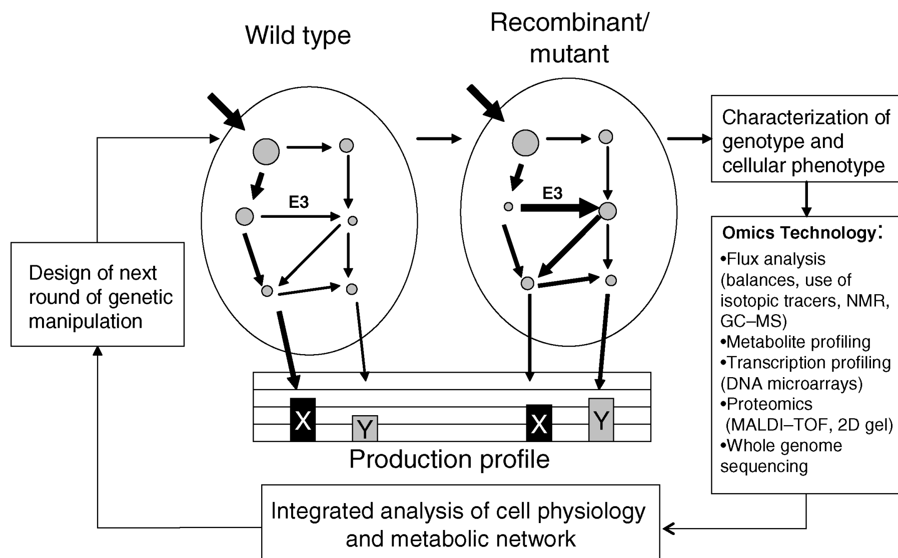

Figure 2-18 The iterative approach of metabolic engineering. Iterative perturbations and systematic phenotype and genotype characterizations yield system insight into the integration of high-throughput data sets. In this schematic, wild-type cells are engineered to overexpress enzyme E3 with the goal of increasing the low yield of product Y. However, because of network interactions, overexpression of E3 has a minimal effect on the accumulation rates of either products Y or X. To improve the yield of product Y, multiple steps in the network will have to be targeted and genetically modified. To identify these targets, various omics technologies will be used to generate an integrated profile of cellular networks in the recombinant or mutant strain. And then, comparative analyses of omics profiles are conducted to identify the target pathways for the overproduction of Y. Gene manipulation is carried out within the suggested candidate genes and characterizations of the new strain will be repeated. The gray circles indicate the pool size of metabolites in the network. Arrow thickness depicts relative flux magnitude of the corresponding reactions (modified with permission from Stephanopoulos G, et al. *Nature Biotechnology* 2004; 32:1262 Copyright 2004 Nature Publishing Group).

engineering is being able to perform the desired genetic perturbations efficiently. A variety of molecular engineering techniques are currently available to create gene deletion or overexpress genes of interest. However, it is essential for metabolic engineer to be able to precisely change the activities of certain enzymes in a desired pathway, as the desired change in activity may not be deletion (no activity) or overexpression driven by a strong promoter (order of magnitude change in activity). In some cases, a deletion is not possible as the enzyme is required for cell survival. Likewise, strong overexpression can result in deleterious outcomes such as the accumulation of toxic intermediates in a pathway.

2.6.2 Molecular and Computational Tools for Metabolic Engineering

Metabolic engineering relies upon methods that perturb the genome, measure fluxes, and analyze the state of the cell, such that the cell's network architecture can be

elucidated and effective targets for genetic manipulation can be identified. Both experimental and computational tools have been developed for identification of targets of metabolic engineering.

2.6.2.1 Mathematical Tools for Analysis *Metabolic flux analysis* (*MFA*) is based on a known biochemistry framework. A linearly independent metabolic matrix is constructed based on the law of mass conservation and on the pseudo-steady-state hypothesis on the intracellular metabolites. The accumulation rate of metabolites in the metabolic network may be expressed as

$$x_i = \sum a_j r_j \qquad (i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n) \tag{2-1}$$

where x_i is the accumulation rate of metabolite i, a_j is a stoichiometric coefficient, and r_j is the flux through reaction j. Equation 2-1 can be expressed in matrix form as

$$Ar = x \tag{2-2}$$

where A is an $m \times n$ matrix of stoichiometric coefficients, r is an n-dimensional flux vector, and x is an m-dimensional metabolite accumulation rate vector. Typically, the system that results is an underdetermined system where $m > n$. However, under certain conditions, some pathways are inoperative and can be neglected. The system may become completely determined or overdetermined and can be solved along with the measurements of external or internal fluxes. Determining the fluxes often requires the measurements made by incorporating ^{13}C-labeling; as the labeled substrate proceeds through the metabolic network, the pools of metabolites that are downstream from the substrate become labeled. At the steady state, the fraction of labeled substrate in a given pool can be used to calculate the flux through the pathway. Noninvasive methods of analysis such as nuclear magnetic resonance can also provide information on the structure of the biochemical network as well as flux measurements.

Metabolic control theory (*MCT*) was independently developed by Kacser and Burns and Heinrich et al. to identify the kinetic constraints in a biochemical network. With MCT, the control structures are quantified as mathematical formulation based on the so-called elasticity coefficients and control coefficients. Especially useful are the flux control coefficients (FCCs), which quantify the influence of the individual reaction rates (or enzyme activities) on the overall flux through the pathway. This approach is used to determine the rate-limiting reaction in a network. However, a single rate-limiting step may not exist and several steps may share the control of the metabolic network. The FCC of an enzyme is defined as the relative effect of modulating the amount of an enzyme on the flux through the desired pathway. Equation 2-3 shows the flux control coefficient C_i^J of an enzyme E_i on the flux J.

$$C_i^J = \frac{\mathrm{d}J}{\mathrm{d}E_i}\left(\frac{E_i}{J}\right) \tag{2-3}$$

The FCC is essentially a sensitivity coefficient of the flux with respect to various enzymes. An important property of the FCC is that summation of all the FCCs affecting a particular flux must equal unity (Eq. 2-4).

$$\sum_i C_i^J = 1 \tag{2-4}$$

An FCC that approached unity would imply a rate-limiting step. MCT is a powerful tool for qualitative studies of metabolic pathways. A serious drawback of the method is, however, the requirement that either the elasticity coefficients or the control coefficients have to be measured. This is not an easy task because it requires independent variation of the activity *in vivo* of all the enzymes within the pathway.

2.6.2.2 Metabolic Profiling and Metabolomics *Metabolomics* is the "systematic study of the unique chemical fingerprints that specific cellular processes leave behind"—specifically, the study of their small-molecule metabolite profiles. Small-molecule metabolites are critical in regulating transcriptional and translational processes and measuring the abundance of small metabolites provides a broad glimpse of the metabolic cellular state. Small molecules possess a wider range of chemical characteristics and are more difficult to measure. Complex profiles obtained using techniques such as nuclear magnetic resonance, liquid chromatography–mass spectrometry (LC–MS) and electrochemical array (EC) are typically analyzed for differences and changes in patterns of small molecules. Statistical analyses, using pattern recognition software even prior to identification of specific metabolites, allow a rapid means of finding specific markers for disease, toxicity, or some other process. Metabolic profiling typically involves the generation of patterns of analysts, containing both known and unknown compounds, to differentiate one sample group from another using these statistical analysis and pattern recognition software tools. However, unlike previously mentioned isotopic-labeling methods, metabolic profiling does not attempt to establish the intracellular flux, making these experiments more convenient. Nevertheless, it may be that the metabolite profiles provide enough similar information such that, when combined with protein and transcript profiles, a fairly complete picture of the cell is obtained that can be used to solve some more complex systemic problems.

2.6.2.3 High-Throughput Methods of Gene Manipulation The recent development in high-throughput methods of gene manipulation provides a way of rapidly screening for new targets of metabolic engineering. In the case of bacteria, the large libraries of knockout mutants can be generated quickly by using transposable elements and subsequently screened for improved physiological performance. High-efficiency transformations can generate libraries of as many as 10^9 genetic variants. Transposon mutagenesis also enables a high-throughput form of mutagenesis where there is only one mutation introduced per cell. In a similar manner for mammalian cells, large-scale screening techniques using genome-wide RNAi provide an

opportunity for metabolic engineering in human and animal cells. This technique can be used to ablate all specific genes in the genome of an organism. Bacteria library containing vectors expressing siRNA against all genes of several model organisms are now commercially available. This technique was found to be very efficient for the study of genes involved in cellular metabolism, confirming not only the several genes already known to be involved in this process but also finding several others not previously linked to the control of cellular metabolism in *Caenorhabditis elegans*.

2.6.3 Application of Metabolic Engineering

Metabolic engineering has many areas of applications. Introduction of new pathways enables us to use nature's diversity to meet human needs in a sustainable way. Strain improvement of microorganisms through metabolic engineering will improve productivity, lower costs, reduce environmental pollution, and generally improve results in a wide variety of industries and areas, including pharmaceuticals, chemical bioprocess, food technology, agriculture, environmental bioremediation, and biomass conversion. The prospective categories of chemicals to be produced by metabolically engineered microorganisms include nutriceuticals, fine chemicals, vitamins, preservatives, sweeteners, minerals, nutrisupplements, pharma intermediates, pharmaceuticals, amino acids, acidulents, cosmetics, and food ingredients.

Using metabolic engineering to redesign plants could potentially improve nutritional value of crops (e.g., essential amino acid supply for storage proteins, vitamin content, modifying lignin to enhance forage digestibility), create new industrial crops (e.g., modified fatty acid composition of seed triglycerides, pharmaceuticals, polyhydroxybutyrate synthesis), alter photosynthate partitioning to increase economic yield, enhance resistance to biotic and abiotic stresses such as infectious disease, and reduce undesired (toxic or unpalatable) metabolites. Furthermore, the potential exists to produce therapeutic proteins (i.e., antibody) in plants, which could eliminate the need for large-scale fermentation or cell culture facilities and could only require purification and formulation processes. There are many opportunities and challenges for metabolic engineers in this area, including increasing protein production, controlling glycosylation, and altering desirable metabolic pathway.

In addition to its application in industrial and agricultural biotechnology, metabolic engineering principles can also be applied to medical research and practice. For instance, flux measurements and metabolite profiling can be conducted on primary cells and/or body fluids isolated from patients' sample for investigating disease initiation, progression and treatment effect. These types of work may lead to the identification of surrogate markers for diagnosis or prognosis of certain diseases as well as molecular targets for new drug development. Metabolic profiling of urine or blood plasma samples can also be used to detect the physiological changes caused by toxic insult of a chemical or drug. Advances in this area promise to contribute to personalized medicine by incorporating patient-specific genetic and metabolite profiles.

2.7 SUMMARY

In this chapter, we first discussed the major types of biomolecules including nucleic acids, carbohydrate, proteins, and lipids; assembly of the biomolecules into subcellular level (organelles) and cell; and the information flow including genetic and epigenetic inheritance and biochemical signaling. We further discussed the major metabolic pathways of biomolecules, linked energy production (bioenergetics) and metabolism, and integrated controls of metabolism and gene expression. Finally, we discussed the application of molecular technology in industry, agriculture, and medicine. We highlighted the important new discoveries and developments in each field. We hope that it will provide a primer and a framework to understand systems biology. The postgenomic era presents new opportunities as new challenges to all fields related to biology. A variety of new technologies allow studying the basic biological processes in the central dogma at whole genome level. Comparative genomics, transcriptomics, proteomics, structural genomics, and metabolomics become active research subjects. However, how to integrate transcriptomics, proteomics, and metabolomics information to draw a complete picture of a cell or a given organism is one of the big challenges. Thus, systems biology emerged. It is likely that the use of postgenomic tools will allow identification of far more complicated functional interactions between protein–protein, protein–DNA, and even protein–metabolite. The incorporation of spectroscopic approaches for metabolic profiling and flux analysis combined with mathematical modeling will contribute to the development of rational metabolic engineering strategies and will lead to the development of new tools to assess temporal and subcellular changes in metabolite pools. New technologies for pathway engineering, including use of heterologous systems, directed enzyme evolution, engineering of transcription factors, and application of molecular/genetic techniques for controlling biosynthetic pathways will move the metabolic engineering approach to the next level. The accumulated new data will provide the basis for systems biology to interpret the function of a biological system at molecular level in the quantitative manner; this in turn will allow modification or redesigning of the biosystems for future applications.

ACKNOWLEDGMENTS

We thank Miss Angel Surdin for her expert editorial assistance and contributions to accomplishing this valuable book. We also thank Miss Elise C. Welsh for her help in preparation of the figures.

SUGGESTED READING

General

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell*, 4th ed. Garland Science, Taylor & Francis Group, 2002.

2. Allis CD, Jenuwein T, Reinberg D, Caparros M. *Epigenetics*. Cold Spring Harbor Laboratory Press, 2007.
3. Cooper GM, Hausman RE. *The Cell: A Molecular Approach*, 4th ed. ASM Press, Sinauer Associates, Inc, 2007.
4. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Darnell J. *Molecular Cell Biology*, 5th ed. W. H. Freeman and Company, 2004.
5. Nelson DL, Cox MM. *Lehninger Principles of Biochemistry*, 4th ed. W. H. Freeman and Company, 2005.

Biomolecules in Living Cells

Nucleic Acids

1. Arnhein N, Erlich H. Polymerase chain reaction strategy. *Annu Rev Biochem* 1992;61:131–156.
2. Brenner S, Jacob F, Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 1961;190:576–581.
3. Brenner S. The end of the beginning. *Science* 2000;24:2173–2174.
4. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* 2003;422:835–847.
5. Dickerson RE. The DNA helix and how it is read. *Sci Am* 1983;249:94–111.
6. Douand JA, Cech TR. The chemical repertoire of natural ribozymes. *Nature* 2002;418:222–228.
7. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931–945.
8. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
9. Lehman IR. Discovery of DNA polymerase. *J Biol Chem* 2003;278:34733–34738.
10. Lehman IR. Wanderings of a DNA enzymologist: from DNA polymerase to viral latency. *Annu Rev Biochem* 2006;75:1–17.
11. Mello CC, Conte D Jr. Revealing the world of RNA interference. *Nature* 2004;431:338–342.
12. Moore PB. Structural motifs in RNA. *Annu Rev Biochem* 1999;68:287–300.
13. Novina CD, Sharp PA. The RNAi revolution. *Nature* 2004;430:161–164.
14. Wang JC, Superhelical DNA. *Trends Biochem Sci* 1980;5:219–221.
15. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 1953;171:737–738.

Proteins

1. Ahn NG, Shabb JB, Old WM, Resing KA. Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem Biol* 2007;2:39–52.
2. Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 1998;92:291–294.

3. Berman HM. The past and future of structure databases. *Curr Opin Biotechnol* 1999;10:76–80.
4. Branden C, Tooze J. *Introduction to Protein Structure*, 2nd ed. New York: Garland Publishing, 1999.
5. Doolittle RF. The multiplicity of domains in proteins. *Annu Rev Biochem* 1995;64:287–314.
6. Dougherty DA. Unusual amino acids as probes of protein structure and function. *Curr Opin Chem Biol* 2000;4:645–652.
7. Gygi SP, Aebersold R. Mass spectrometry and proteomics. *Curr Opin Chem Biol* 2000;4:489–494.
8. Hazbun TR, Fields S. Networking proteins in yeast. *Proc Natl Acad Sci USA* 2001;98:4277–4278.
9. Jones S, Thornton JM. Principles of protein–protein interaction. *Proc Natl Acad Sci USA* 1996;93:13–20.
10. Khosla C, Harbury PB. Modular enzymes. *Nature* 2001;409:247–252.
11. Luque L, Leavitt SA, Freire E. The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu Rev Biophys Biomol Struct* 2002;31:235–256.
12. Marcotte EM, Pellegrini M,et al. Detecting protein function and protein–protein interactions from genome sequences. *Science* 1999;285:751–753.
13. Ponting CP, Russel RR. The natural history of protein domains. *Annu Rev Biophys Biomol Struct* 2002;31:45–71.
14. Resing KA, Ahn NG. Proteomics strategies for protein identification. *FEBS Lett* 2005;579:885–889.
15. Sali A, Kuriyan J. Challenges at the frontiers of structural biology. *Trends Cell Biol* 1999;9: M20–M24.
16. Sanger F. Sequences, sequences, and sequences. *Annu Rev Biochem* 1988;57:1–28.
17. Teichmann SA, Murzin AG, Chothia C. Determination of protein function, evolution and interactions by structural genomics. *Curr Opin Struct Biol* 2001;11:354–363.
18. Toyoshima C, Nakasako M, Nomura H, Ogawa H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 2000;405:647–655.
19. Vale RD. The molecular motor toolbox for intracellular transport. *Cell* 2003;112:467–480.
20. Walsh C. Enabling the chemistry of life. *Nature* 2001;409:226–231.

Polysaccharides

1. Bertozzi CR, Kiessling LL. Chemical glycobiology. *Science* 2001;291:2357–2363.
2. Esko JD, Selleck SB. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu Rev Biochem* 2002;71:435–471.
3. Gabius H-J, Andre S, Kaltner H, Siebert HC. The sugar code: functional lectinomics. *Biochim Biophys Acta* 2002;1572:165–177.
4. Gahmberg CG, Tolvanen M. Why mammalian cell surface proteins are glycoproteins. *Trends Biochem Sci* 1996;21:308–311.
5. Helenius A, Aebi M. Intracellular functions of N-linked glycans. *Science* 2001;291:2364–2369.

6. Loris R. Principles of structures of animal and plant lectins. *Biochim Biophys Acta* 2002;1572:198–208.
7. Selleck S. Proteoglycans and pattern formation: sugar biochemistry meets developmental genetics. *Trends Genet* 2000;16:206–212.
8. Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J. *Essentials of Glycobiology.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1999.

Lipids

1. Christie WW. *Lipid Analysis*, 3rd ed. Bridgwater, UK: The Oily Press, 2003.
2. Clouse SD. Brassinosteroid signal transduction: clarifying the pathway from ligand perception to gene expression. *Mol Cell* 2002;10:973–982.
3. Dowhan W. Molecular basis for membrane phospholipid: why are there so many lipids? *Annu Rev Biochem* 1997;66:199–232.
4. Lemmon MA, Ferguson KM. Signal-dependent membrane targeting by pleckstrin homology (PH) domains. *Biochem J* 2000;350:1–18.
5. Vance DE, Vance JE,editors. *Biochemistry of Lipids, Lipoproteins and Membranes*, New Comprehensive Biochemistry, Vol. 36. New York: Elsevier Science Publishing Co, Inc, 2002.
6. Weber H. Fatty acid-derived signals in plants. *Trends Plant Sci* 2002;7:217–224.

Microstructure and Function of Cells and Organelles

1. Altan-Bonnet N, Sougrat R, Liu W, Snapp EL, Ward T, Lippincott-Schwartz J. Golgi inheritance in mammalian cells is mediated through endoplasmic reticulum export activities. *Mol Biol Cell* 2006;17:990–1005.
2. Battye FL, Shortman K. Flow cytometry and cell-separation procedures. *Curr Opin Immunol* 1991;3:238–241.
3. Baumeister W, Steven AC. Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem Sci* 2000;25:624–630.
4. Bray D. *Cell Movements: From Molecules to Motility.* Garland, 2001.
5. Chen XJ, Butow RA. The organization and inheritance of the mitochondrial genome. *Nat Rev Genet* 2005;6:815–825.
6. Cuervo AM, Dice JF. Lysosomes: a meeting point of proteins, chaperones, and proteases. *J Mol Med* 1998;76:6–12.
7. de Duve C. Exploring cells with a centrifuge. *Science* 1975;189:186–194.
8. Engelman DM. Membranes are more mosaic than fluid. *Nature* 2005;438:578–580.
9. Higgins CF, Linton KJ. Structural biology. The xyz of ABC transporters. *Science* 2001;293:1782–1784.
10. Howell KE, Devaney E, Gruenberg J. Subcellular fractionation of tissue culture cells. *Trends Biochem Sci* 1989;14:44–48.
11. Hruz PW, Mueckler MM. Structural analysis of the GLUT1 facilitative glucose transport. *Mol Membr Biol* 2001;18:183–193.
12. Lamond A, Earnshaw W. Structure and function in the nucleus. *Science* 1998;280:547–553.

13. Kaushik S, Kiffin R, Cuervo AM. Chaperone-mediated autophagy and aging: a novel regulatory role of lipids revealed. *Autophagy* 2007;3:387–389.
14. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. *Science* 1997;276:131–133.
15. Mackinnon R. Structural biology. Membrane protein insertion and stability. *Science* 2005;307:1425–1426.
16. Masters C, Crane D. Recent developments in peroxisome biology. *Endeavour* 1996;20:68–73.
17. Minor DL. Potassium channels: life in the post-structural world. *Curr Opin Struct Biol* 2001;11:408–414.
18. Mason WT. *Fluorescent and Luminescent Probes: Biological Activity*, 2nd ed. Academic Press, 1999.
19. Matsumoto B,editor. *Methods in Cell Biology: Biological Applications of Confocal Microscopy*, Vol. 70 G. Academic Press, 2002.
20. Sachs JN, Engelman DM. Introduction to the membrane protein reviews: the interplay of structure, dynamics, and environment in membrane protein function. *Annu Rev Biochem* 2006;75:707–712.
21. Schekman R, Peroxisomes: another branch of the secretory pathway? *Cell* 2005;122:1–2.
22. Schliwa M, Woehlke G. Molecular motors. *Nature* 2003;422:759–765.
23. Sluder G, Wolf D,editors. *Methods in Cell Biology: Video Microscopy*, Vol. 56. Academic Press, 1998.
24. Sprong H, van der Sluijs P, van Meer G. How proteins move lipids and lipids move proteins. *Nature Rev Mol Cell Biol* 2001;2:504–513.
25. Subramani S. Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. *Physiol Rev* 1998;78:171–188.
26. Taddei A. Active genes at the nuclear pore complex. *Curr Opin Cell Biol* 2007;19:305–310.
27. Tamm LK, Kiessling VK, Wagner ML. *Membrane Dynamics: Encyclopedia of Life Sciences*. Nature Publishing Group, 2001.
28. Vance DE, Vance JE, *Biochemistry of Lipids, Lipoproteins and Membranes*, 4th ed. Elsevier, 2002.
29. Wickner W, Schekman R. Protein translocation across biological membranes. *Science* 2005;310:1452–1456.

Metabolism and Regulation

1. Fell D. *Understanding the Control of Metabolism*. Portland Press, 1997.
2. van de Werve G, Lange A, Newgard C, Mechin MC, Li Y, Berteloot A. New lessons in the regulation of glucose metabolism taught by the glucose 6-phosphatase system. *Eur J Biochem* 2000;267:1533–1549.
3. Harris DA. *Bioenergetics at a Glance*. Blackwell Scientific, 1995.
4. Hurley JH, Dean AM, Sohl JL, Koshland DJ, Stroud RM. Regulation of an enzyme by phosphorylation at the active site. *Science* 1990;249:1012–1016.
5. Lu HP, Xun L, Xie XS. Single-molecule enzymatic dynamics. *Science* 1998;282:1877–1882.

6. Martin BR. *Metabolic Regulation: A Molecular Approach*. Blackwell Scientific, 1987.
7. Miziorko HM. Phosphoribulokinase: current perspectives on the structure/function basis for regulation and catalysis. *Adv Enzymol Relat Areas Mol Biol* 2000;74:95–127.
8. Moser CC, Keske JM, Warncke K, Farid RS, Dutton PL, Nature of biological electron transfer. *Nature* 1992;355:796–802.
9. Page MI, Williams A,editors. *Enzyme Mechanisms*. Royal Society of Chemistry, 1987.
10. Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* 2000;203:229–248.

Genetic Inheritance

DNA Replication

1. Bullock PA. The initiation of simian virus 40 DNA replication *in vitro*. *Crit Rev Biochem Mol Biol* 1997;32:503–568.
2. Cvetic C, Walter JC. Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* 2005;16:343–353.
3. Kornberg A, Baker TA. *DNA Replication*, 3rd ed. University Science Books, 2005.
4. Kunkel TA. DNA replication fidelity. *J Biol Chem* 2004;279:16895–16898.
5. Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* 2007;317:127–130.
6. Waga S, Stillman B. The DNA replication fork in eukaryotic cells. *Annu Rev Biochem* 1998;67:721–751.

Transcriptional Control

1. Bell CE, Lewis M. The *lac* repressor: a second generation of structural and functional studies. *Curr Opin Struct Biol* 2001;11:19–25.
2. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. *Science* 1998;281:60–63.
3. Boeger H, Bushnell DA, Davis R, Griesenbeck J, Lorch Y, Strattan JS, Westover KD, Kornberg RD. Structural basis of eukaryotic gene transcription. *FEBS Lett* 2005;579:899–903.
4. Borukhov S, Lee J, Laptenko O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol* 2005;55:1315–1324.
5. Brasset E, Vaury C. Insulators are fundamental components of the eukaryotic genomes. *Heredity* 2005;94:571–576.
6. Bushnell DA, Westover KD, Davis RE, Kornberg RD. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* 2004;303:983–988.
7. Butler JE, Kadonaga JT. The RNA polymerase core promoter: a key component in the regulation of gene expression. *Genes Dev* 2002;16:2583–2592.
8. Cramer P. Multisubunit RNA polymerases. *Curr Opin Struct Biol* 2002;12:89–97.
9. Courey AJ, Jia S. Transcriptional repression: the long and the short of it. *Genes Dev* 2001;15:2786–2796.

10. Darst SA. Bacterial RNA polymerase. *Curr Opin Struct Biol* 2001;11:155–162.
11. Green MR. Eukaryotic transcription activation: right on target. *Mol Cell* 2005;18:399–402.
12. Harbison CT, et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* 2004;431:99–104.
13. Isogai Y, Takada S, Tjian R, Keles S. Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. *EMBO J* 2007;26:79–89.
14. Kim DK,et al. The regulation of elongation by eukaryotic RNA polymerase II: a recent view. *Mol Cell* 2001;11:267–274.
15. Levine M, Tjian R. Transcription regulation and animal diversity. *Nature* 2003;424:147–151.
16. Luscombe NM,et al. An overview of the structures of protein–DNA complexes. *Genome Biol* 2000;1:1–37.
17. Marr MT, Isogai Y, Wright KJ, Tjian R. Coactivator cross-talk specifies transcriptional output. *Genes Dev* 2006;20:1458–1469.
18. Nirenberg M. Historical review: deciphering the genetic code—a personal account. *Trends Biochem Sci* 2004;29:46–54.
19. Roeder RG. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* 2005;579:909–915.
20. Tupler R, Perini G, Green MR. Expressing the human genome. *Nature* 2001;409:832–833.
21. Woychik NA, Hampsey M. The RNA polymerase II machinery: structure illuminates function. *Cell* 2002;108:453–463.
22. Young BA, Gruber TM, Gross CA. Views of transcription initiation. *Cell* 2002;109:417–420.

Translational and Posttranslational Controls

1. Alexander RW, Schimmel P. Domain–domain communication in aminoacyl-tRNA synthetases. *Prog Nucleic Acid Res Mol Biol* 2001;69:317–349.
2. Garrett RA, et al., editors. *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions*. ASM Press, 2000.
3. Green R. Ribosomal translocation: EF–G turns the crank. *Curr Biol* 2000;10: R369–R373.
4. Holcik M, Pestova TV. Translation mechanism and regulation: old players, new concepts. Meeting on Translational Control and Non-Coding RNA. *EMBO J* 2007;8:639–643.
5. Horan LH, Noller HF. Intersubunit movement is required for ribosomal translocation. *Proc Natl Acad Sci USA* 2007;104:4881–4885.
6. Maguire BA, Zimmermann RA. The ribosome in focus. *Cell* 2001;104:813–816.
7. Nirenberg M, et al. The RNA code in protein synthesis. *Cold Spring Harbor Symp Quant Biol* 1966;31:11–24.
8. Pestova TV, et al. Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci USA* 2001;98:7029–7036.
9. Pickart CM, Cohen RE, Proteasomes and their kin: proteases in the machine age. *Nat Rev Mol Cell Biol* 2004;5:177–187.
10. Ramakrishnan V. Ribosome structure and the mechanism of translation. *Cell* 2002;108:557–572.

11. Sonenberg N, Hershey JWB, Mathews MB,editors. *Translational Control of Gene Expression*. Cold Spring Harbor Laboratory Press, 2000.
12. Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol* 2005;6:599–609.

Epigenetic Inheritance

1. Berger SL. Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 2002;8:142–148.
2. Bestor TH. The DNA methyltransferases of mammals. *Hum Mol Genet* 2000;9: 2395–2402.
3. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell* 2007;128:669–681.
4. Bird A. The essentials of DNA methylation. *Cell* 1992;70:5–8.
5. Craig JM, Bickmore WA. The distribution of CpG islands in mammalian chromosomes. *Nat Genet* 1994;7:376–382.
6. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. *Cell* 2007;128:635–638.
7. Jenuwein T, Allis CD. Translating the histone code. *Science* 2001;293:1074–1080.
8. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002;3:415–428.
9. Murrell A, Rakyan VK, Beck S. From genome to epigenome. *Hum Mol Genet* 2005;15: R3–R10.
10. Richards EJ, Elgin SC. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* 2002;108:489–500.
11. Robertson KD, Jones PA. DNA methylation: past, present and future directions. *Carcinogenesis* 2000;27:461–467.

Cell Signaling and Integrated Controls

1. Andres A, et al. Protein tyrosine phosphatases in the human genome. *Cell* 2004;117:699–711.
2. Bell J. Predicting disease using genomics. *Nature* 2004;429:453–456.
3. Cantley LC, Transcription: translocating tubby. *Science* 2001;292:2019–2021.
4. Cantley L. The phosphoinositide 3-kinase pathway. *Science* 2002;296:1655–1657.
5. Chang L, Karin M. Mammalian MAP kinase signaling cascades. *Nature* 2001;410:37–40.
6. Delmas P, Brown D. Junctional signaling microdomains: bridging the gap between neuronal cell surface and Ca_2^{+} stores. *Neuron* 2002;36:787–790.
7. Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signaling pathways in cancer. *Oncogene* 2007;26:3279–3290.
8. Kirschner MW. The meaning of systems biology. *Cell* 2005;121:503–504.
9. Leonard W. Role of Jak kinases and STATs in cytokine signal transduction. *Int J Hematol* 2001;73:271–277.

10. Lum L, Beachy PA. The Hedgehog response network: sensors, switches, and routers. *Science* 2004;304:1755–1759.
11. Luttrell LM, Daaka Y, Lefkowitz RJ. Regulation of tyrosine kinase cascades by G-protein-coupled receptors. *Curr Opin Cell Biol* 1999;11:177–183.
12. Ma'ayan et al. Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science* 2005;309:1078–1083.
13. Michel JJ, Scott JD. AKAP mediated signal transduction. *Annu Rev Pharmacol Toxicol* 2002;42:235–257.
14. Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signaling networks and analysis of their properties. *Nat Rev Mol Cell Biol* 2005;6:99–111.
15. Pierce KL, Premont RT, Lefkowitz RJ. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol* 2002;3:639–652.
16. Rual JF, et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 2005;437:1173–1178.
17. Simon M. Receptor tyrosine kinases: specific outcomes from general signals. *Cell* 2000;103:13–15.
18. Vo N, Goodman R. CREB-binding protein and p300 in transcriptional regulation. *J Biol Chem* 2001;276:13505–21351.
19. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004;10:789–799.

Genentic Engineering and Recombinant DNA

1. Achenbach TV, Brunner B, Heermeier K. Oligonucleotide-based knockdown technologies, antisense versus RNA interference. *ChemBioChem* 2003;4:928–935.
2. Arnheim N, Erlich H. Polymerase chain reaction strategy. *Annu Rev Biochem* 1992;61:131–156.
3. Capecchi MR. Generating mice with targeted mutations. *Nat Med* 2001;7:1086–1090.
4. Cohen S, Chang A, Boyer H, Helling R. Construction of biologically functional bacterial plasmids *in vitro*. *Proc Natl Acad Sci USA* 1973;70:3240–3244.
5. Felgner PL, Gadek TR, Holm M, Roman R, Chan HW, Wenz M, et al. Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc Natl Acad Sci USA* 1987;84:7413–7417.
6. Freshney RI. *Culture of Animal Cells: A Manual of Basic Technique*, 4th ed. New York: Wiley, 2000.
7. Golgstein DA, Thomas JA. Biopharmaceuticals derived from genetically modified plants. *Q J Med* 2004;97:705–716.
8. Hansen G, Wright MS. Recent advances in the transformation of plants. *Trends Plant Sci* 1999;4:226–231.
9. Hunkapiller T, Kaiser RJ, Koop BK, Hood L. Large-scale and automated DNA sequence determination. *Science* 254:59–67.
10. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
11. Maniatis T, et al. The isolation of structural genes from libraries of eucaryotic DNA. *Cell* 1978;15:687–701.

12. Sanger F, Nickle SN, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463–5467.
13. Scherer LJ, Rossi JJ. Approaches for the sequence-specific knockdown of mRNA. *Nat Biotech* 2003;21:1457–1465.
14. Venter JC, Adams MA, Myers EW, et al. The sequence of the human genome. *Science* 2000;291:1304–1351.
15. Weigel D, Glazebrook J. *Arabidopsis: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2001.
16. Yang NS, Sun WH. Gene gun and other non-viral approaches for cancer gene therapy. *Nat Med* 1995;1:481–483.

Metabolic Engineering

1. Askenazi M, et al. Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nat Biotechnol* 2003;21:150–156.
2. Bailey JE. Toward a science of metabolic engineering. *Science* 1991;252:1668–1675.
3. Bailey JE. Lessons from metabolic engineering for functional genomics and drug discovery. *Nat Biotechnol* 1999;17:616–618.
4. Bailey JE. Complex biology with no parameters. *Nat Biotechnol* 2001;19:503–504.
5. Bailey JE. Reflections on the scope and the future of metabolic engineering and its connections to functional genomics and drug discovery. *Metab Eng* 2001;3:111–114.
6. Cullen LM, Arndt GM. Genome-wide screening for gene function using RNAi in mammalian cells. *Immunol Cell Biol* 2005;83:217–223.
7. Edwards JS, Palsson B. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 2000;97:5528–5533.
8. Fussenegger M, Betenbaugh MJ. Metabolic engineering II. Eukaryotic systems. *Biotechnol Bioeng* 2002;79:509–531.
9. Ideker T, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 2001;292:929–934.
10. Kacser H, Burns JA. The control of flux. *Symp Soc Exp Biol* 1973;27:65–104.
11. Klapa MI, Aon JC, Stephanopoulos G. Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. *Eur J Biochem* 2003;270:3525–3542.
12. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 2000;405:837–846.
13. Park SM, Shaw-Reid CA, Sinskey AJ, Stephanopoulos G. Elucidation of anaplerotic pathways in *Corynebacterium glutamicum* via ^{13}C-NMR spectroscopy and GC–MS. *Appl Microbiol Biotechnol* 1997;47:430–440.
14. Raab RM, Tyo K, Stephanopoulos G. Metabolic engineering. *Adv Biochem Eng Biotechnol* 2005;100:1–17.
15. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* 2002;99:15112–15117.
16. Stephanopoulos G, Vallino JJ. Network rigidity and metabolic engineering in metabolite overproduction. *Science* 1991;252:1675–1681.
17. Stephanopoulos G, Aristidou A, Nielsen J. *Metabolic Engineering: Principles and Methodologies*. San Diego, CA: Academic Press, 1998.

18. Stephanopoulos G, Alper H, Moxley J. Exploiting biological complexity for strain improvement through systems biology. *Nat Biotechnol* 2004;22:1261–1267.
19. Wiechert W, Mollney M, Isermann N, Wurzel M, De Graaf AA. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol Bioeng* 1999;66:69–85.