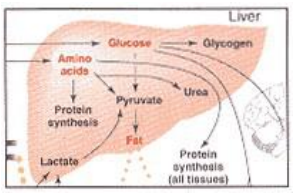
Chapter 13—
Metabolic Interrelationships

Robert A. Harris and David W. Crabb



13.1— Overview

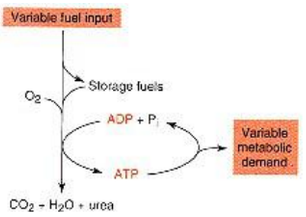


Figure 13.1
Humans are able to use a variable fuel input to meet a variable metabolic demand.

In this chapter the interdependence of metabolic processes of the major tissues of the body will be stressed. Not all of the major metabolic pathways operate in every tissue at any given time. Given the nutritional and hormonal status of a patient, we need to be able to say qualitatively which major metabolic pathways of the body are functional and how these pathways relate to one another.

The metabolic processes with which we are concerned are glycogenesis, glycogenolysis, gluconeogenesis, glycolysis, fatty acid synthesis, lipogenesis, lipolysis, fatty acid oxidation, glutaminolysis, tricarboxylic acid cycle activity, ketogenesis, amino acid oxidation, protein synthesis, proteolysis, and urea synthesis. It is important to know (1) which tissues are most active in these various processes, (2) when these processes are most or least active, and (3) how these processes are controlled and coordinated in different metabolic states.

The best way to gain an understanding of the relationships of the pathways to one another is to become familiar with the changes in metabolism that occur during the **starve–feed cycle.** As shown in Figure 13.1, the starve–feed cycle allows a variable fuel and nitrogen consumption to meet a variable metabolic and anabolic demand. Feed refers to the intake of meals (the variable fuel input) after which we store the fuel (in the form of glycogen and fat) to be used to meet our metabolic demand while we fast. Note the participation of an **ATP cycle** within the starve–feed cycle (Figure 13.1). Adenosine triphosphate

CLINICAL CORRELATION 13.1

Obesity

Obesity is the most common nutritional problem in the United States. It can reduce life span because it is a risk factor in development of diabetes mellitus, hypertension, endometrial carcinoma, osteoarthritis, gallstones, and cardiovascular diseases. Obesity is easy to explain—an obese person has eaten more than he/she required. The accumulation of massive amounts of body fat is not otherwise possible. For unknown reasons, the neural control of caloric intake to balance energy expenditure is abnormal. Rarely, obesity is secondary to a correctable disorder, such as hypothyroidism or Cushing's syndrome. The latter is the result of increased secretion of glucocorticoids, which cause fat deposition in the face and trunk, with wasting of the limbs, and glucose intolerance. These effects are due to increased protein breakdown in muscle and conversion of the amino acids to glucose and fat. Less commonly, tumors, vascular accidents, or maldevelopment of the nervous system hunger control centers in the hypothalamus cause obesity.

Genetic models of obesity in rodents have led to breakthroughs in our understanding of the control of body mass. The obese (*ob/ob*) mouse was discovered in the 1950s, and the defective gene cloned in 1994. This *ob* gene encodes a 146 amino acid secreted protein (alternatively called OB protein or leptin, for its slimming effect) that is produced in adipocytes and detectable in blood. The *ob/ob* mice have a nonsense mutation in the gene and produce no OB protein. Injection of the OB protein into *ob/ob* mice causes increased energy expenditure and reduced eating, with marked weight loss. This effect on appetite is mimicked by intracerebroventricular injection. The OB protein also reduced appetite and weight of normal mice. Obese humans do not generally have defective *on* genes but, like normal mice, may still respond to OB protein used as an anti-obesity drug.

In the most common type of obesity, the number of adipocytes of the body does not increase, they just get large as they become engorged with triacylglycerols. If obesity develops before puberty, however, an increase in the number of adipocytes can also occur. In the latter case, both hyperplasia (increase in cell number) and hypertrophy (increase in cell size) are contributing factors to the magnitude of the obesity. Obesity in men tends to be centered on the abdomen and mesenteric fat, while in women it is more likely to be on the hips. The male pattern, characterized by a high waist to hip circumference ratio, is more predictive of premature coronary heart disease.

The only effective treatment of obesity is reduction in the ingestion or increase in the use of calories. Practically speaking, this means dieting, since even vigorous exercise such as running only consumes 10 kcal/min of exercise. Thus an hour-long run (perhaps 5–6 miles) uses the energy present in about two candy bars. However, exercise programs can be useful to help motivate individuals to remain on their diets. Unfortunately, the body compensates for decreased energy intake with reduced formation of triiodothyronine and a corresponding decrease in the basal metabolic rate. Thus there is a biochemical basis for the universal complaint that it is far easier to gain than to lose weight. Furthermore, about 95% of people who are able to lose a significant amount of weight regain it within one year.

Bray, G. D. Effect of caloric restriction on energy expenditure in obese patients. *Lancet* 2:397, 1969; Bray, G. D. The overweight patient. *Adv. Intern. Med.* 21:267, 1976; and Baringer, M. Obese protein slims mice. *Science* 269:475, 1995.

CLINICAL CORRELATION 13.2

Protein Malnutrition

Protein malnutrition is the most important and widespread nutritional problem among young children in the world today. The clinical syndrome, called kwashiorkor, occurs mainly in children 1–3 years of age and is precipitated by weaning an infant from breast milk onto a starchy, protein-poor diet. The name originated in Ghana, meaning "the sickness of the older child when the next baby is born." It is a consequence of feeding the child a diet adequate in calories but deficient in protein. It may become clinically manifest when protein requirements are increased by infection, for example, malaria, helminth infestation, or gastroenteritis. The syndrome is characterized by poor growth, low plasma protein and amino acid levels, muscle wasting, edema, diarrhea, and increased susceptibility to infection. The presence of subcutaneous fat clearly differentiates it from simple starvation. The maintenance of fat stores is due to the high carbohydrate intake and resulting high insulin levels. In fact, the high insulin level interferes with the adaptations described for starvation. Fat is not mobilized as an energy source, ketogenesis does not take place, and there is no transfer of amino acids from the skeletal muscle to the visceral organs, that is, the liver, kidneys, heart, and immune cells. The lack of dietary amino acids results in diminished protein synthesis in all tissues. The liver becomes enlarged and infiltrated with fat, reflecting the need for hepatic protein synthesis for the formation and release of lipoproteins. In addition, protein malnutrition impairs the function of the gut, resulting in malabsorption of calories, protein, and vitamins, which accelerates the disease. The consequences of the disease depend somewhat on when in development the deficiency occurs. Children with low weight for height are called "wasted" but can make a good recovery when properly fed. Those with low height for weight are called "stunted" and never regain full height or cognitive potential.

Protein–calorie malnutrition is also a problem for the elderly when they become sick. Both the energy requirements and food intake of well elderly decline with age. On a lower calorie diet, there is the risk that insufficient intake of protein and of certain nutrients such as iron, calcium, and vitamins will be lower than needed. Deficiencies in these nutrients may accelerate loss of lean body mass and strength (leading to falls), anemia, loss of bone strength, and rarely, vitamin deficiency states.

Bistrian, B. R., Blackburn, G. L., Vitale, J., and Cochran, D. Prevalence of malnutrition in general medical patients. *JAMA* 235:1567, 1976; Chase, H. P., Kumar, V., Caldwell, R. T., and O'Brien, D. Kwashiorkor in the United States. *Pediatrics* 66:972, 1980; and Schlienger, J. L., Pradignac, A., and Grunenberger, F. Nutrition of the elderly a challenge between facts and needs. *Horm. Res.* 43:46, 1995.

is the energy-transferring agent in the starve–feed cycle, being like money to the cell.

Humans have the capacity to consume food at a rate far greater than their basal caloric requirements, which allows them to survive from meal to meal. We thus store calories as glycogen and fat and utilize them as needed. Unfortunately, an almost unlimited capacity to consume food is matched by an almost unlimited capacity to store it as fat. **Obesity** is the consequence of excess food consumption and is the commonest form of malnutrition in affluent countries (see Clin. Corr. 13.1), whereas other forms of malnutrition are more prevalent in developing countries (see Clin. Corr. 13.2 and 13.3). The regulation of food

CLINICAL CORRELATION 13.3

Starvation

Starvation leads to the development of a syndrome known as marasmus. Marasmus is a word of Greek origin meaning "to waste." Although not restricted to any age group, it is most common in children under 1 year of age. In developing countries early weaning of infants from breast milk is a common cause of marasmus. This may result from pregnancies in rapid succession, the desire of the mother to return to work, or the use of overdiluted artificial formulas (to make the expensive formulas last longer). This practice leads to insufficient intake of calories. Likewise, diarrhea and malabsorption can develop if safe water and sterile procedures are not used.

In contrast to kwashiorkor (see Clin. Corr. 13.2), subcutaneous fat, hepatomegaly, and fatty liver are absent in marasmus because fat is mobilized as an energy source and muscle temporarily provides amino acids to the liver for the synthesis of glucose and hepatic proteins. Low insulin levels allow the liver to oxidize fatty acids and to produce ketone bodies for other tissues. Ultimately, energy and protein reserves are exhausted, and the child starves to death. The immediate cause of death is often pneumonia, which occurs because the child is too weak to cough. Adults can suffer from marasmus as a result of diseases that prevent swallowing (cancer of the throat or esophagus) or interfere with access to food (strokes or dementia).

Waterlow, J. C. Childhood malnutrition—the global problem. *Proc. Nutr. Soc.* 38:1, 1979; and Uvin, P. The state of world hunger. *Nutr. Rev.* 52:151, 1994.

consumption is complex and not well understood. Recent observations suggest that the product of the **leptin gene** *(ob* in mice) expressed in adipocytes is secreted into the blood and regulates energy expenditure and appetite through the hypothalamus (see Clin. Corr. 13.1). The tight control needed is indicated by the calculation that eating two extra pats of butter (~100 cal) per day over caloric expenditures results in a 10-lb weight gain per year. A weight gain of 10 lb may not sound excessive, but multiplied by 10 years it equals obesity!

13.2— Starve–Feed Cycle

In the Well-Fed State the Diet Supplies the Energy Requirements

Figure 13.2 shows the fate of glucose, amino acids, and fat obtained from food. Glucose passes from the intestinal epithelial cells to the liver by way of the portal vein. Amino acids are partially metabolized in the gut before being released into portal blood. Fat, contained in **chylomicrons,** is secreted by the intestinal epithelial cells into lymphatics, which drain the intestine. The lymphatics lead to the thoracic duct, which, by way of the subclavian vein, delivers chylomicrons to the blood at a site of rapid blood flow. This rapidly distributes the chylomicrons and prevents their coalescence.

Liver is the first tissue to have the opportunity to use dietary glucose. Glucose can be converted into glycogen by glycogenesis, into pyruvate and lactate by glycolysis, or can be used in the pentose phosphate pathway for the

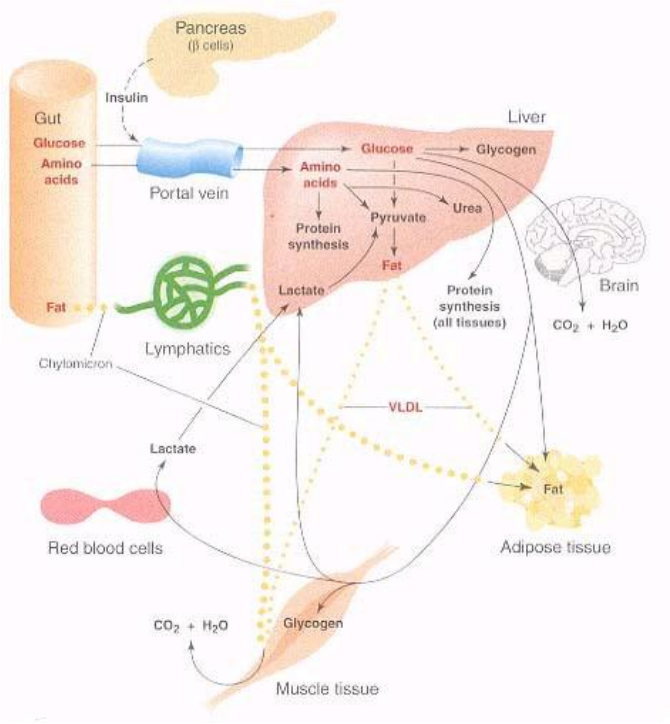


Figure 13.2
Disposition of glucose, amino acids, and fat by various tissues in the well-fed state.

generation of NADPH for synthetic processes. Pyruvate can be oxidized to acetyl CoA, which, in turn, can be converted into fat or oxidized to CO_2 and water by the TCA cycle. Much of the glucose coming from the intestine passes through the liver to reach other organs, including brain and testis, which are almost solely dependent on glucose for the production of ATP, red blood cells and renal medulla, which can only convert glucose to lactate and pyruvate, and the adipose tissue, which converts it into fat. Muscle also has good capacity to use glucose, converting it to glycogen or using it in the glycolytic and the TCA cycle pathways. A number of tissues produce lactate and pyruvate from circulating glucose, which are taken up by the liver and converted to fat. In the very well-fed state, the liver uses glucose and does not engage in gluconeo-genesis. Thus the **Cori cycle** (the conversion of glucose to lactate in the peripheral tissues followed by conversion of lactate back to glucose in liver) is interrupted in the well-fed state.

Dietary protein is hydrolyzed in the intestine, the cells of which use some amino acids as an energy source. Most dietary amino acids are transported into the portal blood, but the intestinal cells metabolize aspartate, asparagine, glutamate, and glutamine and release alanine, lactate, citrulline, and proline into portal blood. Liver then has the opportunity to remove absorbed amino acids from the blood (Figure 13.2). The liver lets most of each amino acid pass through, unless the concentration of the amino acid is unusually high. This is especially important for the essential amino acids, needed by all tissues of the body for protein synthesis. Liver catabolizes amino acids, but the K_m values for amino acids of many of the enzymes involved are high, allowing the amino acids to be present in excess before significant catabolism can occur. In contrast, the tRNA-charging enzymes that generate **aminoacyl-tRNAs** have much lower K_m values for amino acids. This ensures that as long as all the amino acids are present, protein synthesis occurs as needed for growth and protein turnover. Excess amino acids can be oxidized completely to CO_2, urea, and water, or the intermediates generated can be used as substrates for lipogenesis. Amino acids that escape the liver are used for protein synthesis or energy in other tissues.

Glucose, lactate, pyruvate, and amino acids can support hepatic **lipogenesis** (Figure 13.2). Fat formed from these substrates is released from the liver in the form of very low density lipoproteins (VLDLs). Dietary fat is delivered to the bloodstream as **chylomicrons**. Both chylomicrons and **VLDLs** circulate in the blood until they are acted on by an extracellular enzyme attached to the endothelial cells in the lumen of the capillaries. This enzyme, **lipoprotein lipase,** is particularly abundant in the capillaries in adipose tissue. It acts on both the VLDLs and chylomicrons, liberating fatty acids by hydrolysis of the triacylglycerols. The fatty acids are then taken up by the adipocytes, reesterified with glycerol 3-phosphate to form triacylglycerols, and stored as fat droplets. Glycerol 3-phosphate is generated from glucose, using the first half of the glycolytic pathway to generate dihydroxyacetone phosphate, which is reduced to glycerol 3-phosphate by glycerol-3-phosphate dehydrogenase.

The β cells of the pancreas are very responsive to the influx of glucose and amino acids in the fed state. The β cells release insulin during and after eating, which is essential for the metabolism of these nutrients by liver, muscle, and adipose tissue. The role of insulin in the starve–feed cycle is discussed in more detail in Section 13.3.

In the Early Fasting State Hepatic Glycogenolysis Is an Important Source of Blood Glucose

Hepatic glycogenolysis is very important for maintenance of blood glucose during early fasting (Figure 13.3). Lipogenesis is curtailed, and lactate, pyruvate, and amino acids used by that pathway are diverted into formation of glucose, completing the Cori cycle. The **alanine cycle,** in which carbon and nitrogen

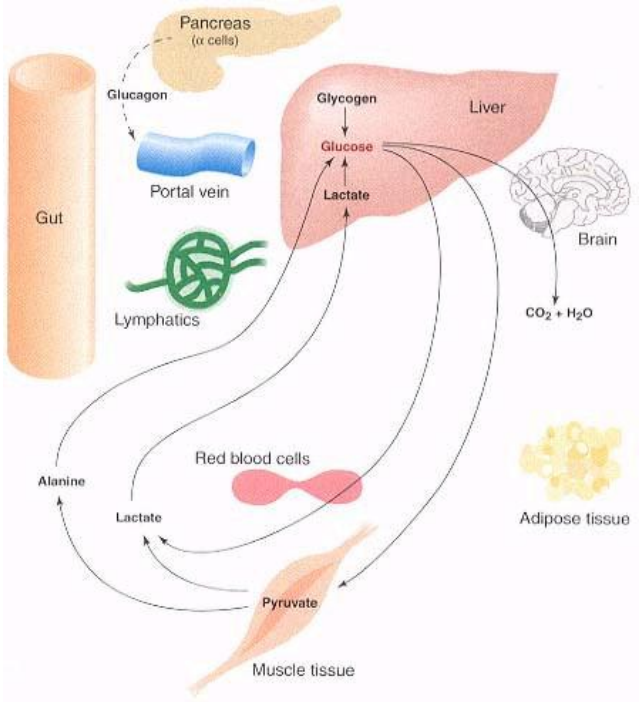


Figure 13.3
Metabolic interrelationships of major tissues in the early fasting state.

return to the liver in the form of alanine, also becomes important. Catabolism of amino acids for energy is greatly diminished in early fasting because less is available.

The Fasting State Requires Gluconeogenesis from Amino Acids and Glycerol

No fuel enters from the gut and little glycogen is left in the liver in the fasting state (Figure 13.4). Tissues that require glucose are dependent on hepatic **gluconeogenesis,** primarily from lactate, glycerol, and alanine. The Cori and alanine cycles play important roles but do not provide carbon for net synthesis of glucose. Glucose formed from lactate and alanine by the liver merely replaces that which was converted to lactate and alanine by peripheral tissues. In effect, these cycles transfer energy from fatty acid oxidation in the liver to peripheral tissues that cannot oxidize fat. The brain oxidizes glucose completely to CO_2 and water. Hence net glucose synthesis from some other source of carbon is mandatory in fasting. Fatty acids cannot be used for the synthesis of glucose, because acetyl CoA obtained by fatty acid catabolism cannot be converted to glucose. Glycerol, a by-product of lipolysis in adipose tissue, is an important substrate for glucose synthesis. However, protein, especially from skeletal muscle, supplies most of the carbon needed for net glucose synthesis. Proteins are hydrolyzed within muscle cells and most amino acids are partially metabolized within muscle cells. Only two amino acids—alanine and glutamine—are re-

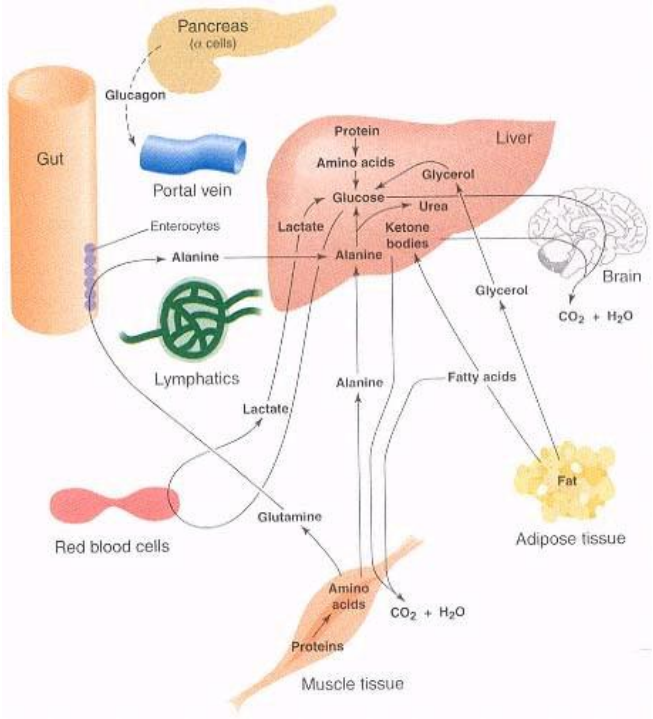


Figure 13.4
Metabolic interrelationships of major tissues in the fasting state.

leased in large amounts. The others are metabolized to give intermediates (pyruvate and α-ketoglutarate), which can yield alanine and glutamine. Branched-chain amino acids are a major source of nitrogen for the production of alanine and glutamine in muscle. Branched-chain α-keto acids produced from the branched-chain amino acids by transamination are partially released into the blood for uptake by the liver, which synthesizes glucose from the keto acid of valine, ketone bodies from the keto acid of leucine, and both glucose and ketone bodies from the keto acid of isoleucine.

Much of the glutamine released from muscle is converted into alanine by the intestinal epithelium. Glutamine is partially oxidized in **enterocytes** to supply energy and precursor molecules for synthesis of pyrimidines and purines, with the carbon and amino groups left over being released back into the bloodstream in part as alanine and NH_4^+. This pathway, sometimes called **glutaminolysis** because glutamine is only partially oxidized, involves formation of malate from glutamine via the TCA cycle and the conversion of malate to pyruvate by malic enzyme (Figure 13.5*a*). Pyruvate then transaminates with glutamate to give alanine, which is released from the cells.

Glutaminolysis is also used by cells of the immune system (lymphocytes and macrophages) to meet a large portion of their energy needs (Figure 13.5*b*). Aspartate rather than alanine is the major end product of glutaminolysis in **lymphocytes.** Enterocytes and lymphocytes use glutamine as their major fuel source as a way to ensure a continuous supply of the precursor molecules

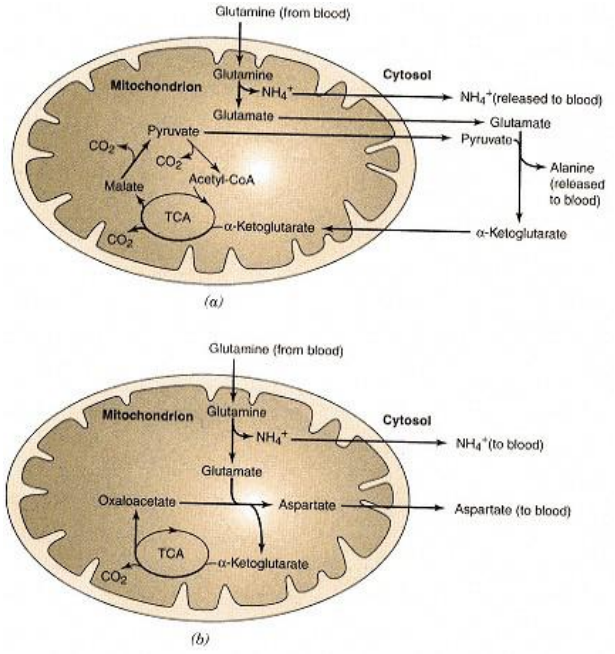


Figure 13.5
Glutamine catabolism by rapidly dividing cells.
(a) Enterocytes.
(b) Lymphocytes.
Redrawn from Duée, P.-H., Darcy-Vrillon, B., Blachier, F., and Morel, M.-T. Fuel selection in intestinal cells. *Proc. Nutr. Soc.* 54:83, 1995.

(glutamine and aspartate) required for synthesis of purines and pyrimidines, which these rapidly dividing cells need for the synthesis of RNA and DNA.

Synthesis of glucose in the liver during fasting is closely linked to synthesis of urea. Most amino acids can give up the amino nitrogen by transamination with α-ketoglutarate, forming glutamate and a new α-keto acid, which can be utilized for glucose synthesis. Glutamate provides both nitrogenous compounds required for urea synthesis: ammonia from oxidative deamination by glutamate dehydrogenase, and aspartate from transamination of oxaloacetate by aspartate aminotransferase. An additional important source of ammonia and precursors of ornithine such as citrulline is the gut mucosa (described in more detail in Section 13.4).

Adipose tissue is also very important in the fasting state. Because of low blood insulin levels during fasting, **lipolysis** is greatly activated in this tissue. This raises the blood level of fatty acids, which are used in preference to glucose by many tissues. In heart and muscle, the oxidation of fatty acids inhibits glycolysis and pyruvate oxidation. In liver, fatty acid oxidation provides most of the ATP needed for gluconeogenesis. Very little acetyl CoA generated by fatty acid oxidation in liver is oxidized completely. The acetyl CoA is converted instead into **ketone bodies** by liver mitochondria. Ketone bodies (acetoacetate and β-hydroxybutyrate) are released into the blood and are a source of energy for many tissues. Like fatty acids, ketone bodies are preferred by many tissues over glucose. Fatty acids are not oxidized by the brain because fatty acids cannot cross the blood–brain barrier. Ketone bodies can penetrate, however, and are oxidized. Once their blood concentration is high enough, ketone bodies function as an alternative fuel for the brain. They are unable, however, to completely replace the need for glucose by the brain. Ketone bodies may also suppress proteolysis and branched-chain amino acid oxidation in muscle and

decrease alanine release. This both decreases muscle wasting and reduces the amount of glucose synthesized in liver. As long as ketone body levels are maintained at a high level by hepatic fatty acid oxidation, there is less need for glucose, less need for gluconeogenic amino acids, and less need for breaking down precious muscle tissue.

CLINICAL CORRELATION 13.4

Reye's Syndrome

Reye's syndrome is a devastating but now rare illness of children that is characterized by evidence of brain dysfunction and edema (irritability, lethargy, and coma) and liver dysfunction (elevated plasma free fatty acids, fatty liver, hypoglycemia, hyperammonemia, and accumulation of short-chain organic acids). It appears that hepatic mitochondria are specifically damaged, which impairs fatty acid oxidation and synthesis of carbamoyl phosphate and ornithine (for ammonia detoxification) and oxaloacetate (for gluconeogenesis). On the other hand, the accumulation of organic acids has suggested that the oxidation of these compounds is defective and that the CoA esters of some of these acids may inhibit specific enzymes, such as carbamoyl phosphate synthetase I, pyruvate dehydrogenase, pyruvate carboxylase, and the adenine nucleotide transporter, all present in mitochondria. The issue has not yet been resolved. The use of aspirin by children with varicella was linked to the development of Reye's syndrome, and parents have been urged not to give aspirin to children with viral infections. The incidence of the syndrome subsequently decreased. The therapy for Reye's syndrome consists of measures to reduce brain edema and the provision of glucose intravenously. Glucose administration prevents hypoglycemia and elicits a rise in insulin levels that may (1) inhibit lipolysis in adipose cells and (2) reduce proteolysis in muscles and the release of amino acids, which (3) reduces the deamination of amino acids to ammonia.

Reye, R. D. K., Morgan, G., and Baval, J. Encephalopathy and fatty degeneration of the viscera, a disease entity in childhood. *Lancet* 2:749, 1963; and Treem, W. R. Inherited and acquired syndromes of hyperammonemia and encephalopathy in children. *Semin. Liver Dis.* 14:236, 1994.

The interrelationships among liver, muscle, and adipose tissue in supplying glucose for the brain are shown in Figure 13.4. Liver synthesizes the glucose, muscle and gut supply the substrate (alanine), and adipose tissue supplies the ATP (via fatty acid oxidation in the liver) needed for hepatic gluconeogenesis. These relationships are disrupted in Reye's syndrome (see Clin. Corr. 13.4) and by alcohol (see Clin. Corr. 7.10). This tissue cooperation is dependent on the appropriate blood hormone levels. Glucose levels are lower in fasting, reducing the secretion of insulin but favoring release of **glucagon** from the pancreas and **epinephrine** from the adrenal medulla. In addition, fasting reduces formation of **triiodothyronine,** the active form of thyroid hormone, from **thyroxine.** This reduces the daily basal energy requirements by as much as 25%. This response is useful for survival but makes weight loss more difficult than weight gain (see Clin. Corr. 13.1).

In the Early Refed State, Fat Is Metabolized Normally, but Normal Glucose Metabolism Is Slowly Reestablished

Figure 13.6 shows what happens soon after fuel is absorbed from the gut. Fat is metabolized as described above for the well-fed state. In contrast, glucose

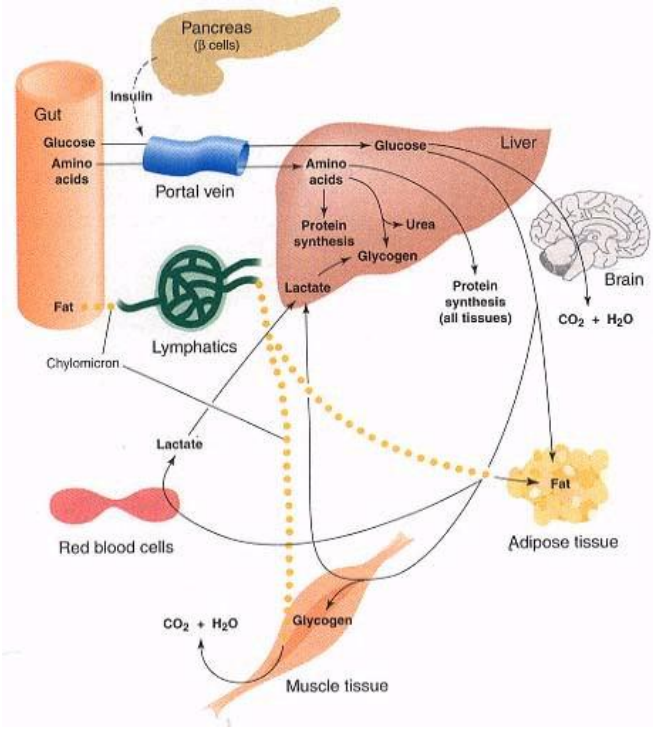
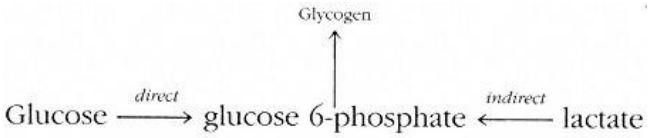


Figure 13.6
Metabolic interrelationships of major tissues in the early refed state.

is poorly extracted by the liver during this period of the starve–feed cycle. In fact, the liver remains in the gluconeogenic mode for a few hours after feeding. Rather than providing blood glucose, however, hepatic gluconeogenesis provides glucose 6-phosphate for glycogenesis. This means that liver glycogen is not repleted after a fast by direct synthesis from blood glucose. Rather, glucose is catabolized in peripheral tissues to lactate, which is converted in the liver to glycogen by the indirect pathway of glycogen synthesis (i.e., gluconeogenesis):



Gluconeogenesis from specific amino acids entering from the gut also plays an important role in reestablishing normal liver glycogen levels by the indirect pathway. After the rate of gluconeogenesis declines, glycolysis becomes the predominant means of glucose disposal in the liver, and liver glycogen is sustained by the direct pathway of synthesis from blood glucose.

Other Important Interorgan Metabolic Interactions

An important pathway exists in the intestinal epithelium for the conversion of glutamine to **citrulline** (Figure 13.7). One of the enzymes (ATP-dependent glutamate reductase) necessary for this conversion is expressed only in enterocytes. Citrulline produced in the gut is metabolized by the kidney to arginine, which can be converted to creatine or released into the blood. The liver uses blood arginine to generate ornithine, which expands the capacity of the urea cycle during periods of increased protein intake. Although perhaps not immedi-

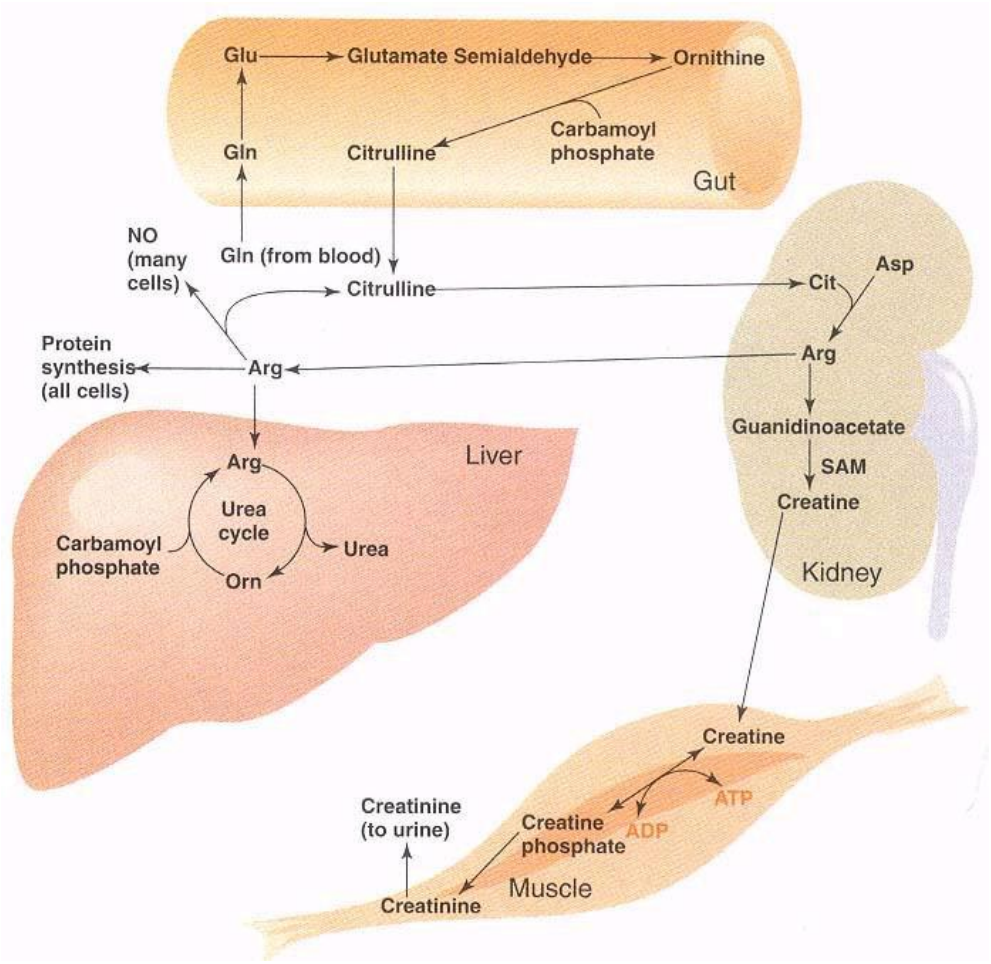


Figure 13.7
Gut and kidney function together in synthesis of arginine from glutamine.
Abbreviations: Cit, citrulline; Arg, arginine; Asp, aspartate; Gln, glutamine; Glu, glutamate; NO, nitric oxide; Orn, ornithine; SAM, *S*-adenosylmethionine.

ately obvious, this pathway is of great importance for urea cycle activity in the liver. The liver contains an enzyme system that irreversibly converts ornithine into glutamate:

Ornithine → glutamate semialdehyde → glutamate

Depletion of ornithine by these reactions inhibits urea synthesis in the liver for want of ornithine, the intermediate of the urea cycle that must recycle. Replenishment of ornithine is necessary and completely dependent on a source of blood arginine. Thus urea synthesis in the liver is dependent on citrulline synthesis by the gut and arginine synthesis by the kidney. Arginine is also used by many cells for the production of **nitric oxide** (NO) (Figure 13.7), an activator of guanylate cyclase that produces cGMP, an important second messenger (see p. 995).

Citrulline participates in another interesting interorgan shuttle. The arginine generated from citrulline in the kidney can be metabolized further to **creatine** (Figure 13.7). The first enzyme in this pathway is glycine transamidinase (GTA), which generates guanidinoacetate from arginine and glycine (see p. 483). GTA is found predominantly in renal cortex, pancreas, and liver. After methylation in a reaction that requires ***S*-adenosylmethionine** (SAM), creatine is formed. This is quantitatively the most important use of **SAM** in the body. One to two grams of creatine are synthesized per day. Creatine then circulates to other tissues, especially muscle, where it serves as a high-energy reservoir when phosphorylated to creatine phosphate. Creatine phosphate undergoes nonenzymatic conversion to **creatinine.** Creatinine is released to the bloodstream and removed from the body by renal filtration. Excretion of creatinine is thus used both as a measure of muscle mass and of renal function.

Two other compounds related to amino acids participate in interorgan shuttles. **Glutathione** (GSH) is a tripeptide that is important in detoxification of endogenously generated peroxides and exogenous chemical compounds (see p. 484). Liver plays a major role in the synthesis of GSH from glutamate, cysteine, and glycine (Figure 13.8). Synthesis is limited by the availability of cysteine. Cystine present in plasma is not taken up well by liver, which utilizes dietary methionine to form cysteine via the cystathionine pathway (see p. 469). Hepatic GSH is released both to the bloodstream and to the bile. Kidney removes a substantial amount of plasma GSH. Enterocytes may be able to take up biliary-excreted GSH from the intestinal lumen. Release to plasma is the same in fed

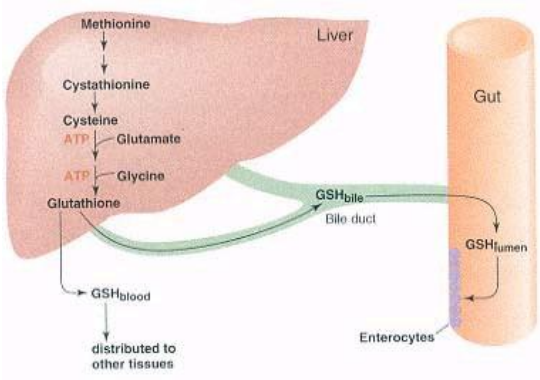


Figure 13.8
Liver provides glutathione for other tissues.

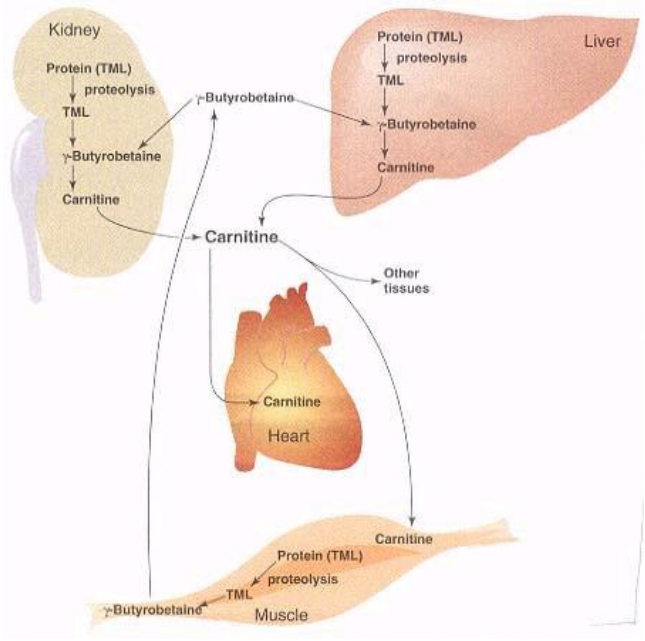


Figure 13.9
Kidney and liver provide carnitine for other tissues.
Abbreviations: Protein (TML), trimethyllysyl residues in protein molecules; TML, free trimethyllysine.

and fasting states, providing a stable source of this compound and its constituent amino acids, especially cysteine, for most tissues of the body.

Carnitine is derived from lysyl residues on various proteins, which are *N*-methylated utilizing SAM to form trimethyllysyl residues (Figure 13.9). Free **trimethyllysine** is released when the proteins are degraded. It is hydroxylated and then cleaved, releasing glycine and γ-butyrobetaine aldehyde. The latter is oxidized to γ-butyrobetaine and then hydroxylated to form carnitine. Both hydroxylation steps require **vitamin C** as a cofactor. Kidney and to a lesser extent liver are the only tissues that can carry out the complete pathway, and thus they supply other tissues, especially muscle and heart, with the carnitine needed for fatty acid oxidation. Skeletal muscle can form γ-butyrobetaine but must release it for its final conversion to carnitine by liver or kidney.

Energy Requirements, Reserves, and Caloric Homeostasis

The average person leading a sedentary life consumes 200–300 g of carbohydrate, 70–100 g of protein, and 60–90 g of fat daily. This meets a daily energy requirement of 1600–2400 kcal. As shown in Table 13.1, the **energy reserves** of an average-sized person are considerable. These reserves are called upon between meals and overnight to maintain blood glucose. Although the ability to mobilize glycogen rapidly is indeed very important, our glycogen reserves are minuscule with respect to our fat reserves (Table 13.1). Fat stores are only called upon during more prolonged fasting. The fat stores of obese subjects can weigh as much as 80 kg, adding another 585,000 kcal to their energy reserves. Protein is listed in Table 13.1 as an energy reserve because it can be used to provide amino acids for oxidation. On the other hand, protein is not inert like stored fat and glycogen. Proteins make up the muscles that allow us

TABLE 13.1 The Energy Reserves of Humans[a]

		Fuel Reserves	
Stored Fuel	***Tissue***	***(g)***	***(kcal)***
Glycogen	Liver	70	280
Glycogen	Muscle	120	480
Glucose	Body fluids	20	80
Fat	Adipose	15,000	135,000
Protein	Muscle	6,000	24,000

[a] Data are for a normal subject weighing 70 kg. Carbohydrate supplies 4 kcal g^{-1}; fat, 9 kcal g^{-1}; protein, 4 kcal g^{-1}.

to move and breathe and the enzymes that carry out metabolism. Hence it is not as dispensable as fat and glycogen and is given up by the body more reluctantly.

The constant availability of fuels in the blood is termed **caloric homeostasis,** which, as illustrated in Table 13.2, means that regardless of whether a person is well-fed, fasting, or starving to death, the blood level of fuels that supply a comparable amount of ATP when metabolized does not fall below certain limits. Note that blood glucose concentrations are controlled within very tight limits, whereas fatty acid and ketone body concentrations in the blood can vary by one or two orders of magnitude, respectively. Glucose is carefully regulated because of the absolute need of the brain for this substrate. If the blood glucose level falls too low (<2.0 mM), coma and death will follow shortly unless the glucose concentration is restored. On the other hand, **hyperglycemia** must be avoided because glucose will be lost in the urine, resulting in dehydration and sometimes hyperglycemic, hyperosmolar coma (see Clin. Corr. 13.5). Chronic hyperglycemia results in glycation of a number of proteins, which contributes to the complications of **diabetes** (see Clin. Corr. 13.6). The changes

CLINICAL CORRELATION 13.5

Hyperglycemic, Hyperosmolar Coma

Type II diabetic patients sometimes develop a condition called hyperglycemic, hyperosmolar coma. This is particularly common in the elderly and can even occur in individuals under severe metabolic stress who were not recognized as having diabetes beforehand. Hyperglycemia, perhaps worsened by failure to take insulin or hypoglycemic drugs, an infection, or a coincidental medical problem such as a heart attack, leads to urinary losses of water, glucose, and electrolytes (sodium, chloride, and potassium). This osmotic diuresis reduces the circulating blood volume, a stress that results in the release of hormones that worsen insulin resistance and hyperglycemia. In addition, elderly patients may be less able to sense thirst or to obtain fluids. Over the course of several days these patients can become extremely hyperglycemic (glucose >1000 mg dL^{-1}), dehydrated, and comatose. Ketoacidosis does not develop in these patients, possibly because free fatty acids are not always elevated or because adequate insulin concentrations exist in the portal blood to inhibit ketogenesis (although it is not high enough to inhibit gluconeogenesis). Therapy is aimed at restoring water and electrolyte balance and correcting the hyperglycemia with insulin. The mortality of this syndrome is considerably higher than that of diabetic ketoacidosis.

Arieff, A. I., and Carroll, H. J. Nonketotic hyperosmolar coma with hyperglycemia. Clinical features, pathophysiology, renal function, acid-base balance, plasma-cerebrospinal fluid equilibria, and the effects of therapy in 37 cases. *Medicine* 51:73, 1972; and Cruz-Caudillo, J. C., and Sabatini, S. Diabetic hyperosmolar syndrome. *Nephron* 69:201, 1995.

TABLE 13.2 Substrate and Hormone Levels in Blood of Well-Fed, Fasting, and Starving Human[a]

Hormone or Substrate (units)	***Very Well Fed***	***Postabsorptive 12 h***	***Fasted 3 days***	***Starved 5 weeks***
Insulin (μU mL^{-1})	40	15	8	6
Glucagon (pg mL^{-1})	80	100	150	120
Insulin/glucagon ratio (μU pg^{-1})	0.50	0.15	0.05	0.05
Glucose (mM)	6.1	4.8	3.8	3.6
Fatty acids (mM)	0.14	0.6	1.2	1.4
Acetoacetate (mM)	0.04	0.05	0.4	1.3
β-Hydroxybutyrate (mM)	0.03	0.10	1.4	6.0
Lactate (mM)	2.5	0.7	0.7	0.6
Pyruvate (mM)	0.25	0.06	0.04	0.03
Alanine (mM)	0.8	0.3	0.3	0.1
ATP equivalent (mM)	313	290	380	537

Source: From Ruderman, N. B., Aoki, T. T., and Cahill, G. F. Jr. Gluconeogenesis and its disorders in man. In: R. W. Hanson and M. A. Mehlman (Eds.), *Gluconeogenesis, Its Regulation in Mammalian Species*. New York: Wiley, 1976, p. 515.

[a]Data are for normal-weight subjects except for the 5-week starvation values, which are from obese subjects undergoing therapeutic starvation. ATP equivalents were calculated on the basis of the ATP yield expected on complete oxidation of each substrate to CO_2 and H_2O: 38 molecules of ATP for each molecule of glucose; 144 for the average fatty acid (oleate), 23 for acetoacetate; 26 for β-ydroxybutyrate; 18 for lactate, 15 for pyruvate, and 13 (corrected for urea formation) for alanine.

CLINICAL CORRELATION 13.6

Hyperglycemia and Protein Glycation

Glycation of enzymes is known to cause changes in their activity, solubility, and susceptibility to degradation. In the case of hemoglobin A, glycation occurs by a nonenzymatic reaction between glucose and the amino-terminal valine of the β chain. A Schiff base forms between glucose and valine, followed by a rearrangement of the molecule to give a 1-deoxyfructose molecule attached to the valine. The reaction is favored by high glucose levels and the resulting protein, called hemoglobin A_{1c}, is a good index of how high a person's average blood glucose concentration has been over the previous several weeks. The concentration of this modified protein increases in an uncontrolled diabetic and is low in patients who control their glucose level closely.

It has been proposed that glycation of proteins may contribute to the medical complications caused by diabetes, for example, coronary heart disease, retinopathy, nephropathy, cataracts, and neuropathy. Increased glycation of lens proteins may contribute to the development of diabetic cataracts. Collagen, laminin, vitronectin, and other matrix proteins can become glycated and undergo alterations in biological properties, such as self-assembly and binding of other matrix molecules. Glycated proteins and lipoproteins can also be recognized by receptors present on macrophages, which are intimately involved in the formation of atherosclerotic plaques. It is likely that these phenomena favor the accelerated atherosclerosis that occurs in diabetics. The compound aminoguanidine inhibits the formation of the glycation products and is being tested for its ability to prevent diabetic complications.

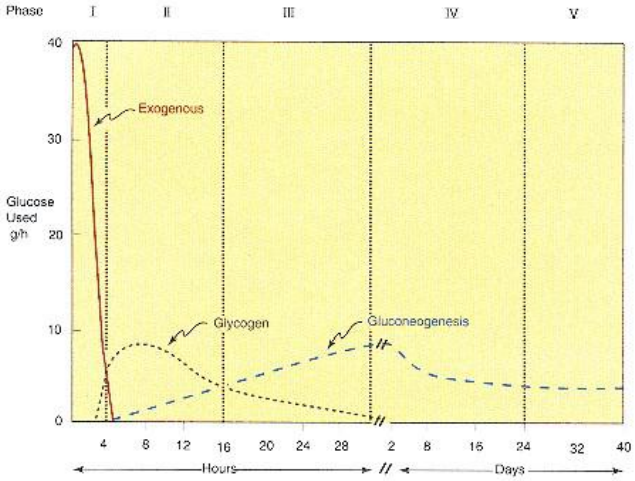
Brownlee, M. Glycation products and the pathogenesis of diabetic complications. *Diabetes Care* 15: 1835, 1992; Vlassara, H. Receptor-mediated interactions with advanced glycosylation end products with cellular components within diabetic tissues. *Diabetes* 41(Suppl 2):52, 1992; and Lyons, T. J. Glycation and oxidation: a role in the pathogenesis of atherosclerosis. *Am. J. Cardiol*. 71:26B, 1993.

in insulin/glucagon ratio shown in Table 13.2 are crucial to the maintenance of caloric homeostasis. Simply stated, well-fed individuals have high insulin/glucagon ratios that favor storage of glycogen and fat, while starving individuals have low insulin/glucagon ratios that stimulate lipolysis, proteolysis, and gluconeogenesis.

Glucose Homeostasis Has Five Phases

Figure 13.10 shows the work of Cahill and his colleagues with obese patients undergoing long-term starvation for weight loss. It illustrates the effects of starvation on those processes that are used to maintain **glucose homeostasis.** For convenience, the time period involved has been divided into five phases. Phase I is the well-fed state in which glucose is provided by dietary carbohydrate. Once this supply is exhausted, hepatic glycogenolysis maintains blood glucose levels during phase II. As this supply of glucose starts to dwindle, hepatic gluconeogenesis from lactate, glycerol, and alanine becomes increasingly important until, in phase III, it is the major source of blood glucose. These changes occur within 20 or so hours of fasting, depending on how well fed the individual was prior to the fast, how much hepatic glycogen was present, and the sort of physical activity occurring during the fast. Several days of fasting move one into phase IV, where the dependence on gluconeogenesis actually decreases. As discussed above, ketone bodies have accumulated to high enough concentrations for them to enter the brain and meet some of its energy needs. Renal gluconeogenesis also becomes significant in this phase. Phase V occurs after very prolonged starvation of extremely obese individuals and is characterized by even less dependence on gluconeogenesis. The energy needs of almost every tissue are met to a large extent by either fatty acid or ketone body oxidation in this phase.

As long as ketone body concentrations are high, proteolysis will be somewhat restricted, and conservation of muscle proteins and enzymes will occur. This continues until practically all of the fat is gone as a consequence of starvation. After all of it is gone, the body has to use muscle protein. Before it is gone—you are gone (see Clin. Corr. 13.3).



Phase	ORIGIN OF BLOOD GLUCOSE	TISSUES USING GLUCOSE	MAJOR FUEL OF BRAIN
I	Exogenous	All	Glucose
II	Glycogen Hepatic gluco-neogenesis	All except liver. Muscle and adipose tissue at diminished rates	Glucose
III	Hepatic gluconeogenesis Glycogen	All except liver. Muscle and adipose tissue at rates intermediate between II and IV	Glucose
IV	Gluconeogenesis, hepatic and renal	Brain, RBCs, renal medulla. Small amount by muscle	Glucose, ketone bodies
V	Gluconeogenesis, hepatic and renal	Brain at a diminished rate, RBCs, renal medulla	Ketone bodies, glucose

Figure 13.10
The five phases of glucose homeostasis in humans.
Reprinted with permission from Ruderman, N. B., Aoki, T. T., and Cahill, G. F. Jr. Gluconeogenesis and its disorders in man. In: R. W. Hanson and M. A. Mehlman (Eds.), *Gluconeogenesis, Its Regulation in Mammalian Species*. New York: Wiley, 1976, p. 515.

13.3— Mechanisms Involved in Switching the Metabolism of Liver between the Well-Fed State and the Starved State

The liver of a well-fed person is actively engaged in processes that favor the synthesis of glycogen and fat; such a liver is glycogenic, glycolytic, and lipogenic. The liver of the fasting person is quite a different organ; it is glycogenolytic, gluconeogenic, ketogenic, and proteolytic. The strategy is to store calories when food is available, but then to mobilize these stores when the rest of the body

is in need. The liver is switched between these metabolic extremes by a variety of regulatory mechanisms: substrate supply, allosteric effectors, covalent modification, and induction–repression of enzymes.

Substrate Availability Controls Many Metabolic Pathways

Because of other, more sophisticated levels of control, the importance of **substrate supply** is often ignored. However, the concentration of fatty acids in blood entering the liver is clearly a major determinant of the rate of ketogenesis. Excess fat is not synthesized unless one consumes excessive amounts of substrates that can be used for lipogenesis. Glucose synthesis by the liver is also restricted by the rate at which gluconeogenic substrates flow to the liver. Delivery of excess amino acids to the liver of the diabetic, because of accelerated and uncontrolled proteolysis, increases the rate of gluconeogenesis and exacerbates the hyperglycemia characteristic of diabetes. In addition, high glucose levels increase the rate of synthesis of sorbitol, which may contribute to diabetic complications. On the other hand, failure to supply the liver adequately with glucogenic substrate (mainly alanine) explains some types of hypoglycemia, such as that observed during pregnancy or advanced starvation.

Another pathway regulated by substrate supply is **urea synthesis.** Amino acid metabolism in the intestine provides a substantial fraction of the ammonia used by the liver for urea production. As discussed above, the intestine also releases citrulline, metabolic precursor of ornithine. A larger ornithine pool permits increased urea synthesis after a high protein meal.

We can conclude that substrate supply is a major determinant of the rate at which virtually every metabolic process of the body operates. However, variations in substrate supply are not sufficient to account for the marked changes in metabolism that must occur in the starve–feed cycle, and finer tuning of the pathways is required.

Negative and Positive Allosteric Effectors Regulate Key Enzymes

Figures 13.11 and 13.12 summarize the effects of negative and positive **allosteric effectors** important in the well-fed and starved states, respectively. As shown in Figure 13.11, glucose inactivates glycogen phosphorylase and activates glycogen synthase (indirectly; see Chapter 7, p. 326), thereby preventing degradation and promoting synthesis of glycogen; fructose 2,6-bisphosphate stimulates 6-phosphofructo-1-kinase and inhibits fructose 1,6-bisphosphatase, thereby stimulating glycolysis and inhibiting gluconeogenesis; fructose 1,6-bisphosphate activates pyruvate kinase, thereby stimulating glycolysis; pyruvate activates pyruvate dehydrogenase (indirectly by inhibition of pyruvate dehydrogenase kinase; see Chapter 6, p. 228); citrate activates acetyl-CoA carboxylase, thereby stimulating fatty acid synthesis; and malonyl CoA inhibits carnitine palmitoyl-transferase I, thereby inhibiting fatty acid oxidation.

As shown in Figure 13.12, acetyl CoA stimulates gluconeogenesis in the fasted state by activating pyruvate carboxylase and inhibiting pyruvate dehydrogenase (a direct allosteric effect and also by stimulation of pyruvate dehydrogenase kinase; see Chapter 7, p. 308); long-chain acyl CoA esters inhibit acetyl-CoA carboxylase, which lowers the level of malonyl CoA and permits greater carnitine palmitoyltransferase I activity and fatty acid oxidation rates; fructose 6-phosphate acts through a regulatory protein to inhibit glucokinase; citrate, which can be increased because of fatty acid oxidation, inhibits 6-phosphofructo-1-kinase as well as 6-phosphofructo-2-kinase (not shown); and NADH produced by fatty acid oxidation inhibits TCA cycle activity.

Although not shown in Figure 13.12, **cAMP** is an important allosteric effector. Its concentration in liver is increased in the starved state. Cyclic AMP is a positive effector of cAMP-dependent protein kinase (also called **protein**

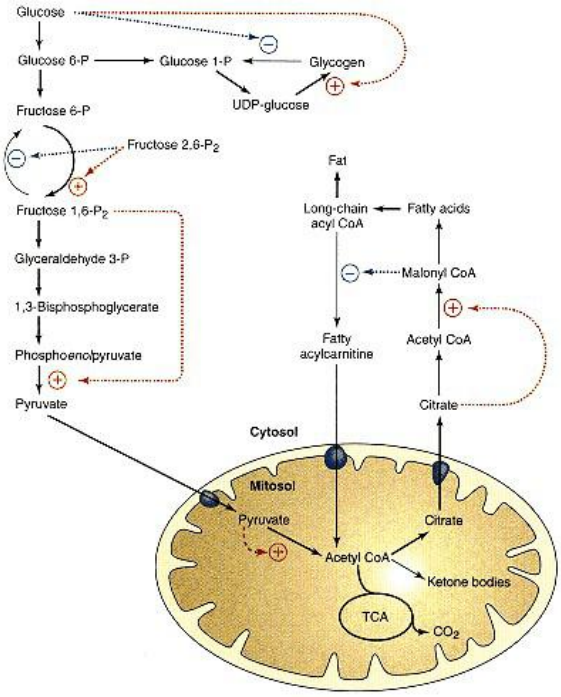
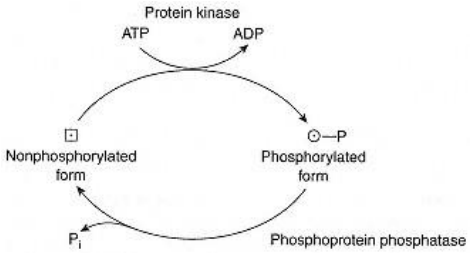


Figure 13.11
Control of hepatic metabolism in the well-fed state by allosteric effectors.

kinase A), which, in turn, is responsible for changing the kinetic properties of several regulatory enzymes by covalent modification, as summarized next.

Covalent Modification Regulates Key Enzymes

Figures 13.13 and 13.14 point out the interconvertible enzymes that play important roles in switching the liver between the well-fed and starved states. The regulation of enzymes by **covalent modification** has been discussed in Chapter 7. Recall that ⊙-P represent interconvertible forms of an enzyme in the nonphosphorylated and phosphorylated states, respectively.



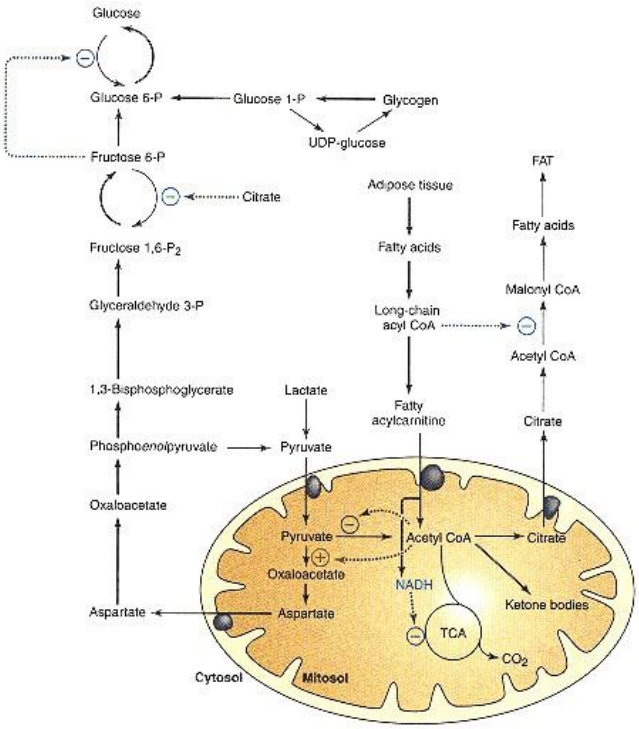


Figure 13.12
Control of hepatic metabolism in the fasting state by allosteric effectors.

The important points are as follows: (1) enzymes subject to covalent modification undergo phosphorylation on one or more serine residues by a protein kinase; (2) the phosphorylated enzyme can be returned to the dephosphorylated state by phosphoprotein phosphatase; (3) phosphorylation of the enzyme changes its conformation and its catalytic activity; (4) some enzymes are active only in the dephosphorylated state, others only in the phosphorylated state; (5) cAMP is the messenger that signals the phosphorylation of many, but not all, of the enzymes subject to covalent modification; (6) cAMP acts by activating protein kinase A; (7) cAMP also indirectly promotes phosphorylation of interconvertible enzymes by signaling inactivation of phosphoprotein phosphatase; (8) glucagon and β-adrenergic agonists (epinephrine) increase cAMP levels by activating adenylate cyclase; (9) insulin (see Chapter 20, p. 879) opposes the action of glucagon and epinephrine, in part by lowering cAMP and in part by mechanisms independent of cAMP; and (10) the action of insulin in general promotes dephosphorylation of interconvertible enzymes.

Hepatic enzymes subject to covalent modification are dephosphorylated in well-fed animals (Figure 13.13). Although not shown, phosphorylase kinase is also dephosphorylated in this state. Insulin/glucagon ratios are high in blood, and cAMP levels are low in liver. This results in low activity of protein kinase A and high activity of **phosphoprotein phosphatase.** Glycogen synthase, glycogen phosphorylase (via phosphorylase kinase), 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (bifunctional enzyme), pyruvate kinase, and

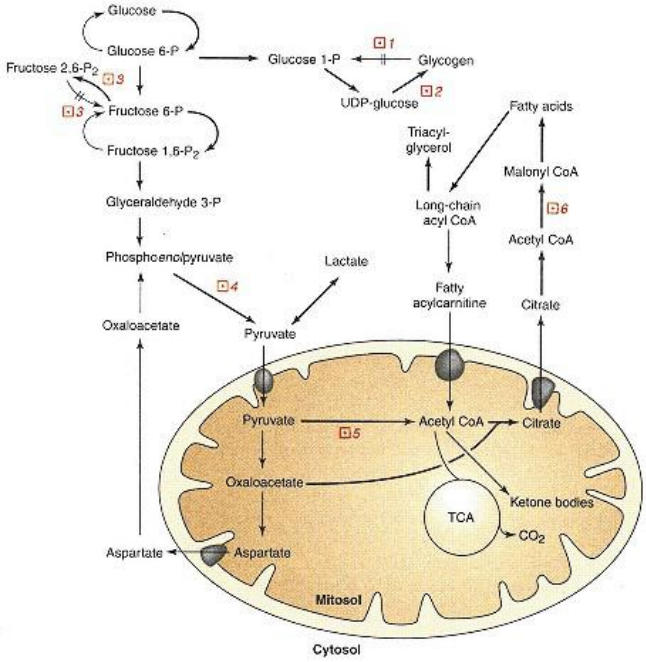


Figure 13.13
Activity and state of phosphorylation of enzymes subject to covalent modification in the lipogenic liver.

The dephosphorylated mode is indicated by the symbol ⊡. Interconvertible enzymes are numbered as follows:
1, glycogen phosphorylase;
2, glycogen synthase;
3, 6-phosphofructo -2-kinase/fructose 2,6-bisphosphatase (bifunctional enzyme);
4, pyruvate kinase;
5, pyruvate dehydrogenase; and
6, acetyl-CoA carboxylase.

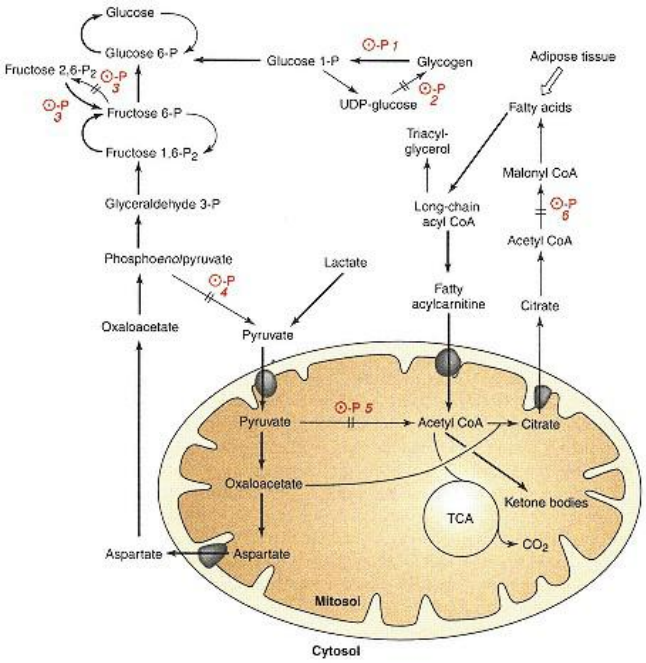


Figure 13.14
Activity and state of phosphorylation of enzymes subject to covalent modification in the glucogenic liver.

Phosphorylated mode is indicated by the symbol ⊙-P. Numbers refer to the same enzymes as in Figure 13.13.

acetyl-CoA carboxylase are phosphorylated by protein kinase A. However, not all interconvertible enzymes are subject to phosphorylation by protein kinase A. No link to protein kinase A for the pyruvate dehydrogenase complex has been established. Only three of the interconvertible enzymes—**glycogen phosphorylase,** phosphorylase kinase, and the **fructose 2,6-bisphosphatase** of the bifunctional enzyme—are inactive when dephosphorylated. All of the other identified interconvertible enzymes are active. Glycogenesis, glycolysis, and lipogenesis are greatly favored when these enzymes are dephosphorylated. On the other hand, the opposing pathways—glycogenolysis, gluconeogenesis, and ketogenesis—are inhibited.

As shown in Figure 13.14 (p. 543), the hepatic enzymes subject to covalent modification are in the phosphorylated mode in the liver of the fasting animal. Insulin is low but glucagon is high in the blood, resulting in an increase in hepatic cAMP levels. This activates protein kinase A and inactivates phosphoprotein phosphatase. The net effect is a greater degree of phosphorylation of interconvertible enzymes than in the well-fed state. In the starved state, three interconvertible enzymes—glycogen phosphorylase, phosphorylase kinase, and the fructose 2,6-bisphosphatase of the bifunctional enzyme—are in the active catalytic state. All the other interconvertible enzymes are inactive in the phosphorylated mode. As a result, glycogenesis, glycolysis, and lipogenesis are shut down almost completely, and glycogenolysis, gluconeogenesis, and ketogenesis predominate.

Two additional hepatic enzymes, **phenylalanine hydroxylase** and **branched-chain α-keto acid dehydrogenase,** are also controlled by phosphorylation/dephosphorylation. These enzymes catalyze rate-limiting steps in the disposal of phenylalanine and the branched-chain amino acids (leucine, isoleucine, and valine), respectively. These enzymes are not included in Figures 13.13 and 13.14 because of special features of their control by covalent modification. Phenylalanine hydroxylase, a cytosolic enzyme, is active in the phosphorylated state, and phosphorylation is stimulated by glucagon via protein kinase A. Branched-chain α-keto acid dehydrogenase, a mitochondrial enzyme, is active in the dephosphorylated state, and its activity is regulated by branched-chain α-keto acid dehydrogenase kinase and a phosphoprotein phosphatase. Phenylalanine acts as a positive allosteric effector for the phosphorylation and activation of phenylalanine hydroxylase by cAMP-dependent protein kinase. Branched-chain α-keto acids activate branched-chain α-keto acid dehydrogenase indirectly by inhibiting branched-chain α-keto acid dehydrogenase kinase. Covalent modification of these enzymes provides a very sensitive means for control of the degradation of phenylalanine and the branched-chain amino acids. The clinical experience with phenylketonuria (see Clin. Corr. 11.5) and maple syrup urine disease (see Clin. Corr. 11.10) emphasizes the importance of regulating blood and tissue levels of these amino acids. Of note, the artificial sweetener **aspartame** (NutraSweet®) is *N*-aspartylphenylalanine methyl ester. The amount in a liter of sweetened drinks may approach the amount of phenylalanine normally obtained from the daily diet. This is of no harm to normal individuals but is a threat to **phenylketonuria** patients on a low phenylalanine diet. Phenylalanine and the branched-chain amino acids cannot be synthesized in humans, making them essential amino acids that must be available continuously for protein synthesis. Thus the activities of phenylalanine hydroxylase and branched-chain α-keto acid dehydrogenase must be carefully controlled to prevent depletion of body stores. Therefore the tissue requirement for these amino acids supersedes the phase of the starve–feed cycle in establishing the phosphorylation and activity state of these interconvertible enzymes.

Adipose tissue responds almost as dramatically as liver to the starve–feed cycle because it also contains enzymes subject to covalent modification. Pyruvate kinase, pyruvate dehydrogenase, acetyl-CoA carboxylase, and hormone-

sensitive lipase (not found in liver) are all in the dephosphorylated mode in the adipose tissue of the well-fed person. As in liver, the first three enzymes are active when dephosphorylated. **Hormone-sensitive lipase** is inactive when dephosphorylated. A high insulin level in the blood and a low cAMP concentration in adipose tissue are important determinants of the phosphorylation state of these enzymes, which favors lipogenesis in the well-fed state. During fasting, as a consequence of the decrease in the insulin level and an increase in epinephrine, adipocytes quickly shut down lipogenesis and activate lipolysis. This is accomplished in large part by the phosphorylation of the enzymes described above. In this manner, adipose tissue is transformed from a fat storage tissue into a source of fatty acids for oxidation in other tissues and glycerol for gluconeogenesis in the liver.

Conservation of glucose as well as three-carbon compounds that can readily be converted to glucose (lactate, alanine, and pyruvate) by the liver is crucial for survival in the starved state. Certain cells, particularly those of the central nervous system, are absolutely dependent on a continuous supply of glucose. Tissues that can use alternative fuels invariably shut down their use of glucose and three-carbon precursors. This is referred to as the glucose–fatty acid cycle in recognition that increased availability of fatty acids for oxidation spares glucose in the starved state. Inactivation of the pyruvate dehydrogenase complex by phosphorylation is an important feature of the **glucose–fatty acid cycle.** This occurs in skeletal muscle, heart, and kidney, but not in the central nervous system, when the alternative fuels (fatty acids and ketone bodies) of the starved state become abundant. Activation of pyruvate dehydrogenase kinase by products of the catabolism of the alternative fuels (acetyl CoA and NADH) is responsible for the greater degree of phosphorylation and therefore lower activity of the pyruvate dehydrogenase complex.

Covalent modification, like allosteric effectors and substrate supply, is a short-term regulatory mechanism, operating on a minute-to-minute basis. On a longer time scale, enzyme activities are controlled at the level of expression.

Changes in Levels of Key Enzymes Are a Longer Term Adaptive Mechanism

The adaptive change in enzyme levels is a mechanism of regulation involving changes in the rate of synthesis or degradation of key enzymes. Whereas allosteric effectors and covalent modification affect either the K_m or V_{max} of an enzyme, this mode of regulation involves the actual quantity of an enzyme in a tissue. Because of the influence of hormonal and nutritional factors, there are more or fewer enzyme molecules present in the tissue. For example, when a person is maintained in a well-fed or overfed condition, the liver improves its capacity to synthesize fat. This can be explained in part by increased substrate supply, appropriate changes in allosteric effectors (Figure 13.11), and the conversion of the interconvertible enzymes into the dephosphorylated form (Figure 13.13). This is not the entire story, however, because the liver also has more of those enzyme molecules that play a key role in fat synthesis (see Figure 13.15). A whole battery of enzymes is induced, including glucokinase, 6-phospho-1-fructokinase, and pyruvate kinase for faster rates of glycolysis; glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, and malic enzyme to provide greater quantities of NADPH for reductive synthesis; and citrate cleavage enzyme, acetyl-CoA carboxylase, fatty acid synthase, and ^{9}-desaturase for more rapid rates of fatty acid synthesis. All of these enzymes are present at higher levels in the well-fed state because of an increase in the blood of the insulin/glucagon ratio and glucose. While these enzymes are induced, there is a decrease in the enzymes that favor glucose synthesis. **Phosphoenolpyruvate carboxykinase,** fructose 1,6-bisphosphatase, **glucose 6-phosphatase,** and

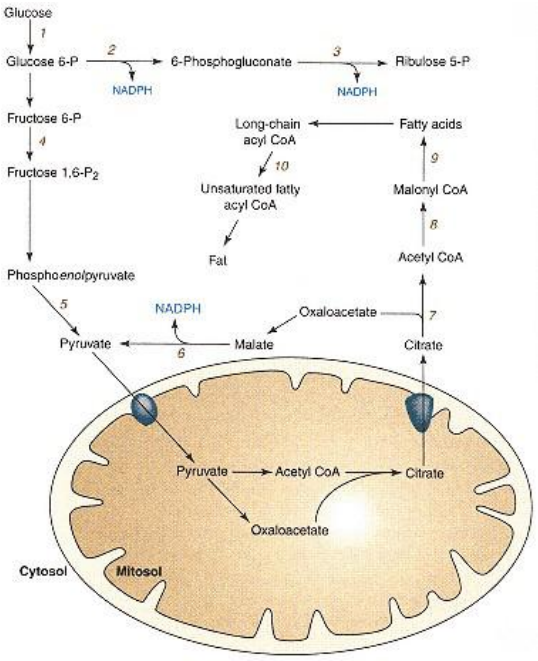


Figure 13.15
Enzymes induced in the liver of the well-fed individual.
Inducible enzymes are numbered as follows:
1, glucokinase;
2, glucose-6-phosphate dehydrogenase;
3, 6-phosphogluconate dehydrogenase;
4, 6-phosphofructo-1-kinase;
5, pyruvate kinase,
6, malic enzyme;
7, citrate cleavage enzyme;
8, acetyl-CoA carboxylase;
9, fatty acid synthase; and
10, ^{9}-desaturase.

some aminotransferases are decreased in amount; that is, their synthesis is reduced or degradation increased in response to increased circulating glucose and insulin.

In fasting, the enzyme pattern of the liver changes dramatically (Figure 13.16). The enzymes involved in lipogenesis decrease in quantity, possibly because their synthesis is decreased or degradation of these proteins is increased. At the same time a number of enzymes (glucose 6-phosphatase, fructose 1,6-bisphosphatase, phosphoenolpyruvate carboxykinase, and various amino-transferases) favoring gluconeogenesis are induced, making the liver much more effective in synthesizing glucose. In addition, the enzymes of the urea cycle and other amino acid-metabolizing enzymes such as liver **glutaminase,** tyrosine aminotransferase, serine dehydratase, proline oxidase, and histidase are induced, possibly by the presence of higher blood glucagon levels. This permits the disposal of nitrogen, as urea, from the amino acids used in gluconeogenesis.

These adaptive changes are clearly important in the starve–feed cycle, greatly affecting the capacity of the liver for its various metabolic processes. The adaptive changes also influence the effectiveness of the short-term regulatory mechanisms. For example, long-term starvation or uncontrolled diabetes decreases the level of acetyl-CoA carboxylase. Taking away long-chain acyl CoA esters that inhibit this enzyme, increasing the level of citrate that activates this enzyme, or creating conditions that activate this interconvertible enzyme by dephosphorylation will not have any effect when the enzyme is virtually absent. Another example is afforded by the **glucose intolerance** of starvation. A chronically starved person cannot effectively utilize a load of glucose because of the absence of the key enzymes needed for glucose metabolism. A glucose

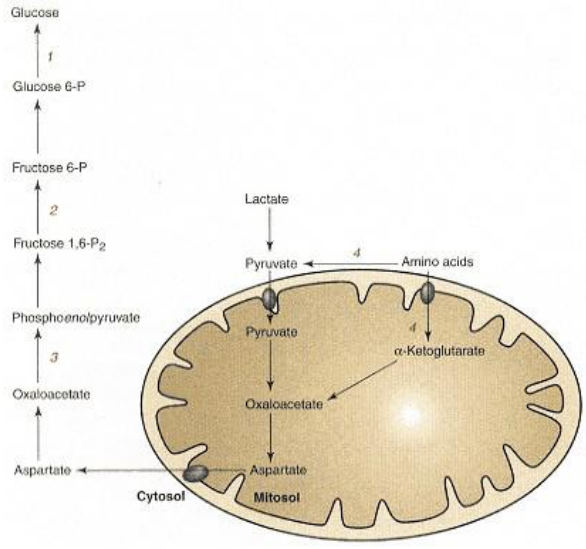


Figure 13.16
Enzymes induced in the liver of an individual during fasting.
Inducible enzymes are numbered as follows:
1, glucose 6-phosphatase;
2, fructose 1,6-bisphosphatase;
3, phosphoenolpyruvate carboxykinase; and
4, various aminotransferases.

load, however, will set into motion the induction of the required enzymes and the reestablishment of short-term regulatory mechanisms.

13.4— Metabolic Interrelationships of Tissues in Various Nutritional and Hormonal States

Many changes that occur in various nutritional and hormonal states are variations on the starve–feed cycle and are completely predictable from what we have learned about the cycle. Some examples are given in Figure 13.17. Others are so obvious that a diagram is unnecessary; for example, in rapid growth of a child, amino acids are directed away from catabolism and into protein synthesis. However, the changes that occur in some physiologically important situations are rather subtle and poorly understood. An example of the latter is **aging,** which seems to lead to a decreased "sensitivity" of the major tissues of the body to hormones. The important consequence is a decreased ability of the tissues to respond normally during the starve–feed cycle. Whether this is a contributing factor to or a consequence of the aging process is unknown.

Staying in the Well-Fed State Results in Obesity and Insulin Resistance

Figure 13.17*a* illustrates the metabolic interrelationships prevailing in an obese person. Most of the body fat of the human is either provided by the diet or synthesized in the liver and transported to the adipose tissue for storage. Obesity is caused by a person staying in such a well-fed state that stored fat does not get used up during the fasting phase of the cycle. The body then has no option other than to accumulate fat (see Clin. Corr. 13.1).

Obesity always causes some degree of **insulin resistance.** Insulin resistance is a poorly understood phenomenon in which the tissues fail to respond to insulin. The number or affinity of insulin receptors is reduced in some patients; others have normal insulin binding, but abnormal postreceptor responses, such

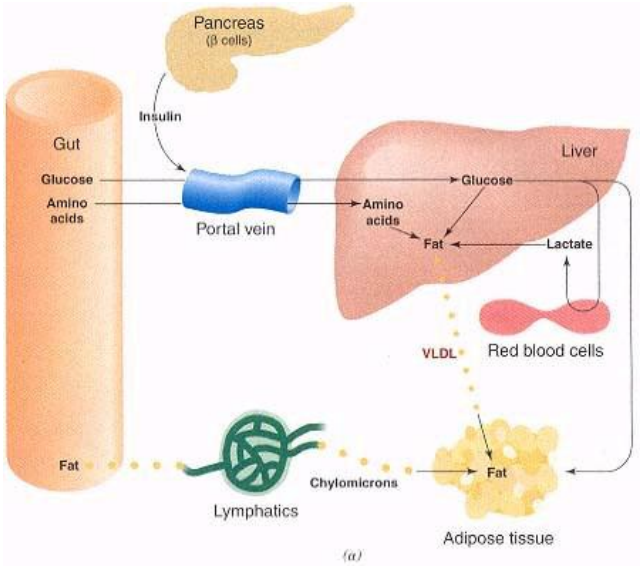


Figure 13.17
Metabolic interrelationships of tissues in various nutritional, hormonal, and disease states.
(a) Obesity.

as the activation of glucose transport. As a general rule, the greater the quantity of body fat, the greater the resistance of normally insulin-sensitive cells to the action of insulin. Current research suggests that high expression of **tumor necrosis factor-α (TNF-α)** in the fat cells of obese individuals contributes to insulin resistance. As a consequence, plasma insulin levels are greatly elevated in the blood of an obese individual. As long as the β cells of the pancreas produce enough insulin to overcome the insulin resistance, an obese individual will have relatively normal blood levels of glucose and lipoproteins. The insulin resistance of obesity can lead, however, to the development of noninsulin-dependent diabetes, as discussed next.

Noninsulin-Dependent Diabetes Mellitus

Figure 13.17b shows the metabolic interrelationships characteristic of a person with **noninsulin-dependent diabetes.** In contrast to insulin-dependent diabetes, insulin is not absent in noninsulin-dependent diabetes (see Clin. Corr. 13.7). Indeed, high levels of insulin may be observed in this form of diabetes, and the problem is primarily resistance to the action of insulin as discussed above for obese individuals. It therefore follows that the majority of patients with noninsulin-dependent diabetes mellitus are obese. Although the insulin levels of noninsulin-dependent diabetic patients may and often are high, they are not as high as those of a nondiabetic but similarly obese person. The pancreases of these diabetic patients do not produce enough insulin to overcome the insulin resistance induced by their obesity. Hence this form of diabetes is also a form of β-cell failure; exogenous insulin will reduce the hyperglycemia and very often must be administered to control blood glucose levels of noninsulin-dependent diabetic patients. Hyperglycemia results mainly because of poor uptake of glucose by peripheral tissues, especially muscle. In contrast to insulin-dependent diabetes, ketoacidosis does not develop because the adipocytes

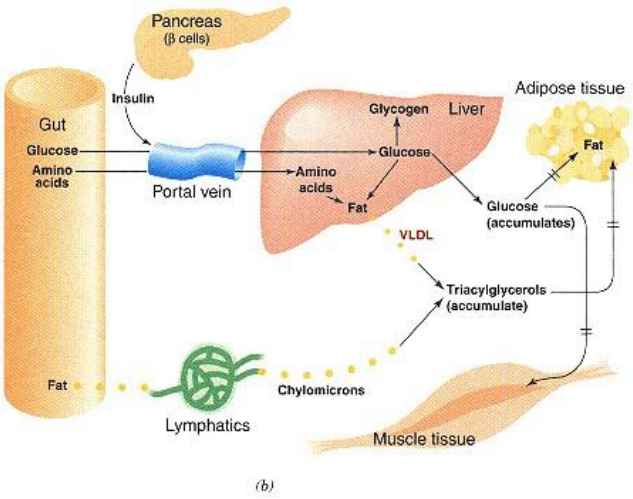


Figure 13.17
(continued)
(b) Noninsulin-dependent diabetes mellitus.

remain sensitive to the effect of insulin on lipolysis. Hypertriglyceridemia is characteristic of noninsulin-dependent diabetes but usually results from an increase in VLDLs without **hyperchylomicronemia.** This is most likely explained by rapid rates of *de novo* hepatic synthesis of fatty acids and VLDLs rather than increased delivery of fatty acids from the adipose tissue.

CLINICAL CORRELATION 13.7

Noninsulin-Dependent Diabetes Mellitus

Noninsulin-dependent diabetes mellitus (NIDDM) accounts for 80–90% of the diagnosed cases of diabetes and is also called maturity-onset diabetes to differentiate it from insulin-dependent, juvenile diabetes. It usually occurs in middle-aged obese people. Noninsulin-dependent diabetes is characterized by hyperglycemia, often with hypertriglyceridemia. The ketoacidosis characteristic of the insulin-dependent disease is not observed. Increased levels of VLDL are probably the result of increased hepatic triacylglycerol synthesis stimulated by hyperglycemia and hyperinsulinemia. Insulin is present at normal to elevated levels in this form of the disease. Obesity often precedes the development of insulin-independent diabetes and appears to be the major contributing factor. Obese patients are usually hyperinsulinemic. Very recent data implicate increased levels of expression of tumor necrosis factor-α (TNF-α) in adipocytes of obese individuals as a cause of the resistance. The greater the adipose tissue mass, the greater the production of TNF-α, which acts to impair insulin receptor function. An inverse relationship between insulin levels and the number of insulin receptors has been established. The higher the basal level of insulin, the fewer receptors present on the plasma membranes. In addition, there are defects within insulin-responsive cells at sites beyond the receptor. An example is the ability of insulin to recruit glucose transporters from intracellular sites to the plasma membrane. As a consequence, insulin levels remain high, but glucose levels are poorly controlled because of the lack of normal responsiveness to insulin. Although the insulin level is high, it is not as high as in a person who is obese but not diabetic. In other words, there is a relative deficiency in the insulin supply from the β cells. Therefore, this disease is caused not only by insulin resistance but also by impaired β-cell function resulting in relative insulin deficiency. Diet alone can often control the disease in the obese diabetic. If the patient can be motivated to lose weight, insulin receptors will increase in number, and the postreceptor abnormalities will improve, which will increase both tissue sensitivity to insulin and glucose tolerance. The noninsulin-dependent diabetic tends not to develop ketoacidosis but nevertheless develops many of the same complications as the insulin-dependent diabetic, that is, nerve, eye, kidney, and coronary artery disease.

Olefsky, J. M., and Kolterman, O. G. Mechanisms of insulin resistance in obesity and non-insulin dependent (type II) diabetes. *Am. J. Med.* 70:151, 1981; Flier, J. S. The adipocyte: storage depot or node on the energy information superhighway? *Cell*80:15, 1995; and Ruderman, N. B., Williamson, J. R., and Brownlee, M. Glucose and diabetic vascular disease. *FASEB J.* 6:2905, 1992.

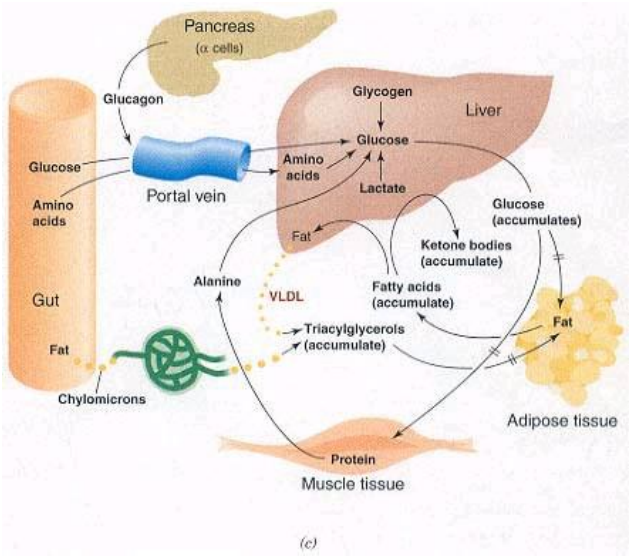


Figure 13.17
(continued)
(c) Insulin-dependent diabetes mellitus.

CLINICAL CORRELATION 13.8

Insulin-Dependent Diabetes Mellitus

Insulin-dependent diabetes mellitus (IDDM) was once called juvenile-onset diabetes because it usually appears in childhood or in the teens, but it is not limited to these patients. Insulin is absent in this disease because of defective or absent β cells in the pancreas. The β cells are destroyed by an autoimmune process. Untreated, IDDM is characterized by hyperglycemia, hyperlipoproteinemia (chylomicrons and VLDLs), and episodes of severe ketoacidosis. Far from being a disease of defects in carbohydrate metabolism alone, diabetes causes abnormalities in fat and protein metabolism in such patients as well. The hyperglycemia results in part from the inability of the insulin-dependent tissues to take up plasma glucose and in part by accelerated hepatic gluconeogenesis from amino acids derived from muscle protein. The ketoacidosis results from increased lipolysis in the adipose tissue and accelerated fatty acid oxidation in the liver. Hyperchylomicronemia is the result of low lipoprotein lipase activity in adipose tissue capillaries, an enzyme dependent on insulin for its synthesis.

Although insulin does not cure the diabetes, its use markedly alters the clinical course of the disease. The injected insulin promotes glucose uptake by tissues and inhibits gluconeogenesis, lipolysis, and proteolysis. The patient has the difficult job of trying to adjust the insulin dose to a variable dietary intake and variable physical activity, the other major determinant of glucose disposal by muscle. Tight control demands the use of several injections of insulin per day and close blood sugar monitoring by the patient. Tight control of blood sugar has now been proved to reduce the microvascular complications of diabetes (renal and retinal diseases).

National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 28:1039, 1977; Atkinson, M. A., and Maclaren, N. K. The pathogenesis of insulin dependent diabetes mellitus. *N. Engl. J. Med.* 331:1428, 1991; and Clark, C. M., and Lee, D. A. Prevention and treatment of the complications of diabetes mellitus. *N. Engl. J. Med.* 332:1210, 1994.

Insulin-Dependent Diabetes Mellitus

Figure 13.17*c* shows the metabolic interrelationships that exist in **insulin-dependent diabetes mellitus** (see Clin. Corr. 13.8 and 13.9). In contrast to noninsulin-dependent diabetes, there is a complete absence of insulin production by the pancreas in this disease. Because of defective β-cell production of insulin, blood levels of insulin do not increase in response to elevated blood glucose levels. Even when dietary glucose is being delivered from the gut, the insulin/glucagon ratio cannot increase, and the liver remains gluconeogenic and ketogenic. Since it is impossible to switch to the processes of glycolysis, glycogenesis, and lipogenesis, the liver cannot properly buffer blood glucose levels. Indeed, since hepatic gluconeogenesis is continuous, the liver contributes to hyperglycemia in the well-fed state. Failure of some tissues, especially muscle, to take up glucose in the absence of insulin contributes further to the hyperglycemia. Accelerated gluconeogenesis, fueled by substrate made available by tissue protein degradation, maintains the hyperglycemia even in the starved state.

The absence of insulin in patients with insulin-dependent diabetes mellitus results in uncontrolled rates of lipolysis in adipose tissue. This increases blood levels of fatty acids and results in accelerated ketone body production by the liver. If ketone bodies are not used as rapidly as they are formed, diabetic ketoacidosis develops due to accumulation of ketone bodies and hydrogen ions. Not all the fatty acid taken up by liver can be handled by the pathway of fatty acid oxidation and ketogenesis. The excess is esterified and directed into VLDL synthesis. **Hypertriglyceridemia** results because VLDLs are synthesized and released by the liver more rapidly than these particles can be cleared from the blood by lipoprotein lipase. The quantity of this enzyme is dependent on the blood insulin level. The defect in lipoprotein lipase also results in hyperchylomicronemia, since lipoprotein lipase is required for chylomicron catabolism in adipose tissue. In summary, in diabetes every tissue continues to play the catabolic role that it was designed to play in starvation, in spite of

delivery of adequate or even excess fuel from the gut. This results in a gross elevation of all fuels in the blood with severe wasting of body tissues and ultimately death unless insulin is administered.

CLINICAL CORRELATION 13.9

Complications of Diabetes and the Polyol Pathway

Diabetes is complicated by several disorders that may share a common pathogenesis. The lens, peripheral nerve, renal papillae, Schwann cells, glomerulus, and possibly retinal capillaries contain two enzymes that constitute the polyol pathway (the term polyol refers to polyhydroxy sugars). The first is aldose reductase, an NADPH-requiring enzyme. It reduces glucose to form sorbitol. Sorbitol is further metabolized by sorbitol dehydrogenase, an NAD^+-requiring enzyme that oxidizes sorbitol to fructose. Aldose reductase has a high K_m for glucose; therefore this pathway is only quantitatively important during hyperglycemia. It is known that in diabetic animals the sorbitol content of lens, nerve, and glomerulus is elevated. Sorbitol accumulation may damage these tissues by causing them to swell. There are now inhibitors of the reductase that prevent the accumulation of sorbitol in these tissues and thus retard the onset of these complications. This is a very controversial area because differences in potency of the inhibitors, experimental designs, length of trials, and the numbers of patients enrolled have resulted in different studies reaching different conclusions. We cannot as yet confidently recommend these drugs to prevent diabetic complications.

Gabbay, K. H. Hyperglycemia, polyol metabolism, and the complications of diabetes mellitus. *Annu. Rev. Med.* 26:521, 1975; Frank, R. N. The aldose reductase controversy. *Diabetes* 43:169, 1994; and Clark, C. M., and Lee, D. A. Prevention and treatment of the complications of diabetes mellitus. *N. Engl. J. Med.* 332:1210, 1994.

Aerobic and Anaerobic Exercise Use Different Fuels

It is important to differentiate between two distinct types of **exercise**—aerobic and anaerobic. Aerobic exercise is exemplified by long-distance running, anaerobic exercise by sprinting or weight lifting. During anaerobic exercise there is really very little interorgan cooperation. The blood vessels within the muscles are compressed during peak contraction, thus their cells are isolated from the rest of the body. Muscle largely relies on its own stored glycogen and **phosphocreatine.** Phosphocreatine serves as a source of high-energy phosphate for ATP synthesis (Figure 13.7) until glycogenolysis and glycolysis are stimulated. Glycolysis becomes the primary source of ATP for want of oxygen. Aerobic exercise is metabolically more interesting (Figure 13.17*d*). For moderate exercise, much of the energy is derived from glycolysis of muscle glycogen. This biochemical fact is the basis for **carbohydrate loading.** Muscle glycogen content can be increased by exhaustive exercise that depletes glycogen, followed by rest and a high-carbohydrate diet. There is also stimulation of branched-chain amino acid oxidation, ammonium production, and alanine release from the exercising muscle. However, a well-fed individual does not store enough glucose and glycogen to provide the energy needed for running long distances. The **respiratory quotient,** the ratio of carbon dioxide exhaled to oxygen consumed, falls during distance running. This indicates the progressive switch from glycogen to fatty acid oxidation during a race. Lipolysis gradually increases as glucose stores are exhausted, and, as in the fasted state, muscles oxidize fatty acids in preference to glucose as the former become available.

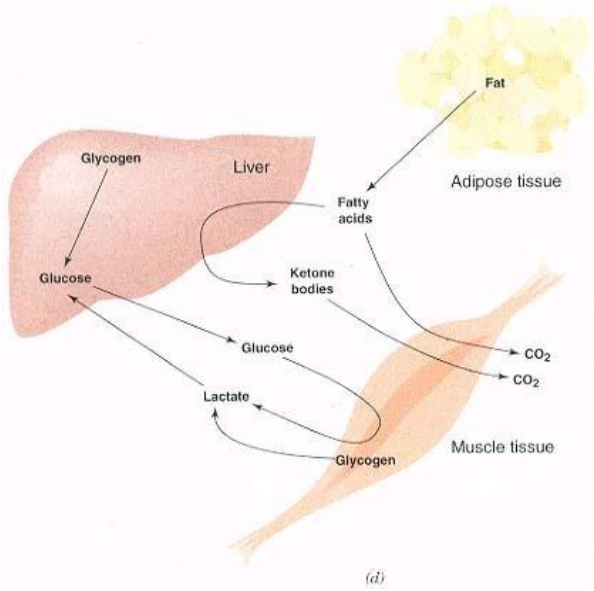


Figure 13.17
(continued)
(d) Exercise.

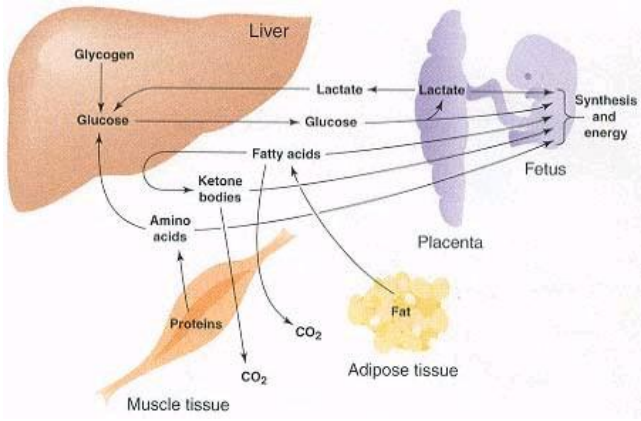


Figure 13.17
(continued)
(e) Pregnancy.

Unlike fasting, there is little increase in blood ketone body concentration. This may reflect a balance between hepatic ketone body synthesis and muscle ketone body oxidation.

Changes in Pregnancy Are Related to Fetal Requirements and Hormonal Changes

The **fetus** can be considered as another nutrient-requiring tissue (Figure 13.17*e*). It mainly uses glucose for energy but may also use amino acids, lactate, fatty acids, and ketone bodies. Lactate produced in the **placenta** by glycolysis goes in two directions. Part of it is directed to the fetus where it serves as a fuel, with the rest returning to the maternal circulation to establish a Cori cycle with the liver. Maternal LDL cholesterol is an important precursor of placental steroids (estradiol and progesterone). During **pregnancy,** the starve–feed cycle is perturbed. The placenta secretes a polypeptide hormone, **placental lactogen,** and two steroid hormones, estradiol and progesterone. Placental lactogen stimulates lipolysis in adipose tissue, and the steroid hormones induce an insulin-resistant state. Thus, in the postprandial state, pregnant women enter the starved state more rapidly than do nonpregnant women. This results from increased consumption of glucose and amino acids by the fetus. Plasma glucose, amino acids, and insulin levels fall rapidly, and glucagon and placental lactogen levels rise and stimulate lipolysis and ketogenesis. The consumption of glucose and amino acids by the fetus may be great enough to cause maternal hypoglycemia. On the other hand, in the fed state pregnant women have increased levels of insulin and glucose and demonstrate resistance to exogenous insulin. These swings of plasma hormones and fuels are even more exaggerated in pregnant diabetic women and make control of their blood glucose difficult.

Lactation Requires Synthesis of Lactose, Triacylglycerol, and Protein

In late pregnancy placental hormones induce lipoprotein lipase in the mammary gland and promote the development of milk-secreting cells and ducts. During **lactation** (see Figure 13.17*f*) the breast utilizes glucose for **lactose** and triacylglycerol synthesis, as well as its major energy source. Amino acids are taken

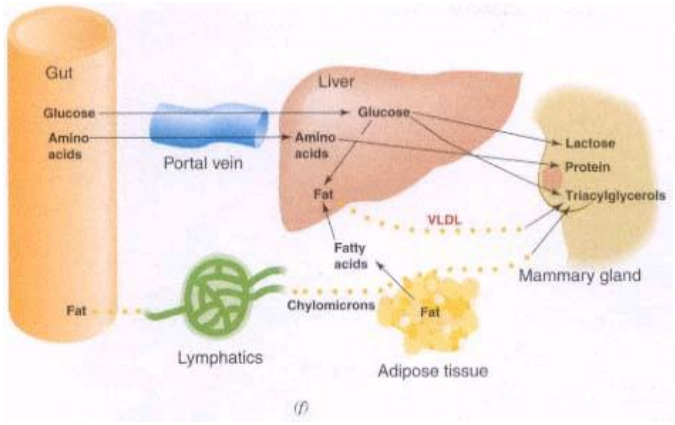


Figure 13.17
(continued)
(f) Lactation.

up for protein synthesis, and chylomicrons and VLDLs are utilized as sources of fatty acids for triacylglycerol synthesis. If these compounds are not supplied by the diet, proteolysis, gluconeogenesis, and lipolysis must supply them, resulting eventually in maternal malnutrition and poor quality milk. The lactating breast also secretes a hormone with some similarity to parathyroid hormone (see Chapter 20). This hormone probably is important for the absorption of calcium and phosphorus from the gut and bone.

Stress and Injury Lead to Metabolic Changes

Physiological stresses include **injury, surgery, renal failure, burns,** and **infections** (Figure 13.17*g*). Characteristically, blood cortisol, glucagon, **catecholamines,** and **growth hormone** levels increase. The patient is resistant to insulin. Basal metabolic rate, blood glucose, and free fatty acid levels are elevated. However, ketogenesis is not accelerated as in fasting. For incompletely understood reasons, the intracellular muscle glutamine pool is reduced, resulting in reduced protein synthesis and increased protein breakdown. It can be very difficult to reverse this protein breakdown, although now it is common to replace amino acids, glucose, and fat by infusing solutions of these nutrients intravenously. However, these solutions lack glutamine, tyrosine, and cysteine because of stability and solubility constraints. Supplementation of these amino acids, perhaps by the use of more stable dipeptides, may help to reverse the catabolic state better than can be accomplished at present.

It has been proposed that the negative nitrogen balance of injured or infected patients is mediated by monocyte and lymphocyte proteins, such as **interleukin-1,** interleukin-6, and **TNF-α** (see Clin. Corr. 13.10). These cytokines are responsible for causing fever as well as a number of other metabolic changes. Interleukin-1 activates proteolysis in skeletal muscle. **Interleukin-6** stimulates the synthesis of a number of hepatic proteins called **acute phase reactants** by the liver. Acute phase reactants include fibrinogen, complement proteins, some clotting factors, and α_2-macroglobulin, which are presumed to play a role in defense against injury and infection. TNF-α suppresses adipocyte fat synthesis, prevents uptake of circulating fat by inhibiting lipoprotein lipase, stimulates lipolysis, inhibits release of insulin, and promotes insulin resistance. These cytokines appear responsible for much of the wasting seen in chronic infections.

CLINICAL CORRELATION 13.10

Cancer Cachexia

Unexplained weight loss may be a sign of malignancy, and weight loss is common in advanced cancer. Decreased appetite and food intake contribute to but do not entirely account for the weight loss. The weight loss is largely from skeletal muscle and adipose tissue, with relative sparing of visceral protein (i.e., liver, kidney, and heart). Although tumors commonly exhibit high rates of glycolysis and release lactate, the energy requirement of the tumor probably does not explain weight loss because weight loss can occur with even small tumors. In addition, the presence of another energy-requiring growth, the fetus in a pregnant woman, does not normally lead to weight loss. Several endocrine abnormalities have been recognized in cancer patients. They tend to be insulin-resistant, have higher cortisol levels, and have a higher basal metabolic rate compared with controls matched for weight loss. Two other phenomena may contribute to the metabolic disturbances. Some tumors synthesize and secrete biologically active peptides such as ACTH, nerve growth factor, and insulin-like growth factors, which could modify the endocrine regulation of energy metabolism. It is also possible that the host response to a tumor, by analogy to chronic infection, includes release of interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor-α(TNF-α) by cells of the immune system. TNF-α is also called cachexin because it produces wasting. TNF-α and IL-1 may act in a paracrine fashion, as plasma levels are not elevated. They do induce the synthesis of IL-6, which has been detected in cachectic patients' sera at increased levels. These cytokines stimulate fever, proteolysis, lipolysis, and the synthesis of acute phase reactants by the liver.

Beutler, B., and Cerami, A. Tumor necrosis, cachexia, shock, and inflammation: a common mediator. *Annu. Rev. Biochem.* 57:1505, 1988; and Tracey, K. J., and Cerami, A. Tumor necrosis factor: a pleiotropic cytokine and therapeutic target. *Annu. Rev. Med.* 45:491, 1994.

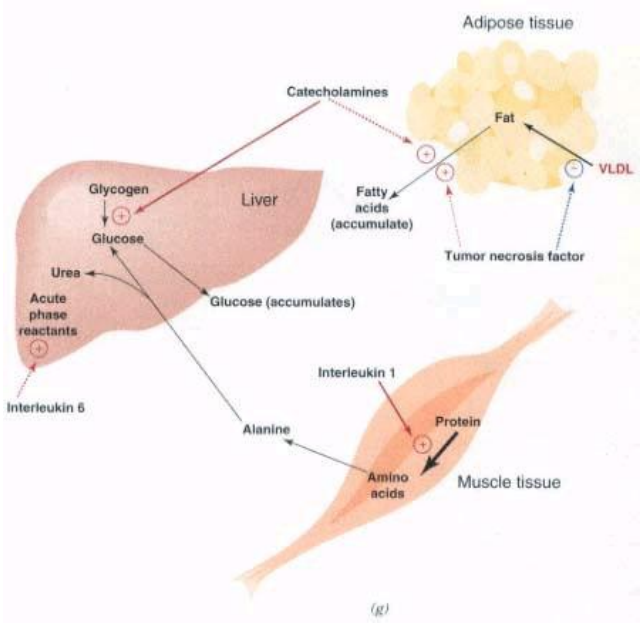


Figure 13.17
(continued)
(g) Stress.

Liver Disease Causes Major Metabolic Derangements

Since the liver is central to the body's metabolic interrelationships, advanced **liver disease** can be associated with major metabolic derangements (Figure 13.17*b*). The most important abnormalities are those in the metabolism of amino acids. The liver is the only organ capable of urea synthesis. In patients with **cirrhosis,** the liver is unable to convert ammonia into urea and glutamine rapidly enough, and the blood ammonia level rises. Part of this problem is due to abnormalities of blood flow in the cirrhotic liver, which interfere with the intercellular **glutamine cycle** (see p. 558). Ammonia arises from several enzyme reactions, such as glutaminase, glutamate dehydrogenase, and adenosine deaminase, during metabolism of amino acids by intestine and liver, and from intestinal lumen, where bacteria split urea into ammonia and carbon dioxide. Ammonia is very toxic to the central nervous system and is a major reason for the coma that sometimes occurs in patients in liver failure.

In advanced liver disease, aromatic amino acids accumulate in the blood to higher levels than branched-chain amino acids, apparently because of defective hepatic catabolism of the aromatic amino acids. This is important because aromatic amino acids and branched-chain amino acids are transported into the brain by the same carrier system. An elevated ratio of aromatic amino acids to branched-chain amino acids in liver disease results in increased brain uptake of aromatic amino acids. Increased synthesis of **neurotransmitters** such as **serotonin** in the brain as a consequence of increased availability of aromatic amino acids has been suggested to be responsible for some of the neurological abnormalities characteristic of liver disease. The liver is also a major source of insulin-like growth factor-I (IGF-I). Cirrhotics suffer muscle wasting because of deficient IGF-I synthesis in response to growth hormone. Finally, in outright

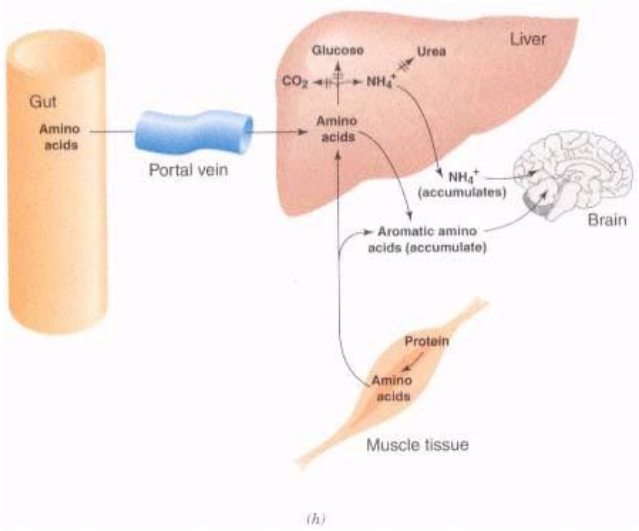


Figure 13.17
(continued)
(h) Liver disease.

liver failure, patients sometimes die of hypoglycemia because the liver is unable to maintain the blood glucose level by gluconeogenesis.

In Renal Disease Nitrogenous Wastes Accumulate

In chronic renal disease, there are many abnormalities of nitrogen metabolism. Levels of amino acids normally metabolized by kidney (glutamine, glycine, proline, and citrulline) increase. Nitrogen end products (e.g., urea, uric acid, and creatinine) also accumulate (Figure 13.17*i*). This accumulation is worsened by high dietary protein intake or accelerated proteolysis. The facts that gut bacteria can split urea into ammonia and that liver uses ammonia and α-keto acids to form nonessential amino acids have been used to control the level of nitrogenous wastes in renal patients. Patients are given a diet high in carbohydrate, and the amino acid intake is limited as much as possible to essential amino acids. Under these circumstances, the liver synthesizes nonessential amino acids from TCA cycle intermediates. This type of diet therapy may extend the time before the patient requires dialysis.

Oxidation of Ethanol in Liver Alters the NAD^+/NADH Ratio

The liver is primarily responsible for the first two steps of the **ethanol** catabolism:

$$\text{Ethanol } (CH_3CH_2OH) + NAD^+ \rightarrow \text{acetaldehyde } (CH_3CHO) + NADH + H^+$$
$$\text{Acetaldehyde } (CH_3CHO) + NAD^+ \rightarrow \text{acetate } (CH_3COO^-) + NADH + H^+$$

The first step, catalyzed by **alcohol dehydrogenases** in the cytosol, generates NADH; the second step, catalyzed by **aldehyde dehydrogenase,** also generates NADH but occurs largely in the mitochondrial matrix space. Liver disposes of NADH generated by these reactions by the only pathway it has available—the mitochondrial electron transport chain. Intake of even moderate amounts of

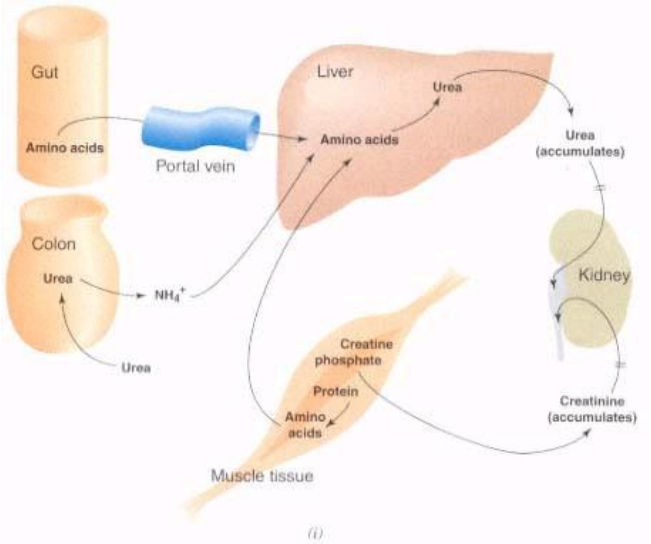


Figure 13.17
(continued)
(i) Kidney failure.

ethanol generates too much NADH. Many enzymes, for example, several involved in gluconeogenesis and fatty acid oxidation, are sensitive to product inhibition by NADH. Thus, during alcohol metabolism, these pathways are inhibited (Figure 13.17*j*), and fasting hypoglycemia and the accumulation of hepatic triacylglycerols (fatty liver) are consequences of alcohol ingestion. Lactate can accumulate as a consequence of inhibition of lactate gluconeogenesis and can result in metabolic acidosis.

Liver mitochondria have a limited capacity to oxidize acetate to CO_2, because the activation of acetate to acetyl CoA requires GTP, a product of the succinyl CoA synthetase reaction. The TCA cycle, and therefore GTP synthesis, are inhibited by high NADH levels during ethanol oxidation. Much of the acetate made from ethanol escapes the liver to the blood. Virtually every other cell with mitochondria can oxidize it to CO_2 by way of the TCA cycle.

Acetaldehyde, the intermediate in the formation of acetate from ethanol, can also escape from the liver. Acetaldehyde is a reactive compound that readily forms covalent bonds with functional groups of biologically important compounds. Formation of acetaldehyde adducts with proteins in tissues and blood of animals and humans drinking alcohol has been demonstrated. Such adducts may provide a marker for past drinking activity of an individual, just as hemoglobin A_{1c} has proved useful as an index of blood glucose control in diabetic patients.

In Acid-Base Regulation, Glutamine Plays a Pivotal Role

Regulation of acid–base balance, like that of nitrogen excretion, is shared by the liver and kidney (Figure 13.17*k*). Metabolism of proteins generates excess hydrogen ions. For example:

$$\text{Methionine or cysteine} \rightarrow SO_4^{2-} + 2\,H^+ + \text{other neutral products}$$

The kidney helps regulate blood pH by excreting hydrogen ions, which is necessary for the reabsorption of bicarbonate and the titration of phosphate

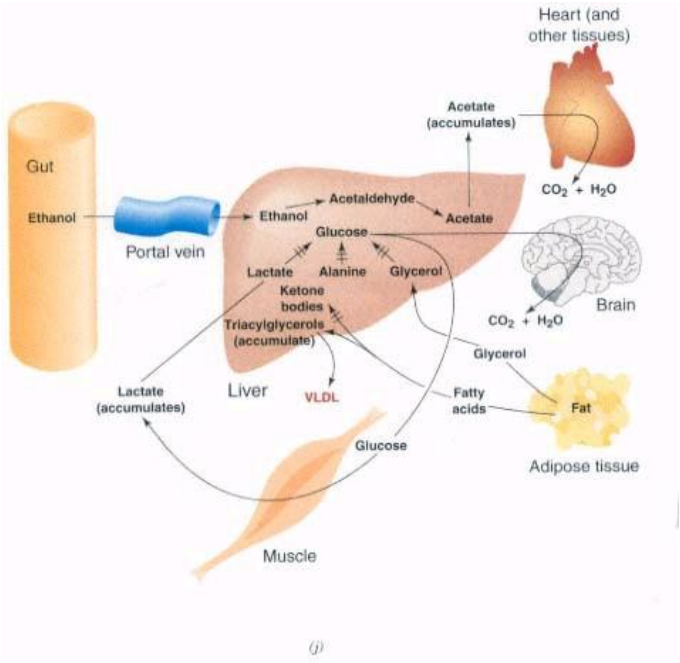


Figure 13.17
(continued)
(j) Ethanol ingestion.

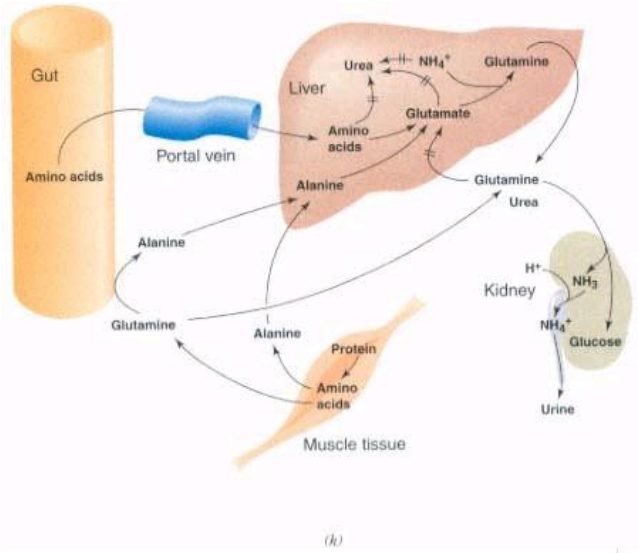


Figure 13.17
(continued)
(k) Acidosis.

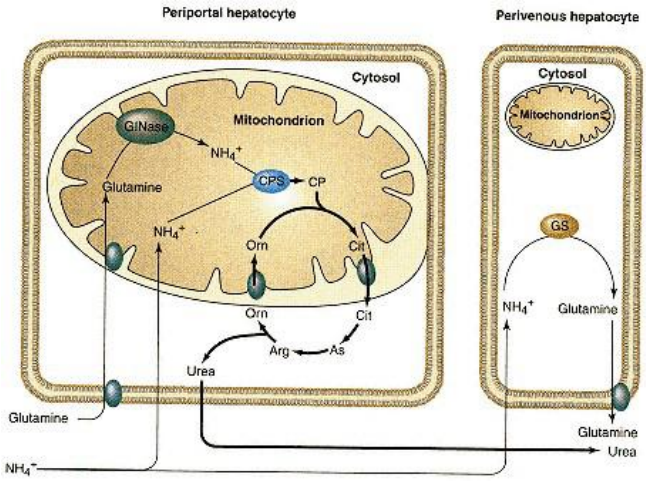


Figure 13.18
Intercellular glutamine cycle of the liver.
Abbreviations: GlNase, glutaminase; GS, glutamine synthetase; CPS, carbamoyl phosphate synthetase I; CP, carbamoyl phosphate; Cit, citrulline; AS, argininosuccinate; Arg, arginine; Orn, ornithine.
Redrawn from Häussinger, D. Glutamine metabolism in the liver: overview and current concepts. *Metabolism* 38(Suppl. 1):14, 1989.

and ammonia in the tubular filtrate (see Chapter 25, p. 1045). Glutamine is the precursor of renal ammonia production. In chronic metabolic **acidosis,** the activities of renal glutaminase, glutamate dehydrogenase, phosphoenolpyruvate carboxykinase, and the mitochondrial glutamine transporter increase and correlate with increased urinary excretion of ammonium ions and increased renal gluconeogenesis from amino acids. Liver participates by synthesizing less urea, which makes more glutamine available for the kidney. In **alkalosis,** urea synthesis increases in the liver, and gluconeogenesis and ammonium ion excretion by the kidney decrease.

An intercellular glutamine cycle enables the liver to play a central role in the regulation of blood pH. The liver is composed of two types of hepatocytes involved in glutamine metabolism: **periportal hepatocytes** near the hepatic arteriole and portal venule and **perivenous hepatocytes** located near the central venule (Figure 13.18). Blood enters the liver by the hepatic artery and

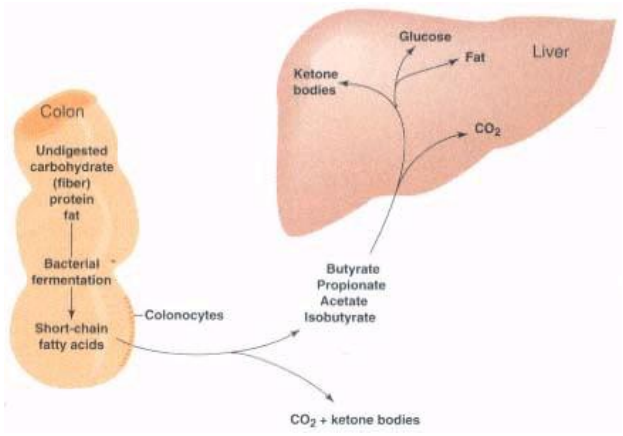


Figure 13.19
Bacterial fermentation generates fuel for colonocytes.

portal vein and leaves by way of the central vein. Glutaminase and urea cycle enzymes are concentrated in the periportal hepatocytes, while glutamine synthetase is found exclusively in perivenous hepatocytes (see p. 450). During alkalosis, glutamine enters the periportal cells and is hydrolyzed to contribute ammonium ion for urea synthesis. The bulk of glutamine and ammonium nitrogen entering the liver leaves the liver as urea. The perivenous cellular location of glutamine synthetase is important because some ammonium ions escape conversion to urea. This enzyme traps much of this toxic compound in the form of glutamine. Thus glutamine is released from the liver and circulates back to the liver where it reenters the glutamine cycle in the periportal hepatocytes. Thus, in liver, both donation of ammonium ion by glutamine for urea synthesis and the synthesis of glutamine are important in maintaining low blood ammonium levels. In acidosis, glutaminase of the periportal hepatocytes (unlike the renal glutaminase isozyme) is less active and much of the blood glutamine escapes hydrolysis in the liver. Likewise, carbamoyl phosphate synthetase of periportal hepatocytes is less active in acidosis, permitting perivenous cells to convert more ammonium ion to glutamine, which is then available for metabolism by the kidney to yield hydrogen ions that need to be eliminated in the urine.

The Colon Salvages Energy from the Diet

Unlike the small intestine, which uses glutamine for its major energy source, the **colon** utilizes short-chain fatty acids: **butyrate, propionate, isobutyrate,** and acetate (Figure 13.19). It obtains most of these fatty acids from the lumen of the colon, where bacteria produce them by fermentation of unabsorbed dietary components. These short-chain fatty acids would otherwise be lost in stool, so their use by cells of the colon (colonocytes) represents a way of gaining as much energy from dietary sources as possible. If produced in excess of the needs of the colon, short-chain fatty acids pass into the portal blood for use by the liver. Interestingly, colonocytes can produce ketone bodies from butyrate, presumably releasing them into the portal blood for use as fuel by extrahepatic tissues. When surgery is performed that bypasses the colon (e.g., an ileostomy), some patients develop a form of colitis called diversion **colitis**. In some cases, providing enemas containing the short-chain fatty acids has healed the colitis.

Bibliography

Amiel, S. A. Organ fuel selection: brain. *Proc. Nutr. Soc.* 54:151, 1995.

Brownlee, M., Vlassara, H., and Cerami, A. Nonenzymatic glycosylation and the pathogenesis of diabetes complications. *Ann. Intern. Med.* 101:527, 1984.

Cahill, G. F. Jr. Diabetes mellitus: a brief overview. *Johns Hopkins Med. J.* 143:155, 1978.

Calder, P. C. Fuel utilization by cells of the immune system. *Proc. Nutr. Soc.* 54:65, 1995.

Cerami, A., and Koenig, R. H. Hemoglobin A_{1c} as a model for the development of the sequelae of diabetes mellitus. *Trends Biochem. Sci.* 3:73, 1978.

Cohen, P., and Cohen, P. T. W. Protein phosphatases come of age. A review. *J. Biol. Chem.* 264:21435, 1989.

Crabb, D. W., and Lumeng, L. Metabolism of alcohol and the pathophysiology of alcoholic liver disease. In G. Gitmck (Ed.), *Principles and Practice of Gastroenterology and Hepatology. New York: Elsevier, 1988, p. 1163.*

Curthoys, N. P., and Watford, M. Regulation of glutaminase activity and glutamine metabolism. *Annu. Rev. Nutr.* 15:133, 1995.

DeLeve, L. D., and Kaplowitz, N. Glutathione metabolism and its role in hepatotoxicity. *Pharmacol. Ther.* 52:287, 1991.

Denton, R. M., and Pogson, C. I. *Metabolic Regulation*. New York: Wiley, 1976.

Duée, P.-H., Darcy-Vrillon, B., Blachier, F., and Morel, M.-T. Fuel selection in intestinal cells. *Proc. Nutr. Soc.* 54:83, 1995.

Elia, M. General integration and regulation of metabolism at the organ level. *Proc. Nutr. Soc.* 54:213, 1995.

Foster, D. W. Banting lecture 1984. From glycogen to ketones and back. *Diabetes* 33:1188, 1984.

Foster, D. W. Diabetes mellitus. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Vallee. (Eds), *The Metabolic Basis of Inherited Disease*. New York:McGraw-Hill, 1989.

Frayn, K. N., Humphreys, S. M., and Coppack, S. W. Fuel selection in white adipose tissue. *Proc. Nutr. Soc.* 54:177, 1995.

Fulop, M. Alcoholism, ketoacidosis, and lactic acidosis. *Diabetes Metab. Rev.* 5:365, 1989.

Geelen, M. J. H., Harris, R. A., Beynen, A. C., and McCune, S. A. Short-term hormonal control of hepatic lipogenesis. *Diabetes* 29:1006, 1980.

Goldberg, A. L., Baracos, V., Rodemann, P., Waxman, L., and Dinarello, C. Control of protein degradation in muscle by prostaglandins, calcium, and leukocytic pyrogen (interleukin 1). *Fed. Proc.* 43:1301, 1984.

Harris, R. A., Popov, K. M., and Zhao, Y. Nutritional control of the protein kinases responsible for the phosphorylation of the α-ketoacid dehydrogenase complexes. *J. Nutr.* 125:1758S, 1995.

Haussinger, D. Glutamine metabolism in the liver: overview and current concepts. *Metabolism* 38(Suppl. 1):14, 1989.

Henriksson, J. Muscle fuel selection: effect of exercise and training. *Proc. Nutr. Soc.* 54:125, 1995.

Hers, H. G., and Hue, L. Gluconeogenesis and related aspects of glycolysis. *Annu. Rev. Biochem.* 52:617, 1983.

Ingebritsen, T. S., and Cohen, P. Protein phosphatases: properties and role in cellular regulation. *Science* 221:331, 1983.

Katz, J., and McGarry, J. D. The glucose paradox. Is glucose a substrate for liver metabolism? *J. Clin. Invest.* 74:1901, 1984.

Krebs, H. A. Some aspects of the regulation of fuel supply in omnivorous animals. *Adv. Enzyme Regul.* 10:387, 1972.

Krebs, H. A., Williamson, D. H., Bates, M. W., Page, M. A., and Hawkins, R. A. The role of ketone bodies in caloric homeostasis. *Adv. Enzyme Regul.* 9:387, 1971.

Kurkland, I. J., and Pilkis, S. J. Indirect and direct routes of hepatic glycogen synthesis. *FASEB J.* 3:2277, 1989.

Lamer, J. *Intermediary Metabolism and Its Regulation*. Englewood Cliffs, NJ: Prentice Hall, 1971.

MacDonald, I. A., and Webber, J. Feeding, fasting and starvation: factors affecting fuel utilization. *Proc. Nutr. Soc.* 54:267, 1995.

McGarry, J. D., Woeltje, K. F., Kuwajima, M., and Foster, D. W. Regulation of ketogenesis and the renaissance of carnitine palmitoyltransferase. *Diabetes Metab. Rev.* 5:271, 1989.

Newsholme, E. A., and Leech, A. R. *Biochemistry for the Medical Sciences*. New York: Wiley, 1983.

Newsholme, E. A., and Start, C. *Regulation in Metabolism*. New York: Wiley, 1973.

Nosadini, R., Avogaro, A., Doria, A., Fioretto, P., Trevisan, R., and Morocutti, A. Ketone body metabolism: a physiological and clinical overview. *Diabetes Metab. Rev.* 5:299, 1989.

Pedersen, O. The impact of obesity on the pathogenesis of non-insulin-dependent diabetes mellitus: a review of current hypotheses. *Diabetes Metab. Rev.* 5:495, 1989.

Pedersen, O., and Beck-Nielsen, H. Insulin resistance and insulin-dependent diabetes mellitus. *Diabetes Care* 10:516, 1987.

Pilkis, S. J., Ed-Maghrabi, M. R., and Claus, T. H. Hormonal regulation of hepatic gluconeogenesis and glycolysis. *Annu. Rev. Biochem.* 57:755, 1988.

Randle, P. J. Metabolic fuel selection: general integration at the whole-body level. *Proc. Nutr. Soc.* 54:317, 1995.

Shulman, G. I., and Landau, B. R. Pathways of glycogen repletion. *Physiol. Rev.* 72:1019, 1992.

Sugden, M. C., Holness, M. J., and Palmer, T. N. Fuel selection and carbon flux during the starved-to-fed transition. A review article. *Biochem. J.* 263:313, 1989.

Toth, B., Bollen, M., and Stalmans, W. Acute regulation of hepatic phosphatases by glucagon, insulin, and glucose. *J. Biol. Chem.* 263:14061, 1988.

Ulwin, P. The state of world hunger. *Nutr. Rev.* 52:151, 1994.

Williamson, D. H., and Lund, P. Substrate selection and oxygen uptake by the lactating mammary gland. *Proc. Nutr. Soc.* 54:165, 1995.

Questions

C. N. Angstadt and J. Baggott

Refer to the following for Questions 1 and 2

A. well-fed state

B. early fasting state

C. fasting state

D. early refed state

1. Hepatic glycogenolysis is a primary source of blood glucose during this period.

2. Ketone bodies supply a significant portion of the brain's fuel.

3. The fact that the K_m of aminotransferases for amino acids is much higher than that of aminoacyl-tRNA synthetases means that:

   A. at low amino acid concentrations, protein synthesis will take precedence over amino acid catabolism.

   B. the liver cannot accumulate amino acids.

   C. amino acids will undergo transamination as rapidly as they are delivered to the liver.

   D. any amino acids in excess of immediate needs for energy must be converted to protein.

   E. amino acids can be catabolized only if they are present in the diet.

4. Branched-chain amino acids:

   A. are normally completely catabolized by muscle to CO_2 and H_2O.

   B. can be catabolized by liver but not muscle.

   C. are the main dietary amino acids metabolized by intestine.

   D. are in high concentration in blood following the breakdown of muscle protein.

   E. are a major source of nitrogen for alanine and glutamine produced in muscle.

5. In the early refed state:

   A. the fatty acid concentration of blood rises.

   B. liver no longer carries out gluconeogenesis.

   C. liver replenishes its glycogen by synthesis of glucose-6-phosphate from lactate.

   D. glucose being fed is converted directly to glycogen by the liver.

   E. amino acids cannot be used.

6. All of the following statements about interorgan interactions are correct EXCEPT:

   A. ornithine for the urea cycle is synthesized from glutamate in the kidney.

   B. citrulline leads to the formation of arginine in both liver and kidney.

   C. kidney uses arginine in the synthesis of creatine for use by muscle.

   D. arginine synthesized by the kidney is the source of nitric oxide for many cells.

   E. creatinine cleared by the kidney is generated from creatine phosphate in muscle.

7. Carnitine:

   A. is formed in all cells for their own use.

   B. is synthesized from free lysine.

   C. formation requires that lysyl residues in protein be methylated by *S*-adenosylmethionine.

   D. formation is inhibited by vitamin C.

   E. is cleaved to γ-butyrobetaine.

8. The largest energy reserve (in terms of kilocalories) in humans is:

   A. blood glucose.

   B. liver glycogen.

   C. muscle glycogen.

   D. adipose tissue triacylglycerol.

   E. muscle protein.

9. All of the following represent control of a metabolic process by substrate availability EXCEPT:

A. increased urea synthesis after a high-protein meal.

B. rate of ketogenesis.

C. hypoglycemia of advanced starvation.

D. response of glycolysis to fructose 2,6-bisphosphate.

E. sorbitol synthesis.

10. Which of the following would favor gluconeogenesis in the fasted state?

A. fructose 1,6-bisphosphate stimulation of pyruvate kinase

B. acetyl CoA activation of pyruvate carboxylase

C. citrate activation of acetyl-CoA carboxylase

D. malonyl CoA inhibition of carnitine palmitoyltransferase I

E. fructose 2,6-bisphosphate stimulation of 6-phosphofructo-1-kinase

11. Conversion of a nonphosphorylated enzyme to a phosphorylated one:

A. always activates the enzyme.

B. is always catalyzed by a cAMP-dependent protein kinase.

C. is signaled in the liver by insulin.

D. is more likely to occur in the fasted than in the well-fed.

E. usually occurs at threonine residues of the protein.

12. Adipose tissue responds to low insulin/glucagon ratio by:

A. dephosphorylating the interconvertible enzymes.

B. stimulating the deposition of fat.

C. increasing the amount of pyruvate kinase.

D. stimulating hormone-sensitive lipase.

E. stimulating phenylalanine hydroxylase.

13. Changing the level of enzyme activity by changing the number of enzyme molecules:

A. is considerably slower than allosteric or covalent modification methods.

B. may involve enzyme induction.

C. may override the effectiveness of allosteric control.

D. may be caused by hormonal influences or by changing the nutritional state.

E. all of the above are correct.

14. Muscle metabolism during exercise:

A. is the same in both aerobic and anaerobic exercise.

B. shifts from primarily glucose to primarily fatty acids as fuel during aerobic exercise.

C. uses largely glycogen and phosphocreatine in the aerobic state.

D. causes a sharp rise in blood ketone body concentration.

E. uses only phosphocreatine in the anaerobic state.

15. In noninsulin-dependent diabetes mellitus:

A. hypertriglyceridemia does not occur.

B. ketoacidosis in the untreated state is always present.

C. results because the β cells of the pancreas can no longer make insulin.

D. may be accompanied by high levels of insulin in the blood.

E. results in severe weight loss.

16. The elevated liver concentration of NADH produced by ingestion of ethanol:

A. is restricted to the mitochondria.

B. may lead to an acidosis by inhibiting gluconeogenesis from lactate.

C. leads to "fatty liver" by stimulating fatty acid synthesis.

D. increases the conversion of acetate to acetyl CoA.

E. arises solely from the conversion of ethanol to acetaldehyde.

17. Glutaminase:

A. in renal cells is unaffected by blood pH.

B. in liver is confined to perivenous hepatocytes.

C. activity is low in liver during alkalosis.

D. activity is more active in both liver and kidney in acidosis.

E. in renal cells increases in acidosis.

Answers

1. B The response of glycogenolysis to fasting is rapid, and during this period there is still glycogen present. In fasting, the glycogen is depleted and in the other two states, glycogenesis would occur (p. 529).

2. C If ketone body concentration in blood is high, ketone bodies can cross the blood–brain barrier and they are a good fuel. High ketone body concentrations do not occur in the other states (p. 532).

3. A A high K_m means that a reaction will proceed slowly at low concentration, whereas a low K_m means the reaction can be rapid under the same circumstances. Protein synthesis requires only that all amino acids be present. Unless amino acids are in high enough concentration, the liver does not catabolize them (p. 529).

4. E A and B: Muscle has high levels of the aminotransferases for branched-chain amino acids, whereas liver has high levels of enzymes for the catabolism of the branched-chain α-keto acids. C: Intestine metabolizes several dietary amino acids but not these. D and E: When branched-chain amino acids are derived from muscle protein, transamination transfers the nitrogen to alanine or glutamine, which are transported to the liver and kidney (p. 531).

5. C This is the indirect pathway. A, B, and D: Fat metabolism is normal but glucose metabolism is not normal yet. E: Amino acids are also used for gluconeogenesis (p. 534).

6. A This is a pathway of intestinal epithelium (p. 534). B: This is part of the urea cycle in liver but arginine from kidney is an important source of arginine for liver. C: The reaction requires *S*-adenosylmethionine. E: Creatinine is thus a measure of both muscle mass and renal function (pp. 535–536).

7. C These trimethyllysines are released when protein is hydrolyzed. A: Only liver and kidney have the complete synthetic pathway. D: There are two hydroxylations that require this vitamin. E: This is a precursor (p. 536).

8. D The caloric content of adipose tissue fat is more than five times as great as that of muscle protein and almost 200 times as great as that of the combined carbohydrates (Table 13.1). A: Blood glucose must be maintained but is a relatively minor reserve. B and C: Glycogen is a rapidly mobilizable reserve of energy but not a large one. E: Protein can be used for energy, but that is not its primary role.

9. D Fructose 2,6-bisphosphate is an allosteric effector (activates the kinase and inhibits the phosphatase) of the enzyme controlling glycolysis. A: After a high-protein meal, the intestine produces ammonia and precursors of ornithine for urea synthesis. B: Ketogenesis is dependent on the availability of fatty acids. C: This represents lack of gluconeogenic substrates. E: This leads to complications in diabetes (p. 540).

10. B Pyruvate carboxylase is a key gluconeogenic enzyme. A and E: Stimulation of these enzymes stimulates glycolysis, opposing gluconeogenesis. C and D: Malonyl CoA inhibits transport of fatty acids into mitochondria for β-oxidation, a necessary source of energy for gluconeogenesis (p. 540).

11. D In the well-fed state, insulin/glucagon ratio is high and cAMP levels are low. A: Some enzymes are active when phosphorylated; for others the reverse is true. B: This is the most common, though not only, mechanism of phosphorylation. C: Insulin does not signal the phosphorylation of the enzymes involved. E: The most common site for phosphorylation is serine (pp. 541–544).

12. D A: Low insulin/glucagon ratio means high cAMP and, thus, high activity of cAMP-dependent protein kinase and protein phosphorylation. B and D: Phosphorylation activates hormone-sensitive lipase to mobilize fat. C: cAMP works by stimulating covalent modification of enzymes. E: This is a liver enzyme (pp. 544–545).

13. E A: Adaptive changes are examples of long-term control. B and D: Both hormonal and nutritional effects are involved in inducing certain enzymes and/or altering their rate of degradation. C: If there is little or no enzyme because of adaptive changes, allosteric control is irrelevant. This is important to keep in mind in refeeding a starved person (pp. 545–546).

14. B This is indicated by the drop in the respiratory quotient. A: Anaerobically exercised muscle uses glucose almost exclusively; aerobically exercised muscle uses fatty acids and ketone bodies. D: Ketone bodies are good aerobic substrates so the blood concentration does not increase greatly. E: Phosphocreatine is only a short-term source of ATP (p. 551).

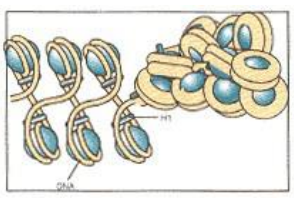
15. D A: Hypertriglyceridemia is characteristic. B: Ketoacidosis is common only in the insulin-dependent type. C and D: The problem is insulin resistance, not failure to produce insulin. E: Most patients are obese because adipocytes remain sensitive to insulin (pp. 548 and 550).

16. B Failure to oxidize lactate to pyruvate because of the unfavorable NAD^+/NADH ratio leads to lactate accumulation. A and E: The oxidation of ethanol, which also produces NADH, is cytosolic. Acetaldehyde oxidation is mitochondrial. C: "Fatty liver" is a consequence of inhibition of fatty acid oxidation by high NADH. D: Acetate activation requires GTP from the TCA cycle. Why is this cycle inhibited (pp. 555–556)?

17. E A and E: Glutamine in kidney is the primary source of ammonia for excretion of protons. B: This is the site of glutamine synthesis. C: This is when urea is formed. D: The liver activity of liver glutaminase during acidosis permits glutamine to escape liver for the kidney's use (pp. 556–559).

Chapter 14—
DNA I:
Structure and Conformation

Stelios Aktipis

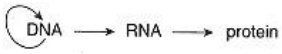


14.1— Overview

One of the striking aspects of natural order is the sense of unity that exists between members of successive generations in each species. An almost totally stable bank of information must always be preserved and passed from one generation to the next if individual species are to maintain their identities relatively unchanged over millions of years. It is well established that this bank of genetic information takes the form of a macromolecule, deoxyribonucleic acid (DNA), which serves as the carrier of genetic information in both prokaryotes and eukaryotes. DNA exhibits a rare purity of function by being one of the few macromolecules known to perform, with only minor exceptions, the same basic functions across species barriers.

Properties of cells are to a large extent determined by their constituent proteins. Many proteins are indispensable structural components of cells. Other proteins, such as enzymes and certain hormones, are functional and determine many of the biochemical properties of the cell. As a result, factors that control which proteins a cell may synthesize, at what quantities, and with which sequence are the same factors that primarily determine function and destiny of every living cell.

DNA is the macromolecule that ultimately controls every aspect of cellular function, primarily through protein synthesis. DNA exercises this control as suggested by the sequence



Flow of biological information is clearly from one class of nucleic acid to another, from DNA to RNA, with only minor exceptions, and from there to protein. For this transfer of information to occur faithfully, each preceding macromolecule serves as a structure-specifying template for the synthesis of the subsequent member in the sequence.

In addition to regulating cellular expression, DNA plays an exclusive role in heredity. This role is suggested by a circular arrow engulfing DNA, which depicts DNA as a replicon, a molecule that can undergo self-replication. Replication permits DNA to make copies of itself as a cell divides and bestows them to daughter cells, which can thus inherit every property and characteristic of the original cell. Thus DNA ultimately determines the properties of a living cell by regulating expression of biological information, primarily by control of protein synthesis, and transfers biological information from one generation to the next; that is, it transmits genetic information.

DNA Can Transform Cells

These universally accepted principles were rejected outright not long ago. In fact, prior to the 1950s the general view was that nucleic acids are substances of somewhat limited cellular importance. The first convincing suggestion that DNA is the genetic material was made during the mid-1940s. The experiment involved transformation of a type of pneumococcus, surrounded by a polysaccharide capsule and referred to as S form because of its property of forming colonies with smooth-looking cellular perimeters, to a mutant without capsule, called R form, which forms colonies with rough-looking outlines. These two forms are genetically distinct and cannot interconvert spontaneously. **Transformation** experiments demonstrated that an extract of pure DNA from S form, when incorporated into R form of pneumococcus, conveyed to R form the specific property of synthesizing the characteristic polysaccharide capsule. Furthermore, bacteria transformed from R form to S form maintained the property

of synthesizing capsule over succeeding generations. It was thus demonstrated that DNA was the **transforming agent,** as well as the material responsible for transmitting genetic information from one generation to the next. Almost three-quarters of a century elapsed from the time nucleic acids were discovered until their important biological role was generally recognized. Clinical Correlation 14.1 describes current studies in transforming mammalian cells with DNA.

DNA's Information Capacity Is Enormous

A striking characteristic of DNA is its ability to encode an enormous quantity of biological information. An undifferentiated mammalian fetal cell contains only a few picograms (10^{-12} g) of DNA. Yet this minute amount of material is sufficient to direct synthesis of as many as 100,000 distinct proteins that will determine the form and biochemical behavior of a large variety of differentiated tissues in adult animals.

The compactness of information storage in DNA is unique. Even sophisticated memory elements of contemporary computers appear pitifully inadequate by comparison. How does DNA achieve such a supreme **coding effectiveness?** Answers must obviously be sought in the nature of its chemical structure. It turns out that this structure is not only consistent with the unique efficiency of DNA as a "memory bank" but also provides the basis for understanding how DNA eventually "translates" this information into proteins.

CLINICAL CORRELATION 14.1

DNA Vaccines

Traditional procedures of vaccination have used purified components of an infectious organism, dead or attenuated intact cells or viruses, to provide individuals with active immunity by eliciting production of specific antibodies. Many have been successful in providing protection against diseases such as polio, smallpox, whooping cough, typhoid fever, and diphtheria.

A prototype DNA vaccine has been developed. It consists of a naked DNA that encodes the nucleoprotein of the influenza virus. This gene is the same or very similar in many strains of this virus and should afford protection against all or most of them. Naked DNA, that is, DNA freed of all its naturally associated proteins, is used. It enters cells and can be expressed without need of a complex virus system. Results of its use in mice and nonhuman primates have been very encouraging. Naked-DNA vaccines appear to stimulate cell-mediated immunity and an antibody response.

McDonnell, W. M., and Askari, F. K. DNA vaccines. *N. Engl. J. Med.* 334:42, 1996.

14.2— Structure of DNA

DNA is a polynucleotide produced by polymerization of deoxyribonucleotides. The structure of nucleotides and their constituent purine and pyrimidine bases are presented in Chapter 12.

The **base composition of DNA** varies considerably among species, particularly prokaryotes, which have a range of 25–75% in adenine–thymine content. This range narrows with evolution, reaching limiting values of about 45–53% in mammals.

DNA contains various **methylated bases.** These methylated derivatives are present in all prokaryotic DNA molecules examined to date but are absent in certain eukaryotes such as yeast and insects. As a rule these bases are generated by action of methylases, Dam and Dem, following synthesis of DNA. Methyl groups are transferred from *S*-adenosylmethionine. **Dam methylase** selects adenine residues on GATC sequences for methylation. **Dem methylase** acts on cytosine residues on opposite strands in the sequence

```
C C A G G
⋮ ⋮ ⋮ ⋮ ⋮
G G T C C
```

Such methylated sites are recognized by proteins involved in DNA functions such as recombination and initiation of DNA synthesis.

A base may be methylated prior to incorporation into DNA, as in transformation of cytosine to **5-hydroxycytosine. Glycosylated 5-hydroxycytosine** is found as a constituent of T-even phages of *Escherichia coli*. Other unusual base changes include the presence of uracil, a constituent of RNA, in certain *Bacillus subtilis* phages, instead of thymine. Structures of some of these bases are shown in Figure 14.1.

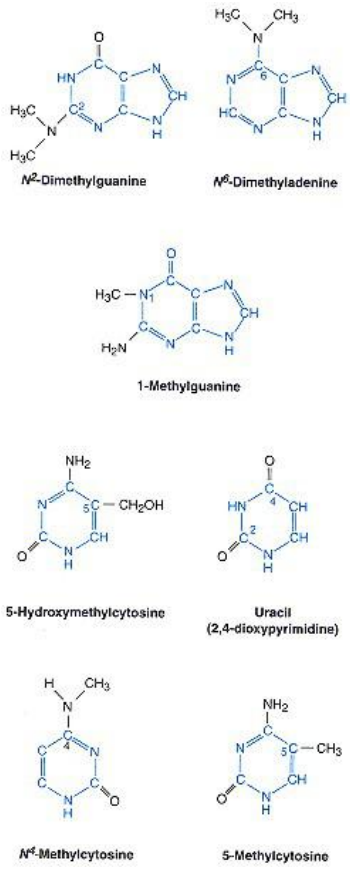


Figure 14.1
Structures of some less common bases occurring in DNA.

Nucleotides Joined by Phosphodiester Bonds Form Polynucleotides

Polynucleotides are formed by joining of nucleotides by **phosphodiester bonds.** The phosphodiester bond is the formal analog of the peptide bond in proteins. It joins by esterification of two of the three OH groups of phosphoric acid, two adjoining nucleotide residues. Deoxyribose contains two free OH groups on the C-3 and C-5 atoms that can participate in formation of a phosphodiester bond. Indeed, the nucleotide residues in DNA are joined by **3′,5′-phosphodiester bonds,** as shown in Figure 14.2.

Many polynucleotides are linear polymers. The last nucleotide residues at opposite ends of the polynucleotide chain serve as the two terminals of the chain. It is apparent that these terminals are not structurally equivalent, since one of the nucleotides must terminate at a 3 -OH group and the other at a 5 -OH group. These ends of the polynucleotides are referred to as 3 and 5 termini, and they may be viewed as corresponding to the amino and carboxyl termini in proteins. Polynucleotides also exist as cyclic structures, which contain

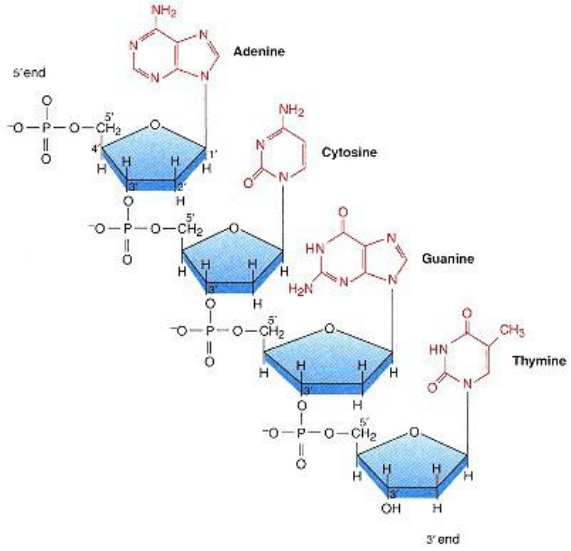


Figure 14.2
Structure of a DNA polynucleotide segment.
Shown is a tetranucleotide. Generally, polymers containing less than 30–40 nucleotides are referred to as oligonucleotides.

no free terminals. Esterification between the 3 -OH terminus of a polynucleotide with its own 5 -phosphate terminus can produce a cyclic polynucleotide.

Long polymers of nucleotides joined by phosphodiester bonds are called polynucleotides. **Oligonucleotides** are shorter nucleotide-containing polymers. According to formal rules of nomenclature, however, polynucleotides are named by using roots derived from the names of corresponding nucleotides, and using the ending ylyl. Polynucleotide sequences are always read in the 5 3 direction, unless specified otherwise. For example, the polynucleotide segment in Figure 14.2, in which the 5 terminal is on the left of each nucleotide residue, should be named from left to right as

...deoxyadenylyl, deoxycytidylyl, deoxyguanylyl, deoxythymidylyl...

However, use of complete chemical names is cumbersome and abbreviations are generally preferred. For example, the oligonucleotide shown in Figure 14.2 is usually referred to as dAdCdGdT, and a polynucleotide containing only one kind of nucleotide, for example, dA, may be written as poly(dA). Oligo- and polynucleotide structures are also written out in shorthand, as shown in Figure 14.3.

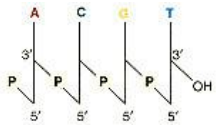


Figure 14.3
Shorthand form for structure of oligonucleotides.
The convention used in writing the structure of an oligo- or polynucleotide is a perpendicular bar representing the deoxyribose moiety, with the 5 -OH position of the sugar located at the bottom of the bar and the 3 -OH at a midway position. Bars joining the 3 and 5 positions represent the 3 ,5 -phosphodiester bond, and the P on the left side of the perpendicular bar represents a 5 -phosphate ester. A 3 -phosphate ester is represented by placing the phosphate group on the right side of the bar and the base by its initial.

The specific sequence of bases along a polynucleotide chain determines its biological properties. Although the structure of the nucleic acid bases had been known for many years, the polymeric structure initially proposed for DNA was one of the classical errors in the history of biochemistry. Experimental data obtained from partially degraded samples of DNA, and several misconceptions, led to the erroneous conclusion that DNA consisted of repeating tetranucleotide units. Each tetranucleotide supposedly contained equimolar quantities of the four common bases. These impressions persisted to some degree until the late 1940s and early 1950s, when they were clearly shown to be in error. In the interim, however, these misconceptions were responsible for setting back acceptance of the concept that DNA of chromosomes carried genetic information. The monotonous structure of repeating tetranucleotides appeared to lack the

versatility to encode for the enormous number of messages necessary to convey hereditary traits. Instead proteins, which can be ordered in an almost unlimited number of amino acid sequences, were favored as the most suitable candidates for a hereditary function. Transformation experiments carried out in the mid-1940s, and the finding that DNA consists of polynucleotide and not tetranucleotide chains, were responsible for general acceptance of the hereditary role of DNA that followed.

Nucleases Hydrolyze Phosphodiester Bonds

The nature of the linkage between nucleotides to form polynucleotides was elucidated primarily by use of exonucleases, enzymes that hydrolyze these polymers in a selective manner. Exonucleases cleave the last nucleotide residue at either of the two terminals of an oligonucleotide. Oligonucleotides can thus be degraded by stepwise removal of individual nucleotides or small oligonucleotides from either the 5 or 3 terminus. **Nucleases** sever bonds in one of two nonequivalent positions indicated in Figure 14.4 as proximal (p) or distal (d) to the base, which occupies the 3 end of the bond. For example, treatment of an oligodeoxyribonucleotide with snake venom diesterase, an enzyme obtained from snake venom, yields deoxyribonucleoside 5 -phosphates. In contrast, treatment with a diesterase isolated from animal spleen produces deoxyribonucleoside 3 -phosphates.

Other nucleases that cleave phosphodiester bonds located in the interior of polynucleotides are designated as endonucleases and behave similarly. For instance, DNase I cleaves only p linkages, while DNase II cleaves d linkages. Points of cleavage along an oligonucleotide chain are indicated by arrows in Figure 14.4. Some endonucleases have been particularly useful in development of methodologies for sequencing of DNA polynucleotides and have provided the basis for development of recombinant DNA techniques.

Many nucleases do not exhibit any specificity with respect to the base adjacent to the linkage that is hydrolyzed. Others, however, act very discriminately only next to specific types of bases or even specific bases. **Restriction endonucleases** act only on sequences of bases specifically recognized by each restriction enzyme. Nucleases also exhibit specificities with respect to overall structure of polynucleotides. For instance, some nucleases act on both single- or double-stranded polynucleotides, whereas others discriminate between these two structures. In addition, some nucleases exclusively designated as **phosphodiesterases** will act on either DNA or RNA, whereas other nucleases will limit their activity to only one type of polynucleotide. Nucleases listed in Table 14.1 illustrate some of the properties of these enzymes.

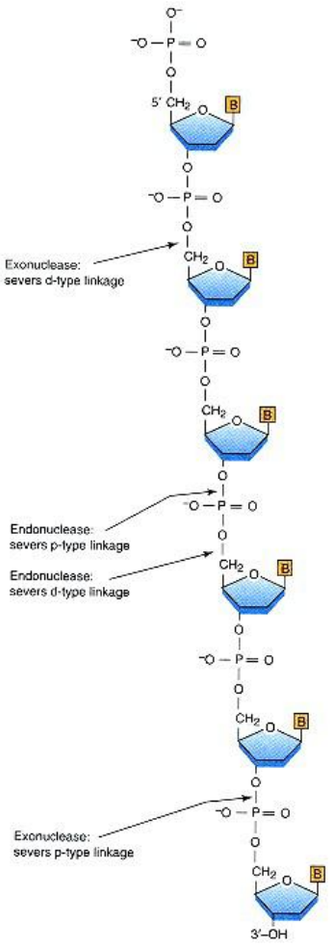


Figure 14.4
Specificities of nucleases.
Exonucleases remove nucleotide residues from either terminal of a polynucleotide, depending on their specificity. Endonucleases hydrolyze interior phosphodiester bonds. Both endo- and exonucleases hydrolyze either d- or p-type linkages (see text for explanation of d- and p-type linkages).

Periodicity Leads to Secondary Structure of DNA

Polypeptide chains of protein are often arranged in space so as to form periodic structures. For instance, in the α-helix each residue is related to the next by a translation of 1.5 A along the helix axis and a rotation of 100°. This places 3.6 amino acid residues in each complete turn of the polypeptide helix. The property of **periodicity** is also encountered with polynucleotides, which usually occur in the form of **helices.** Such preponderance of helical conformations among macromolecules is not surprising. Formation of helices tends to accommodate effects of intramolecular forces, which in a helix can be distributed at regular intervals. The alternative, that is, a hypothetical extended linear conformation, would place successive base pairs at 0.68 nm apart and allow water molecules to be inserted between hydrophobic base pairs. Clearly such an arrangement would be thermodynamically unfavorable. The precise geometry of the polynucleotide helices varies, but the helical structure invariably results from stacking

TABLE 14.1 Specificities of Various Types of Nucleases

Enzyme	*Substrate*	*Specificity*[a]
EXONUCLEASES		
Snake venom phosphodiesterase	DNA or RNA single-stranded only	Cleaves all p-type linkages, starting with a free 3 -OH group and moving toward the 5 terminal; releases nucleoside 5 -phosphates; has no base specificity
Bovine spleen phosphodiesterase	DNA or RNA single-stranded only	Cleaves all d-type linkages, starting at the free 5 -OH and proceeding to the 3 terminal; releases nucleoside 3 -phosphates; has no base specificity
ENDONUCLEASES		
Bovine pancreas deoxyribonuclease (DNase I)	DNA single- or double-stranded	Cleaves all p-type linkages but prefers those between purine and pyrimidine bases
Calf thymus deoxyribonuclease (DNase II)	DNA single- or double-stranded	Cleaves all d-type linkages randomly

[a] See text for explanation of d- and p-type linkages.

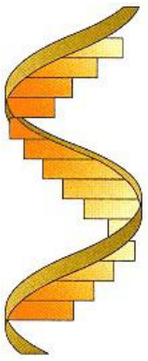


Figure 14.5 Conformation of a hypothetical, perfectly helical, single-stranded polynucleotide. The helical band represents the phosphate backbone of the polynucleotide. Bases are shown in a side view as solid blocks in tight contact with their neighbors, above and below each base. Surfaces of the rings are in contact with each other and are not visible in the perspective.

of bases along the helix axis. In many instances stacking produces helices in which bases are more or less perpendicularly oriented along the helix and touch one another. This arrangement leaves no free space between two successive neighboring bases (Figure 14.5). Such stacked single-stranded helices, however, are not commonly encountered in cells. Rather, polynucleotide helices tend to associate with one another to form double helices.

Forces That Determine Polynucleotide Conformation

The hydrophobic properties of the bases are, to a large extent, responsible for forcing polynucleotides to adopt helical conformations. Molecular models of bases reveal that the edges of the rings contain polar groups (i.e., amino and OH groups) that interact with other polar groups or surrounding water molecules. The faces of the rings, however, are unable to participate in such interactions and tend to avoid any contact with water. Instead they tend to interact with one another, producing a **stacked conformation.** The stability of this arrangement is further reinforced by an interchange between electrons that circulate in π orbitals located above and below the plane of each ring.

Clearly then, single-stranded polynucleotide helices are stabilized by **hydrophobic** and **dipole-induced dipole interactions** involving the π orbitals of bases, which collectively produce base stacking. The stability of helical structures is somewhat decreased by potential repulsion among charged phosphate residues of the polynucleotide backbone. These repulsive forces introduce a certain degree of rigidity to the structure of polynucleotides. Under physiological conditions, that is, at neutral pH and relatively high concentrations of salts, the charges on the phosphate residues are partially shielded by the cations present, such as Mg^{2+}, and the structure can be viewed as a fairly flexible coil. Under more extreme conditions stacking of bases is disrupted and the helix collapses. A collapsed helix is commonly described as a **random coil.** Conversion between a stacked helix and an unstacked conformation is depicted in Figure 14.6.

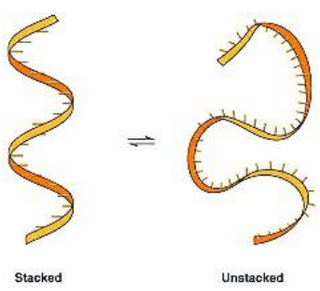


Figure 14.6 Stacked and unstacked conformations of a polynucleotide. Stacking of bases decreases flexibility of a polynucleotide and tends to produce a more extended, often helical, structure.

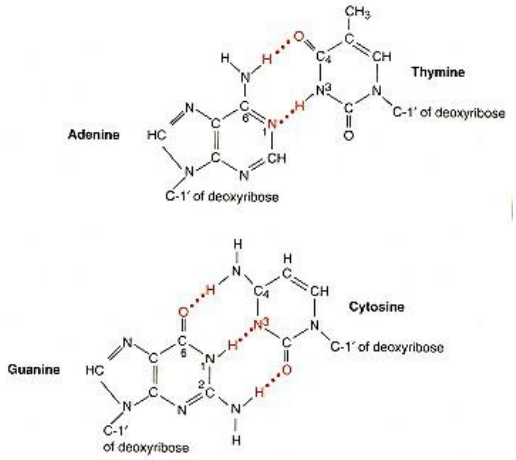


Figure 14.7
Formation of hydrogen bonds between complementary bases in double-stranded DNA.
Interaction between polynucleotide strands is a highly selective process. Complementarity depends not only on the geometric factors that allow the proper fitting between the complementary bases of the two strands, but also on the electronic specificity of interaction between complementary bases. Thus specificity of interaction between purines and pyrimidines has also been noted both in solution and in the crystal form, and it is expressed in terms of strong hydrogen bonding between monomers of adenine and uracil or monomers of guanine and cytosine.

DNA Double Helix

Although some forms of cellular DNA exist as **single-stranded structures,** the most widespread DNA structure is the **double helix.** The double helix can be visualized as resulting from interwinding of two right-handed helical polynucleotide strands around a common axis. The two strands achieve contact through hydrogen bonds, which are formed at the hydrophilic edges of their bases. These bonds extend between purine residues in one strand and pyrimidine residues in the other, so that the two types of resulting pairs are always adenine–thymine and guanine–cytosine. A direct consequence of these hydrogen-bonding specificities is that double-stranded DNA contains equal amounts of purines and pyrimidines. Examination of space-filling models clearly indicates structural compatibility of these bases in forming linear hydrogen bonds.

This relationship between bases in the double helix is described as **complementarity.** Bases are complementary because every base of one strand is matched by a complementary hydrogen-bonding base on the other strand. For instance, for each adenine projecting toward the common axis of the double helix, a thymine must be projected from the opposite chain so as to fill exactly the space between strands by hydrogen bonding with adenine. Neither cytosine nor guanine fits precisely in the available space across from adenine in a manner that allows formation of hydrogen bonds across strands. These **hydrogen-bonding specificities** (Figure 14.7) ensure that the entire base sequence of one strand is complementary to that of the other strand.

The double helix exists in various geometries designated as **DNA A, B, and C.** Formation of these different **conformations** depends on the base composition of DNA and on physical conditions. These forms share certain common characteristics. Specifically, the phosphate backbones are always located on the outside of the helix. Also, because diesters of phosphoric acid are fully ionized at neutral pH, the exterior of the helix is negatively charged. Bases are well packed in the interior of the helix, where their faces are protected from contact with water. In this environment the strength of hydrogen bonds that connect bases can be maximized. Interwinding of two strands produces a structure having two helical grooves that separate the winding phosphate backbone ridge.

However, the precise **geometry of the double helix** varies among the

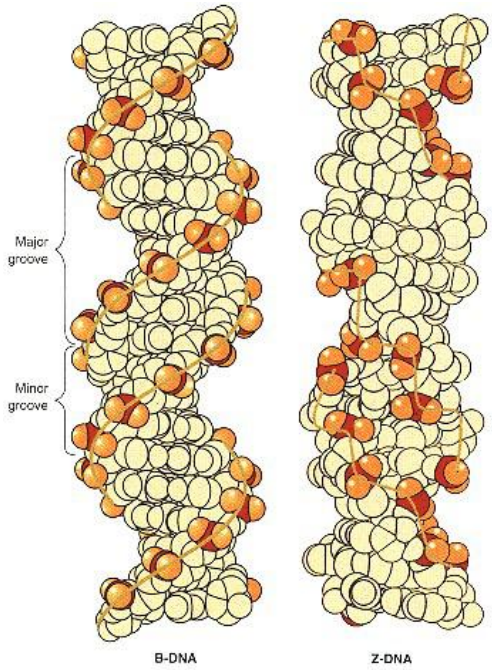


Figure 14.8
Space-filling molecular models of B- and Z-DNA.
The double helix is referred to as the Watson and Crick model, although this structure has been substantially refined since it was proposed. B-DNA may be the most typical form occurring in cells. Z-DNA may be present in cells as small stretches, consisting of alternating purines and pyrimidines, incorporated between long stretches of B-DNA. The zigzag nature of the Z-DNA backbone is illustrated by the heavy lines that connect phosphate residues along the chain.
Redrawn based on figure from Rich, A. *J. Biomol. Struct. Dyn.* 1:1, 1983.

different forms. The original X-ray data obtained with highly oriented DNA fibers suggested occurrence of a form, later designated as B, which appears to be that commonly found in solution and *in vivo* (Figure 14.8). A characteristic of this form is that one of its grooves is wider **(major groove)** than the other **(minor groove).** Disparity in width between these two grooves results from the characteristic geometry of base pairs (bp). Glycosidic bonds between sugars and bases of each base pair are not arranged directly opposite to one another. Instead the edge of the helix, that is more than 180° from glycosidic bond to glycosidic bond, is the edge that forms part of the major groove. Clearly, the opposite edge corresponds to the minor groove. The nucleotide sequence of a polynucleotides can be discerned without dissociating the double helix by looking inside these grooves. As each of the four bases has its own orientation with respect to the rest of the helix, each base always shows the same atoms through the grooves. C-6, N-7, and C-8 of the purine rings and C-4, C-5, and C-6 of the pyrimidine rings line up in the major groove. The minor groove is paved with C-2 and N-3 of the purine and C-2 of the pyrimidine rings. Forms A and C differ from B in the pitch of the base pairs relative to the helix axis as shown in Figure 14.9, as well as in other geometric parameters of the double helix, including conformation of sugar residues, which is one of the more flexible components of the DNA molecule. Alternative forms of the double helix are the result of **conformational variations** of the sugar-phosphate groups that form the backbone of constituent polynucleotides (Figure 14.10). The conformation of the furanose ring of sugar residues exists in nonplanar (puckered) forms. This ring may be visualized in the form of an envelope with four carbon atoms at the corner of the envelope. Oxygen is positioned at the

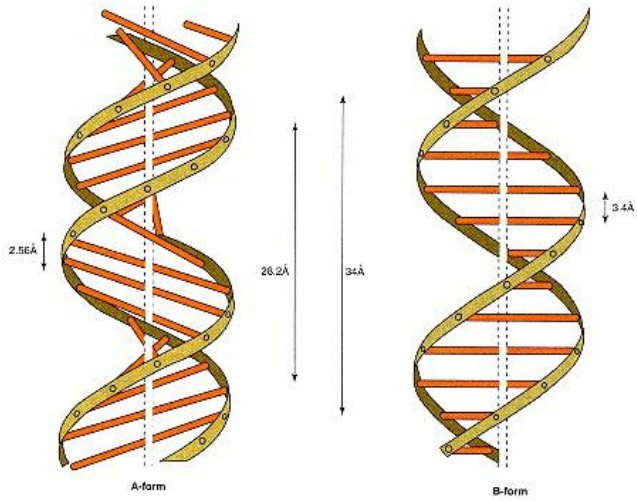


Figure 14.9
Various geometries of DNA double helix.
Depending on conditions, the double helix can acquire various forms of distinct geometries. In the B form of DNA the centers of the bases are about 34 Å apart and produce a complete turn of a helix with a pitch of 34 Å. Such an arrangement results in a complete turn of the helix for every 10 bp. The diameter of the helix is 20 Å. Form C (not shown) is very similar to the B structure, with a pitch of 33 Å and 9 bp per turn. Form A, which is obtained from form B when the relative humidity of the fiber is reduced to 75%, differs from B in that the base pairs are not perpendicular to the helical axis but are tilted. This tilt results in a pitch of 28.2 Å and a shortening of the helix by the packing of 11 pairs per helical turn.
Redrawn based on figure from Guschelbauer, W. *Nucleic Acid Structure.* Berlin: Springer-Verlag, 1976.

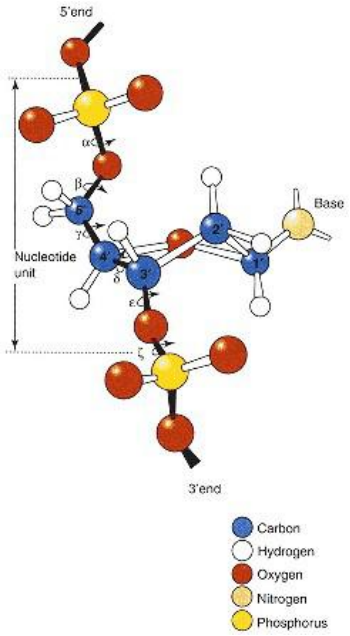


Figure 14.10
Structure of ribose–phosphate backbone of polynucleotides.
The polynucleotide backbone has six degrees of freedom on rotation along the bonds identified by Greek letters to . However, steric hindrance and electrostatic repulsion between the oxygen atoms of the phosphate residue restrict the number of conformational variants that can be generated by rotation along some of these bonds. Rotation is particularly limited, but still possible, around the C-5 –O bond (bond) and the C-3 –C-4 bond (bond).

top of the envelope flap and therefore may bend out of the envelope body. The main body of the envelope may also be twisted. Twisting the C-2 and the C-3 atoms relative to the other atoms produces two distinct forms. C-2 twists up from the plane and results in the C-2 **endo** form. As a rule, atoms that are positioned on the same side of the plane as C-5 have by definition the **endo conformation.** The C-2 **endo** and C-3 **endo** are the most common conformers found in nucleic acids, while free nucleotides in solution are characterized by a rapid equilibrium between these conformers. Another variation in nucleic acid conformations arises from rotations about the C-1 –*N*-glycosidic bond that is responsible for variants known as **syn** and **anti** forms (see Figure 14.11). **Anti** conformations are the predominant forms in nucleic acids, while in free nucleotides in solution **syn–anti** equilibrium depends on the nature of the base. Generally, purine nucleotides are characterized by a rapid **syn–anti** equilibrium while pyrimidines usually adopt **anti** conformations.

Finally, conformational variations in DNA may result from relative orientations of the planes of the bases between strands. Differences in orientation between planes of H-bonded bases may produce double helix variants with different base **tilt, roll, twist,** or **propeller twist.** For example, DNA forms A and B differ drastically in base **tilt** and deviations of **tilt** and **roll** angles, occurring in phage tracts of adenine residues, are responsible for extensive bending of the double helix axis over certain functionally important regions of DNA. Under conditions of low salt concentration and humidity, the thin B-DNA double helix shifts to a conformation characterized by a thicker helix. In this conformation nucleotides move off center toward the major edge of each base pair, generating A-DNA, which has a narrower and deeper major groove and a wider and shallower minor groove than B-DNA. The parameters for these different DNA conformations, listed in Table 14.2, have been determined of DNA by X-ray diffraction methods. While the numbers provide very accurate information about molecular geometry and dimensions of crystalline samples, they give only average dimensions for monomeric units present in a noncrystalline macromolecule. Therefore these parameters are listed as such and the listing does not imply that the same geometry characterizes each and every individual base pair in DNA. Rather, depending on base sequence, considerable local variation in conformation of individual nucleotides may occur. Such varia-

TABLE 14.2 Structural Features of A-, B-, and Z-DNA

Features	***A-DNA***	***B-DNA***	***Z-DNA***
Helix rotation	Right-handed	Right-handed	Left-handed
Base pair per turn (crystal)	10.7	9.7	12
Base pair per turn (fiber)	11	10	—
Base pair per turn (solution)	—	10.5	—
Pitch per turn of helix	24.6 Å	33.2 Å	45.6 Å
Proportions	Short-end broad	Longer and thinner	Elongated and thin
Helix packing diameter	25.5 Å	23.7 Å	18.4 Å
Rise per base pair (crystal)	2.3 Å	3.3 Å	3.7
Rise per base pair (fiber)	2.6	3.4 Å	—
Base pair tilt	+19°	–1.2° (but varies)	–9°
Propeller twist	+18°	+16°	0°
Helix axis rotation	Major groove	Through base pairs	Minor groove
Sugar ring conformation (crystal)	C-3 *endo*	Variable	Alternating
Sugar ring conformation (fiber)	C-3 *endo*	C-2 *endo*	—
Glycosyl bond conformation	anti	anti	anti at C, syn at G

tions may be important in regulation of gene expression, since they influence the extent of DNA binding with various types of regulatory proteins.

A form of DNA, which was discovered more recently, has geometric characteristics radically different from those of conventional forms. In this DNA, called **Z-DNA,** the polynucleotide phosphodiester backbone assumes a "zigzag" arrangement rather than the smooth conformation that characterizes other double-stranded forms. The Z-DNA structure is longer and much thinner than that of B-DNA and completes one turn in 12 bp rather than the 10 bp in a B-DNA turn. It forms a single groove as opposed to two grooves that characterize B-DNA. Therefore the conformation of Z-DNA may be viewed as the result of the major groove of B-DNA having "popped out" in order to form the outer convex surface of Z-DNA. This change places the stacked bases on the outer part of Z-DNA rather than in their conventional positions in the interior of the double helix. Another highly unusual property of the Z structure is that it consists of left-handed rather than right-handed helices, which characterize conventional forms. These major structural differences between B-DNA and Z-DNA (Figure 14.8) are partly the result of different conformations in nucleotide residues between the two forms. Specifically, in B- and A-DNA sugars and bases are arranged in the extended **anti** conformation. In contrast, in Z-DNA some nucleotides rotate into **syn** conformation, which places the sugar and base on the same side of the glycosidic bond (Figure 14.11). DNA sequences that consist of alternating GC nucleotides are the most prone to acquire Z conformation, which places glycosidic bonds of each G in syn, with C residues maintaining the anti conformation. The zigzag arrangement of the phosphate backbone reflects sudden turns of the backbone, as it follows the alternating arrangement of syn and anti geometries.

The **biological function of Z-DNA** is not known with certainty. Some evidence exists suggesting that Z-DNA influences gene expression and regulation. Apparently small stretches of DNA approximately 12–24 bp long with the potential of forming Z-DNA are more commonly found at the 5 end of genes, that is, in regions that regulate transcriptional activities. These stretches consist of alternating purines and pyrimidines that favor formation of the Z conformation. Z-DNA may have a role in genetic recombination. Sites of genetic recombination in eukaryotic cells appear to be associated with DNA regions with the potential of Z-DNA formation. The Z form of DNA is stabilized by the presence of cations or polyamines and by methylation of either guanine residues in C-8 and N-7 positions or cytosine residues in C-5 position. Sequences that are not strictly alternating pu-pyr may also acquire the Z conformation as a result of methylation. For instance, the hexanucleotide ^{m5}GAT^{m5}CG, which contains two internal adjacent pairs of pu and py, forms Z-DNA. This outcome is not surprising because in Z-DNA hydrophobic methyl groups do not protrude unfavorably into the aqueous environment surrounding the double helix, as is the case with B-DNA. On this basis it might be expected that *in vivo* methylation of cytosine also induces a B Z transition in cellular DNA. The suggestion that Z-DNA may have a role in gene regulation is supported by modification in methylation patterns that accompany the process of gene expression.

An important structural characteristic of double-stranded DNA is that its strands are **antiparallel.** Polynucleotides are asymmetric structures with an intrinsic sense of **polarity** built into them (Figure 14.12). The two strands are aligned in opposite directions; if two adjacent bases in the same strand, for example, thymine and cytosine, are connected in the 5 3 direction, their complementary bases adenine and guanine will be linked in the 3 5 direction (directions are defined by linking the 3 and 5 positions within the same nucleotide). This antiparallel alignment produces a stable association between strands to the exclusion of the alternate parallel arrangement.

The double-stranded structure of DNA was proposed in 1953, partly based on previously available X-ray diffraction studies suggesting that the structures

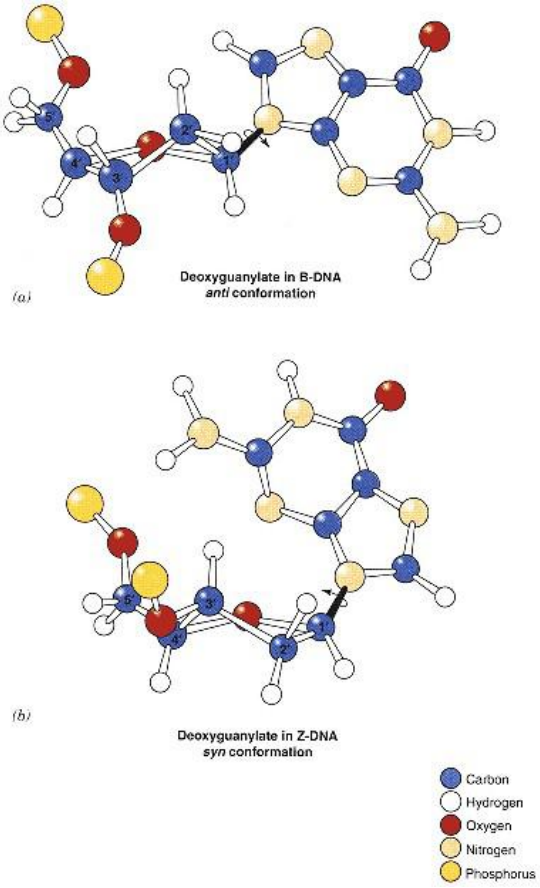


Figure 14.11
Conformational variants of nucleotides.
Rotation of the base plane around the C-1 –N-9 glycosyl bond gives rise to two distinct nucleotide conformations, the so-called anti and syn conformations. The anti conformation is characteristic of B-DNA. In Z-DNA the glycosyl bond rotates as shown to give the syn conformation. The B Z DNA transformation is also accompanied by a change in the conformation of the ribose ring from the C-2 endo to C-3 endo conformation.

of DNA from various sources exhibited remarkable similarities. These studies also suggested that DNA had a helical structure containing two or more polynucleotides. Evidence of central importance to the proposal was the clarification of the quantitative base composition of DNA, indicating the molar equivalence between purines and pyrimidines essential for the complementarity between the two strands.

Many Factors Stabilize DNA Structure

Factors that stabilize single-stranded polynucleotides—that is, **hydrophobic interactions** and **van der Waals forces**—are also instrumental in stabilizing the double helix. Van der Waals interactions generate attractive forces among atoms that are optimally situated, that is, neither too close nor too far apart

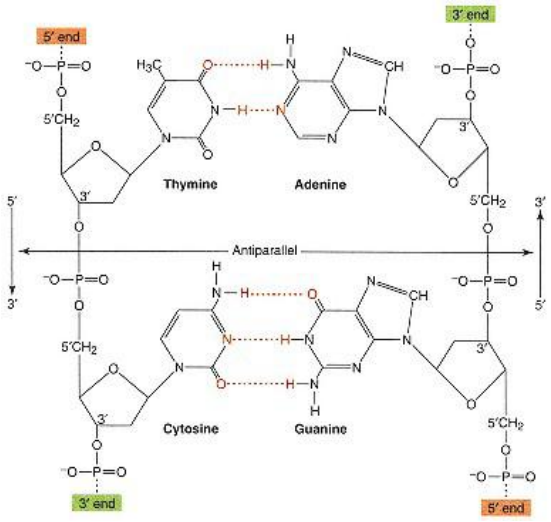


Figure 14.12
Antiparallel nature of DNA strands.
Note the opposite direction of the strands of a double-stranded DNA. The geometry of the helices does not prevent a parallel alignment, but such an arrangement is not found in DNA.

relative to one another, within a molecular structure. These forces are the result of **dipole–dipole interactions** and **London dispersion interactions** (transient dipole interactions) between adjacent bases. Hydrophobic interactions are also very important in stabilizing polynucleotide structures and especially the double helix. The separation between the hydrophobic core of the stacked bases and the hydrophilic exterior of the charged sugar–phosphate groups is even more striking in the double helix than with single-stranded helices. This explains the preponderance of the DNA double helix. The stacking tendency of single-stranded polynucleotides may be viewed as resulting from a tendency of the bases to avoid contact with water. The double-stranded helix is a more favorable arrangement, permitting the phosphate backbone to be highly solvated by water while the bases are essentially removed from the aqueous environment.

Collectively, hydrophobic and van der Waals forces are referred to as **stacking interactions** because they produce the stacked arrangement of the bases typical of the double helix. Stacking interactions are estimated to generate 4–15 kcal mol^{-1} for each adjacent pair of stacked bases.

Additional stabilization of both single-stranded DNA as well as the double helix results from extensive networks of **cooperative hydrogen bonding.** Typically, hydrogen bonds are relatively weak (3–7 kcal mol^{-1}) and are even weaker in DNA (2–3 kcal mol^{-1}) because of geometric constraints within the double helix. Cumulatively, however, H bonds provide substantial energies of stabilization for the double helix although the stabilization is less than what is provided by stacking interactions. However, hydrogen bonding, in contrast to stacking forces, does not confer to any significant degree preferential stabilization to the double helix relative to its constituent single-stranded polynucleotides, which can form equally effective hydrogen bonds with water molecules in an aqueous environment.

Hydrogen bonds have important biochemical consequences for the functions in which the double helix participates. In contrast to stacking forces, hydrogen bonds are highly directional and are able to provide a discriminatory function for choosing between correct and incorrect base pairs. Because of

TABLE 14.3 Effects of Various Reagents on the Stability of the Double Helix[a]

Reagent	*Adenine Solubility × 10^{-3} (in 1 M reagent)*	*Molarity Producing 50% Denaturation*
Ethylurea	22.5	0.60
Propionamide	22.5	0.62
Ethanol	17.7	1.2
Urea	17.7	1.0
Methanol	15.9	3.5
Formamide	15.4	1.9

Source: Data from Levine, L., Gordon, J., and Jencks, W. P. *Biochemistry* 2:168, 1963.

[a] The destabilizing effect of the reagents listed below on the double helix is independent of the ability of these reagents to break hydrogen bonds. Rather, the destabilizing effect is determined by the solubility of adenine. Similar results would be expected if the solubilities of the other bases were examined.

their directionality, hydrogen bonds tend to orient the bases in a way that favors stacking. Therefore the contribution of hydrogen bonds is essential for the stability of the double helix.

The relative importance of hydrogen bonding and stacking forces in stabilizing the double helix was not always appreciated. The effects of various reagents on the stability of the double helix have suggested that the destabilizing effect of a reagent is not related to the ability of the reagent to break hydrogen bonds. Rather, the stability of the double helix is determined by the solubility of the free bases in the reagent, the stability decreasing as the solubility increases. Some of these findings, summarized in Table 14.3, emphasize the importance of hydrophobic forces in maintaining the structure of double-stranded DNA.

A direct consequence of the conclusion that the relative stability of the double helix versus the single-stranded DNA depends almost exclusively on stacking forces is that differences in the stabilities of various segments of the double helix reflect variabilities in the stacking energies of different base sequences. Indeed, a large degree of variability exists among the stacking energies of various pairs of stacked bases as shown in Table 14.4. As a rule, stacking interaction involving dimers of G-C base pairs are stronger than interactions between stacked dimers of A-T base pairs.

Ionic forces also have an effect on the stability and conformation of the double helix. At physiological pH, the electrostatic intrastrand repulsion between negatively charged phosphates is potentially destabilizing and it forces the double helix into a relatively rigid rod-like conformation. In addition, this repulsion tends to separate the complementary strands. In distilled water, DNA strands will separate at room temperature; near the physiological salt concentration, cations, particularly Mg^{2+} (in addition to other charged groups, e.g., the basic side chains of proteins), shield the phosphate groups and decrease repulsive forces. Therefore the flexibility of the double helix is partially restored and its stability is enhanced.

Denaturation

The double helix is disrupted during almost every important biological transformation in which DNA participates, including DNA replication, transcription, repair, and recombination. Therefore the forces that hold the two strands together are adequate for providing stability and yet weak enough to allow facile

TABLE 14.4 Base Pair Stacking Energies

Dinucleotide Base Pairs	*Stacking Energies (kcal mol^{-1} per stacked pair)[a]*
(GC) · (GC)	−14.59
(AC) · (GT)	−10.51
(TC) · (GA)	− 9.81
(CG) · (CG)	− 9.69
(GG) · (CC)	− 8.26
(AT) · (AT)	− 6.57
(TG) · (CA)	− 6.57
(AG) · (CT)	− 6.78
(AA) · (TT)	− 5.37
(TA) · (TA)	− 3.82

[a] Data from Ornstein, R. L., Reim, R., Breen, D. L., and Mc Elroy, R. D. *Biopolymers* 17:2341, 1978.

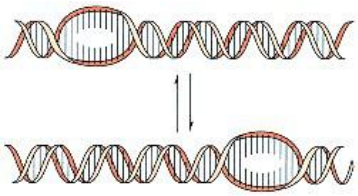


Figure 14.13
"Zipper" model for DNA double helix.
DNA contains short sections of open-strandedness that can "move" up and down the helix.

strand separation. In fact, the double helix is stabilized relative to the single strands by about 1 kcal per base pair. Therefore a relatively minor perturbation can produce disruption in double strandedness, provided that only a short section of the DNA is involved. As soon as the relatively few base pairs have separated, they close up again and release free energy, and then the adjacent base pairs unwind. In this manner minor disruptions of double strandedness can be propagated along the length of the double helix. Thus, at any particular moment, the large majority of the bases of the double helix remain hydrogen bonded, but all bases can pass through the single-stranded state, a few at a time. This **dynamic state** of the double helix is characterized by the movement of an "open-stranded" portion up and down the length of the helix, as indicated in Figure 14.13. The "dynamic" nature of this structure is an essential prerequisite for the biological functions of DNA as it undergoes repair or recombination.

Separation of DNA strands can be studied by increasing the temperature in solution. At relatively low temperatures a few base pairs will be disrupted, creating one or more "open-stranded bubbles." These "bubbles" form initially in sections that contain relatively higher proportions of adenine and thymine pairs because of the lower stacking energies of dimers of such pairs. As the temperature is raised, the size of the "bubbles" increases and eventually the thermal motion of the polynucleotides overcomes the forces that stabilize the double helix. This transformation is depicted in Figure 14.14. At even higher temperatures the strands separate physically and acquire a random-coil conformation (Figure 14.15). The process is most appropriately described as a **helix-to-coil transition,** but it is commonly called **denaturation.** This is accompanied by a number of physical changes, including a buoyant density increase, reduction in viscosity, change in ability to rotate polarized light, and changes in absorbancy.

Changes in absorbance are frequently used to follow the process of denaturation experimentally. DNA absorbs in the UV region due to the heterocyclic aromatic nature of its purine and pyrimidine constituents. Although each base has a unique absorption spectrum, all bases exhibit maxima at or near 260 nm. This property is responsible for the absorption of DNA at 260 nm. However,

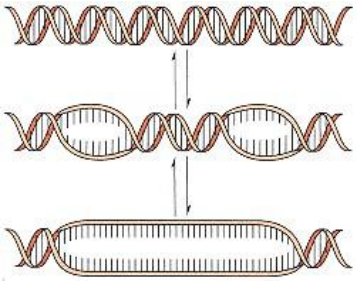


Figure 14.14
Structure of double-stranded DNA at increasing temperatures.
Disruptions of the double-stranded structure appear first in regions of relatively high adenine–thymine content. The size of these "bubbles" increases with increasing temperatures, leading to extensive disruptions in the structure of the double helix at elevated temperatures.

this absorbancy can be as much as 40% lower than that expected from adding up the absorbancy of each of the base components of DNA. This property of DNA, referred to as **hypochromic effect,** results from the stacking of the bases along the DNA helices. In this arrangement, interactions between the π electrons of neighboring bases produce a decrease in absorbancy. However, as the ordered structure of the double helix is disrupted at increasing temperatures, stacking interactions are gradually decreased. Therefore a totally disordered polynucleotide approaches an absorbance not very different from the sum of the absorbancies of its purine and pyrimidine constituents.

Slow heating of double-stranded DNA in solution is accompanied by a gradual change in absorbancy as the strands separate. However, since the interactions between the two strands are cooperative, the transition from double-stranded to random-coil conformation occurs over a narrow range of temperatures, as indicated in Figure 14.16. Before the rise of the **melting curve,** DNA is double stranded. In the rising section of the curve an increasing number of base pairs are interrupted as the temperature rises. Strand separation occurs at a critical temperature corresponding to the upper plateau of the curve. However, if the temperature is decreased before the complete separation of the strands, the native structure is completely restored.

The **midpoint temperature,** T_m, of this process, under standard conditions of concentration and ionic strength, is characteristic of the base content of each DNA. The higher the guanine–cytosine content, the higher the transition temperature between the double-stranded helix and the single strands. This difference in T_m values is attributed to the increased stability of guanine-cytosine pairs, as a result of the higher stacking interactions between dimers of G-C pairs relative to the dimers of A-T pairs.

Rapid cooling of a heated DNA solution normally produces denatured DNA, a structure that results from the reformation of some hydrogen bonds either between the separate strands or between different sections of the same strand. The latter must contain complementary base sequences. By and large denatured DNA is a disordered structure containing substantial amounts of **random-coil** and single-stranded regions.

DNA can also be denatured at a pH above 11.3 as the charge on several substituents on the rings of the bases is changed, preventing these groups from participating in hydrogen bonding. **Alkaline denaturation** is often used as an experimental tool in preference to heat denaturation to prevent breakage of phosphodiester bonds that can occur to some degree at high temperatures or low pH. Denaturation can also be induced at low ionic strengths, because of enhanced interstrand repulsion between negatively charged phosphates, as well as by various denaturing reagents, that is, compounds that weaken or break

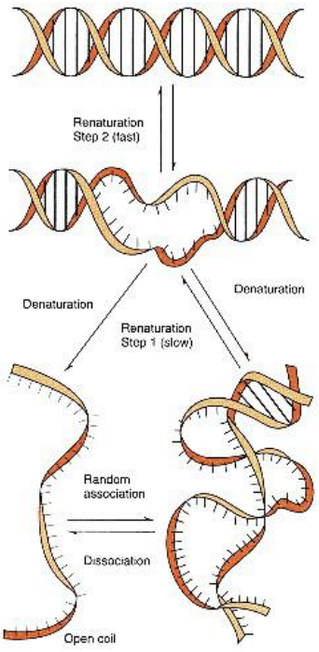


Figure 14.15
Denaturation of DNA.
At high temperatures the double-stranded structure of DNA is completely disrupted, with the eventual separation of the strands and the formation of single-stranded open coils. Denaturation also occurs at extreme pH ranges or at extreme ionic strengths.

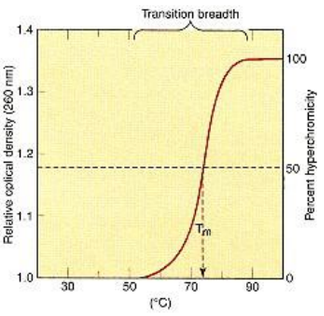


Figure 14.16
Temperature–optical density profile for DNA.
When DNA is heated, the optical density increases with rising temperature. A graph in which optical density versus temperature is plotted is called a "melting curve." Relative optical density is the ratio of the optical density at the temperature indicated to that at 25°C. The temperature at which one-half of the maximum optical density is reached is the midpoint temperature (T_m).
Redrawn based on figure from Freifelder, D. *The DNA Molecule: Structure and Properties.* San Francisco: Freeman, 1978.

hydrogen bonds. A complete denaturation curve similar to that shown in Figure 14.16 can be obtained at a relatively low constant temperature, for instance, room temperature, by variation of the concentration of an added denaturant.

Renaturation

Complementary DNA strands, separated by denaturation, can reform a double helix if appropriately treated. This is called **renaturation** or **reannealing.** If denaturation is not complete and only a few bases remain hydrogen bonded between the two strands, the helix-to-coil transition is rapidly reversible. **Annealing** is possible even after the complementary strands have been completely separated. Under these conditions the renaturation process depends on the meeting of complementary DNA strands in an exact manner that can lead to the reformation of the original structure, and it is a slow, concentration-dependent process. As a rule, maintaining DNA at 10–15°C below its T_m, under conditions of moderate ionic strength (0.15 M), provides the maximum opportunity for renaturation. At lower salt concentrations, the charged phosphate groups repel one another and prevent the strands from associating. As renaturation begins, some of the hydrogen bonds formed are extended between short tracts of polynucleotides that might have been distant in the original native structure. Renaturation is facilitated by the presence of short sequences, consisting of four to six base pairs, reiterated many times within every DNA strand. A large number of much longer nucleotide sequences are repeated many times within the eukaryotic genome. Such sequences provide sites for initial base pairing that produces a partially hydrogen-bonded double helix. These randomly base-paired structures are short-lived because the bases that surround the short complementary segments cannot pair and lead to the formation of a stable fully hydrogen-bonded structure. However, once the correct bases begin to pair by chance, the double helix over the entire DNA molecule is rapidly reformed. Renaturation is a two-step process. The first step determines the rate of association, involves the chance meeting of two complementary sequences on different strands, and is therefore a second-order reaction. The rate of renaturation is thus proportional to the product of the concentrations of the two homologous dissociated strands and is expressed as $dt/dc = -kc^2$, where k is the rate constant for the association. Integration of this equation gives $C/C_0 = 1/(1 + kC_0t)$, where C is the concentration of single-stranded DNA expressed as moles of nucleotide per liter at time t, and C_0 is the concentration of DNA at time zero. A plot of C/C_0 (which is proportional to DNA that is single stranded or of the DNA fraction that is reassociated) versus C_0t can be constructed (Figure 14.17), and a $C_0t_{0.5}$ **(Cot-a-half)** value, which corresponds to $C/C_0 = 0.5$ can be determined. The $C_0t_{0.5}$ value is proportional to the complexity of the genome. **Complexity** is equal to the molecular mass of the genome provided that the genome consists of unique nucleotide sequences. For example, both the complexity and molecular mass of a hypothetical genome consisting of three unique nucleotide sequences that may be represented as N_1, N_2, and N_3 is equal to the sum $N_1 + N_2 + N_3$. However, in eukaryotic genomes, which contain both unique as well as reiterated sequences, the complexity of the genome is significantly lower than the molecular mass. If, for instance, a eukaryotic genome contains 10^5 copies of sequence N_3, 10^3 copies of sequence N_2, and 1 copy of sequence N_1, the complexity will still be $N_1 + N_2 + N_3$ but the molecular mass will be equal to $10^5N_3 + 10^3N_2 + N_1$. Thus complexity may be defined as the minimum length of DNA that contains a single complete copy of all the single and **reiterated sequences** that are represented within the genome.

The C_0t curves of eukaryotic genomes with reiterated DNA segments show several kinetic components, each representing those parts that have similar reiteration frequencies (Figure 14.18). Highly reiterated sequences will reassociate the fastest; unique sequences are the slowest. Thus C_0t curves provide information on **genome complexity,** on the number of **repetitive classes,**

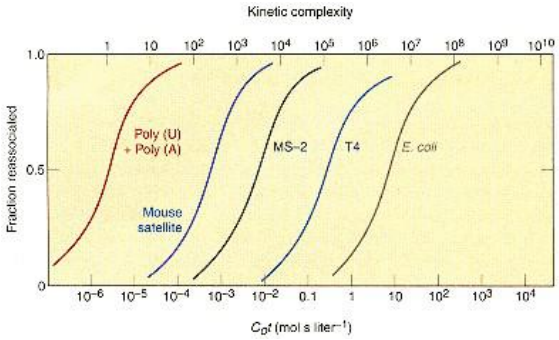
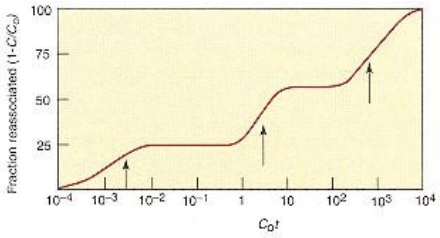


Figure 14.17
Reassociation kinetics for DNA isolated from various sources.
Each DNA is first fragmented to segments of approximately 400 nucleotides. The denatured segments are subsequently allowed to renature. The fraction of each poly-nucleotide reassociated, calculated from changes in hypochromicity, is plotted against the total concentration of nucleotides multiplied by the renaturation time (C_0t). The top scale shows the kinetic complexity of each DNA sample. Whenever a DNA contains reiterated sequences, these sequences are present in the fragments at higher concentrations than they would have been if a unique sequence had been fragmented. As a result, renaturation of fragments, obtained from DNAs containing reiterated sequences, proceeds more rapidly the higher the degree of repetition. This is exemplified by the rates of renaturation of fragments obtained from the synthetic double-stranded polynucleotide poly(A)–poly(U) and mouse satellite DNA, a DNA that contains many repeated sequences. For a homogeneous DNA, which contains a distribution of different extents of reiterated sequences, kinetic complexity can be defined as the minimum length of DNA needed to contain a whole single copy of the reiterated sequence.
Based on figure in Britten, R. J., and Kohne, D. E. *Science* 161:529, 1968.



	Fast Component	Intermediate Component	Slow Component
Percent of genome	25	30	45
$C_0t_{1/2}$	0.0013	1.9	630
Complexity, bp	340	6.0×10^5	3.0×10^8
Repetition frequency	500,000	350	1

Figure 14.18
Reassociation kinetics of eukaryotic DNA.
This idealized $C_0t_{1/2}$ plot represents a eukaryotic DNA that consists of three distinct components with three different C_0t values. The percentage to which each one of these components is present in the DNA can be read from the ordinate (fraction-reassociated axis) of the z igure. Repetition frequencies and complexities are calculated based on the principles discussed in the text. In practice, the experimental separation of different DNA components is not as pronounced and their identification not as clear-cut as shown in this hypothetical example.

and on the proportion of the total genome represented by those classes. Since most genes occur only once within a genome, separation of DNA into different repetitive classes facilitates the search for individual genes by narrowing the search within the single copy component of DNA.

Hybridization

Self-association of complementary polynucleotide strands has provided the basis for development of the technique of **hybridization.** This depends on the association between two polynucleotide chains, which may be of the same or of different origin or length, provided that a base complementarity exists between these chains. Hybridization can take place not only between DNA chains but also between complementary RNA chains as well as DNA–RNA combinations.

Appropriate techniques have been developed for measuring the maximum amount of polynucleotide that can be hybridized as well as the rates of hybridization. These techniques are important basic tools of contemporary molecular biology and are being used for the following: (1) determining whether or not a certain sequence occurs more than once in the DNA of a particular organism, (2) demonstrating a genetic or evolutionary relatedness between different organisms, (3) determining the number of genes transcribed in a particular mRNA (clearly DNA–RNA hybridizations are needed for accomplishing the last goal), and (4) determining the location of any given DNA sequence by annealing with a complementary polynucleotide, called a **probe,** that is appropriately tagged for easy detection of the hybrid.

DNA to be tested for hybridization is denatured. The resulting single strands are immobilized by binding to a suitable polymer, which is then used to pack a chromatography column. DNA formed in the presence of labeled precursors, usually tritiated thymidine, is allowed to run through the column that contains the bound, unlabeled DNA. The rate at which radioactivity is retained by the column equals the rate of annealing between complementary strands.

Determination of the maximum amount of DNA that can be hybridized can establish homologies between DNA of different species since the base sequences in each organism are unique. On this basis annealing can be used to compare the degree to which DNAs isolated from different species are related to one another. The observed homologies serve as indexes of **evolutionary relatedness** and have been particularly useful for defining phylogenies in prokaryotes. Hybridization studies between DNA and RNA have, in addition, provided very useful information about the biological role of DNA, particularly the mechanism of transcription.

Hybridization techniques using membrane filters, usually made of nitrocellulose, have found increasing application. In general, hybridization can be quantitated by either measuring the amount of hybrid in equilibrium or the rate of hybrid formation under conditions in which one nucleic acid is present in large excess. The approach used for the latter determination is analogous to the C_0t procedure and when it is used for DNA–RNA hybridization and RNA is present in excess it is referred to as the ***R_0t* method,** or the ***D_0t* method** when DNA is in excess.

A variant of filter hybridization, known as the **Southern transfer,** can be used for identifying the location of specific genes (see p. 774). Since a gene sequence represents a very small percentage of total DNA, the gene must be separated from the remaining DNA and the DNA detected by using appropriate probes. Another variation of hybridization known as ***in situ* hybridization** uses intact DNA molecules within metaphase chromosomes. The chromosomes are spread on slides and subjected to denaturation and then exposed to a probe labeled with a fluorescent molecule. The DNA sequence of interest is located by observation with a fluorescence microscope.

DNA Probes

Probes are short single-stranded RNA or DNA oligonucleotides that are complementary to specific sequences of interest in genomic DNA. Under proper conditions probes interact only with a segment of interest, indicating whether the segment is present in a particular sample of DNA. Probes synthesized by chemical means may appear to be limited by the degree to which the desired genomic nucleotide sequence is known, but in fact this approach has much wider applicability. As an example, if the protein product of a gene is known, the nucleotide sequence of the desired gene can be approximated by using a mixture of different synthetic oligonucleotides that represent alternate mRNA sequences that, because of degeneracy of the code, can encode for the same protein. One of these oligonucleotide sequences is therefore complementary to the desired gene. When the gene of interest is transcribed to mRNA molecules that are abundant and easily purified, mRNA can be used. Probes need to be at least 15 nucleotides long because shorter sequences may occur randomly along genomic DNA. To achieve easy detection, probes are labeled by the incorporation of ^{32}P or are identified by the use of biotin-containing nucleotides that are incorporated into the probe and serve as fluorescent labels. Probes are useful for definitive and rapid diagnosis of genetic disorders, infectious disease, and cancer as described briefly in Clin. Corr. 14.2.

Heteroduplexes

Hybridization is the basis for a technique that has permitted construction of precise physical maps of DNA genes. This technique depends on direct visualization under the electron microscope of single-stranded loops in the structures of artificially formed double-stranded DNA molecules known as **heteroduplexes** constructed by hybridization of two complementary DNA strands. One strand is selected on the basis that, as the result of a known mutation, it misses the

CLINICAL CORRELATION 14.2

Diagnostic Use of Probes in Medicine

A probe is a molecule with a strong affinity for a specific target, which can easily be detected after its interaction with the target. The specificity of DNA probes is based on interaction between complementary polynucleotide strands. Probes can be obtained by amplification of naturally occurring DNA sequences or by chemical synthesis. Use of DNA-based techniques is becoming increasingly important in laboratory diagnosis of many genetic diseases and certain types of cancers. The method is used selectively in diagnosis of bacterial infections for bacteria that are slow growing or difficult to identify by conventional culture-base methods, such as bacteria causing Lyme disease (*Borrelia burgdorferi*), certain types of syphilis (*Treponema pallidum*), or pneumonia (*Chlamydia pneumoniae*). In addition, DNA probes are indispensable for identification of bacteria that are extremely difficult or impossible to grow in culture, such as organisms responsible for Lepra (*Mycobacterium leprae*) and Whipple's disease (*Tropheryma whippellii*). DNA-based techniques also have the potential to provide faster, more versatile, and less expensive diagnostic applications for detection of more common bacterial infections. Hybridization procedures generally begin with amplification of target DNA (bacterial DNA) by cloning or more commonly by a technique known as the polymerase chain reaction (PCR). The probe hybridized with target DNA is typically detected by the Southern blot technique.

Probes are very useful for identification of mutant alleles responsible for genetic diseases, especially if the mutations are stable and few in number. Some genetic disorders are due to mutations in a single gene and in some instances appear to correlate well with a particular phenotype or symptom. Detection of mutations can be of diagnostic value. One approach used for direct identification of mutations involves hybridization with an allele-specific probe (ASP). Examples of diseases diagnosed by probes are sickle cell anemia, hemoglobin C disease, and phenylketonuria. The first two are the result of a single base change in genes coding for β-globin. By using three different probes, corresponding to the sequence of normal and two mutated hemoglobins, the presence of mutated β-globin genes can be detected. Similarly, a probe can identify a mutation in the phenylalanine 4-monooxygenase gene that is responsible for phenylketonuria. Many other genetic diseases, including cystic fibrosis, Gaucher's disease, β-thalassemia, and Tay-Sachs disease, can be diagnosed using DNA-based techniques.

Keller, G. H., and Manak, M. M. *DNA Probes*. New York: Stockton Press, 1993.

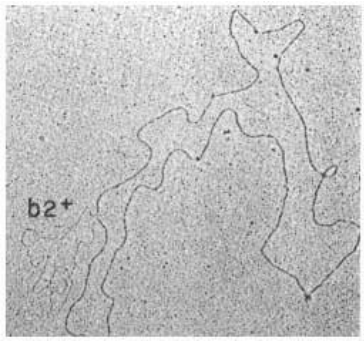


Figure 14.19
Heteroduplex formation in bacteriophage.
Electron micrograph of a heteroduplex DNA molecule constructed from complementary strands of bacteriophage and a bacteriophage deletion mutant (bacteriophage b2). In b2 a segment of DNA has been deleted, producing, at the site of deletion, a loop labeled b2+.
Reprinted with permission from Westmoreland, B. C., Szybalski, W., and Ris, H. *Science 163:1343, 1969.*

gene being mapped. As shown in Figure 14.19, the complementary strands of the heteroduplex pair perfectly throughout the length of the molecule except that across from the position of the missing gene in the mutant strand the complementary strand forms a visible loop. The position of the loop identifies the location of the deleted gene.

14.3— Types of DNA Structure

Only the essential features common to all DNAs have been presented so far. The specific structural features of DNA vary, depending on the origin and function of each DNA molecule. Molecules of DNA differ in size, conformation, and topology.

Size of DNA Is Highly Variable

The length of DNA varies from a few thousand base pairs for DNA of the small viruses, to millions for chromosomal DNA of bacteria, and to billions for the chromosomal DNA of animals. DNA size can be expressed as number of base pairs, molecular mass, the length of the strands, and even the actual mass of DNA. The units used in these expressions, however, can easily be interconverted, since a DNA of mol wt 1×10^6 contains approximately 1500 bp and is 0.5 nm long. DNA mass can be converted to molecular mass by division with the average molecular mass of a DNA nucleotide pair.

The amount of DNA per cell increases as the complexity of the cellular function increases (Table 14.5). Although mammalian cells contain some of the

TABLE 14.5 DNA Cell Content of Some Species

Type of Cell	*Organism*	*DNA per Cell (pg)*[a]
Phage	T4	2.4×10^{-4}
Bacterium	*E. coli*	4.4×10^{-3}
Fungus	*N. crassa*	1.7×10^{-2}
Avian erythrocyte	Chicken	2.5
Mammalian leukocyte	Human	3.4

Source: From Lewin, B. *Gene Expression*, Vol. 2, 2nd ed. New York: Wiley, 1980, p. 958.

[a] pg, picograms.

highest amounts of DNA per cell, some amphibian, fish, and plant cells may contain even higher amounts. In fact, lung fish cells contain more than 40 times the amount of DNA in human cells, but such extraordinary amounts of DNA reflect a **reiteration of nucleotide sequences** within the DNA macromolecule and do not represent an actual increase in the size of DNA in terms of unique sequences, that is, **DNA complexity.** The size of the DNA of higher cells is very large indeed. The DNA contained within a single human cell is packaged in the form of 46 chromatin fibers or chromosomes. In its most condensed state, that is, during metaphase, the largest of these chromosomes is about 10 μm. If the DNA packaged within this chromosome were stretched out in the conventional B-DNA form, it would be over 8 cm long, that is, 8000 times longer than it is when packed within the chromosome. This suggests that the polynucleotides are exquisitely packed in order to fit within the minute dimensions of the cell nucleus.

Because of their extraordinary length, relative to the total mass, DNA molecules are extremely sensitive to shearing forces that develop during ordinary laboratory manipulations. Even careful pipetting may shear a DNA molecule. During the process of isolation it is difficult to prevent with absolute confidence the disruption of some phosphodiester bonds by contaminating endonucleases (nicking). For these reasons the precise size of DNA of higher species could not be determined until special handling techniques were developed, both for the isolation of DNA and the measurement of its molecular mass.

Techniques for Determining DNA Size

Classical methods for determining size in proteins proved to be unsuitable for measuring the molecular mass of even relatively small DNAs. Custom-tailored methods were devised. **Equilibrium centrifugation** in a density gradient (usually a concentrated cesium chloride solution), **electron microscopy**, and **electrophoresis** in agarose gels are among the principal methods providing reliable information about the molecular masses of DNAs. Electron microscopy provides a measure of the length of DNA strands. Molecular masses can be calculated from known values of the mass per unit length. The DNA can be visualized under the electron microscope if it is first coated with protein and a metal film. Determination of molecular masses by electrophoresis depends on the molecular sieving effect of porous agarose gels. Over a limited range of molecular masses the mobility of DNA is directly proportional to the logarithm of the molecule's weight.

To determine the molecular mass of DNA by equilibrium centrifugation a small portion of a DNA solution to be analyzed is layered on top of a gradient in a centrifuge tube. Upon centrifugation, the molecules of DNA sediment to equilibrium through the gradient. Under these conditions a homogeneous high molecular mass DNA will form a Gaussian band centered at a position in the gradient that corresponds to the density of the DNA. Molecules with different densities are resolved into a series of bands that sediment independently of one another, as shown in Figure 14.20. A relationship can be demonstrated between the width of the bands at equilibrium and the molecular masses.

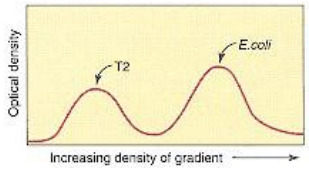


Figure 14.20
Equilibrium gradient centrifugation of DNA.
The DNA macromolecules travel into the increasingly dense regions of the gradient driven by centrifugal forces. The macromolecules equilibrate as soon as they reach an area of the gradient of density equal to their own. For example, bacteriophage T2 DNA and *E. coli* DNA can be resolved into two distinct bands. The width of the bands at equilibrium is related to the molecular weight of DNA.

Labeling of the terminals of DNA has been used successfully for determining molecular masses. DNA is treated with the enzyme alkaline phosphatase, which converts the 5 -phosphate nucleotide terminals of double-stranded DNA to the corresponding OH groups. These terminals are then esterified, using [γ-^{32}P]ATP with the enzyme polynucleotide kinase. The free 5 terminus of each polynucleotide chain becomes labeled as shown in Figure 14.21. The labeled DNA is then analyzed by **zonal centrifugation** and detected from both its absorbancy at 260 nm and ^{32}P content (Figure 14.22). The molecular mass is calculated from the ratio of the amount of ^{32}P to the absorbancy, both measured at the coinciding peaks of the bands.

Gel electrophoresis (see page 773) has replaced electron microscopy and

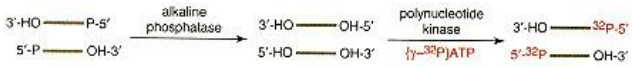


Figure 14.21
End-group labeling procedure.
The 5 terminals on the opposite ends of DNA are labeled with ^{32}P by treatment with alkaline phosphatase and esterification of the resulting 5 -hydroxyl groups with ATP.

centrifugation-based methods for the routine determination of DNA molecular weights. The above methods have permitted determination of DNA molecular masses with an accuracy of at least 10%, but the usefulness of each method is limited within certain molecular mass ranges. Electrophoresis is most suitable between 7.5×10^5 and 1.5×10^7. Electron microscopy is useful for up to 2×10^8 molecular mass. The most versatile method is equilibrium centrifugation, the range of which extends between 2×10^5 and 10^9.

DNA May Be Linear or Circular

DNAs of several small viruses are linear double-stranded helices of equal size. Some DNAs have naturally occurring interior single-stranded breaks. The breaks found in natural bacteriophage molecules result mostly from broken phosphodiester bonds, although occasionally a deoxyribonucleoside may be missing. DNA of coliphage T5 consists of one intact strand and a complementary strand, which is really four well-defined complementary fragments ordered perfectly along the intact strand. A similar regularity in the points of strand breaks is noted, for example, in *Pseudomonas aeruginosa* phage B3, but generally interior breaks seem to be randomly distributed. The double helix structure is maintained because the breaks in one strand are generally in different locations from breaks in the complementary strand.

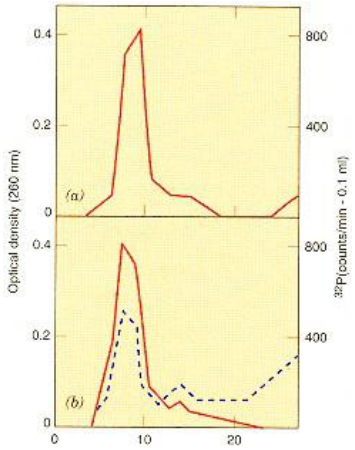


Figure 14.22
Zonal centrifugation profiles of denatured T7 DNA treated by the end-group labeling procedure.
Sedimentation is from right to left.
(a) Untreated DNA.
(b) DNA treated by the end-group labeling procedure. Zonal centrifugation is performed on a sucrose density gradient and should be distinguished from density gradient centrifugation. The latter is an equilibrium centrifugation with the macromolecules reaching equilibrium at regions within the tube at which their density equals the density of the environment. With zonal centrifugation the macromolecules move continuously until they reach the bottom of the tube or until the centrifuge is stopped. The molecular mass is calculated from the ratio of ^{32}P (dotted line) to optical density (solid line) at the peak of the curve. Redrawn from Richardson, C. C. *J. Mol. Biol.* 15:49, 1966.

Double-Stranded Circles

Most naturally occurring DNAs exist in circular form. In some instances **circular DNA** exists as **interlocked circles** or **catenates**. Provided that suitable precautions are taken to avoid shearing the DNA, the circular form can be isolated intact and observed by electron microscopy. The circular structure results from the circularization of a linear DNA by formation of a phosphodiester bond between the 3 and 5 terminals of a linear polynucleotide. Circular structures present many advantages for chromosomal DNA, protecting it from the action of exonucleases and facilitating the process of DNA replication.

The circular nature of small phage ϕX174 DNA was suspected from studies showing that no ends were available for reactions with exonucleases. Sedimentation studies also revealed that endonuclease cleavage yielded one rather than two polynucleotides. These suspicions were later confirmed by direct observation with electron microscopy.

After the circular nature of the DNA chromosome of *E. coli* was demonstrated, it became apparent that many other DNAs (e.g., those of mitochondria, chloroplasts, bacterial plasmids, and mammalian viruses) also existed as closed circles. Obviously, the strands of a circular DNA cannot be irreversibly separated by denaturation because they exist as intertwined closed circles. The absence of 3 or 5 termini provides an evolutionary advantage because it endows the circular DNA with complete resistance toward exonucleases, which ensures the longevity of DNA.

DNA of some bacteriophages exists in a linear double-stranded form that circularizes when it enters the host cell. The linear DNA of bacteriophage λ of *E. coli*, for instance, has single-stranded 5 terminals consisting of 20 nucleotides each. These have complementary sequences, so that an open circle structure

can be formed when the linear molecule acquires a circular shape, which allows the overlap of these complementary sequences. Subsequently, the enzyme DNA ligase, which forms phosphodiester bonds between properly aligned polynucleotides, joins the 3 - and 5 -terminal residues of each strand and forms a covalently closed circle (Figure 14.23).

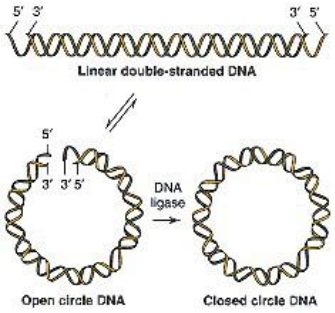


Figure 14.23
Circularization of λ DNA.
The DNA of bacteriophage exists in both linear and circular forms, which are interconvertible. The circularization of DNA is possible because of the complementary nature of the single-stranded 5 terminals of the linear form.

Single-Stranded DNA

With the exception of a few small bacteriophages (e.g., ϕX174 and G4) that can acquire a single-stranded form, most circular and linear DNAs exist as double-stranded helices. The single-stranded nature of the nonreplicative form of ϕX174 DNA was suspected when it was discovered that the base composition of this DNA did not conform to the base equivalence rules; that is, A did not equal T and G did not equal C.

Circular DNA Is a Superhelix

Double-stranded circular DNA, with few exceptions, has an intriguing **topology.** The circular structure contains twists, referred to as **supercoils,** which are visualized by electron microscopy. In principle, linear DNA could be converted to a circular molecule. Circular DNA may be formed by bringing together, and joining by a phosphodiester bond, the free terminals of linear DNA. If no other manipulations are introduced, the resulting circular DNA will be **relaxed;** that is, it will have a thermodynamically favored structure of the linear double helix (B-DNA), which accommodates one complete turn of the helix for approximately 10 base pairs. However, if before sealing the circle, one DNA terminus is held steady while the other terminus is rotated in a direction that unwinds the double helix, the resulting structure will be strained. This strained structure, which is characterized by a deficit of turns, is known as **negative superhelical DNA** (Figure 14.24). Negatively supercoiled DNA is underwound in that it has fewer helical turns than what the molecule would accommodate as a linear or as a relaxed structure. The underwinding results in participation of more base pairs per helical turn, which produces a decrease in the angle of twist between adjacent base pairs. Therefore underwinding generates **torsional tension.** Torsional strain increases the standard free energy of DNA by about 10 kcal mol^{-1} per each supercoil that is introduced into the structure. The strain produced by this deficit of turns is accommodated by the disruption of hydrogen bonds

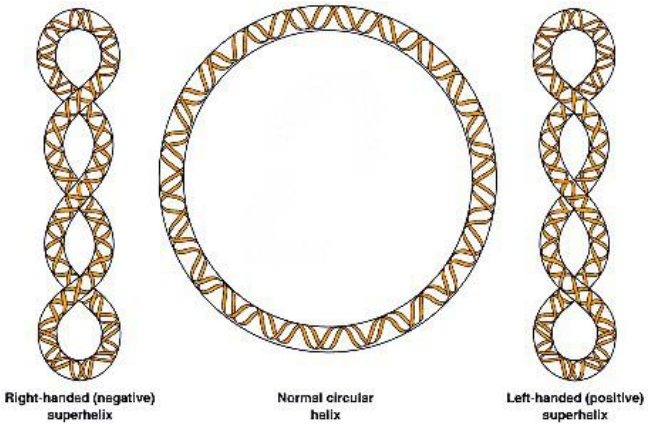


Figure 14.24
Relaxed and supercoiled DNA.
Relaxed DNA can be converted to either right- or left-handed superhelical DNA. Right-handed DNA (negatively supercoiled DNA) is the form normally present in cells. Left-handed DNA may also be transiently generated as DNA is subjected to enzymatically catalyzed transformations (replication, recombination, etc.) and it is also present stably in certain bacterial species. The distinctly different patterns of folding for right- and left-handed DNA are apparent in this representation of the two types of superhelices.
Redrawn from Darnell, J., Lodish, H., and Baltimore, D. *Molecular Cell Biology.* New York: Freeman, 1986.

and the opening of the double helix over a small region of the macromolecular structure. The resulting structure may be viewed as consisting of a small-stranded loop along with regions of regularly spaced double helical turns with a geometry similar to relaxed B-DNA. If, however, hydrogen bonds are not disrupted, the circular DNA will twist in a direction opposite to the one in which it was rotated initially in order to relieve the strain induced by the unwinding. Thus the rotational strain that was introduced before the circularization of DNA can be accommodated either by the disruption of H bonding or by the formation of tertiary structures with visible supercoils (Figure 14.25). These two representations of the negative superhelix should be viewed as two manifestations of the same underlying phenomenon. In general, a dynamically imposed compromise, determined by the environment and the status of circular DNA, is reached between hydrogen-bond disruption and supertwisting. In practice, this means that supercoiled DNA may consist of twisted structures with enhanced tendency to generate regions with disrupted hydrogen bonding (bubbles).

In a circular DNA that is initially relaxed, the transient strand unwinding would tend to introduce compensating supertwists. However, if DNA is superhelical to begin with, the density of the superhelix will obviously tend to fluctuate with the "breathing" of the helix. All naturally occurring DNA molecules contain a deficit of helical turns; that is, they exist as negative superhelices with a superhelical density that remains remarkably constant among different DNAs. Normally one negative twist is found for every 20 turns of the helix.

If one of the terminals of the linear polynucleotides is rotated in the direction of **overwinding** rather than **unwinding** the double helix, the resulting DNA will contain positive superhelices. While negatively superhelical DNA can accommodate unwinding stress either by unwinding (accompanied by the interruption of hydrogen bonds) or by formation of negative superhelices, the only available option for overwound DNA is to accommodate the stress by acquiring positive superhelices. Positive supercoils can be generated by specialized enzymes, the **topoisomerases,** and may be present transiently **in vivo** but are rarely present in cellular DNA.

Positive and negative supercoils can, in principle, coexist transiently within the same DNA molecule. Yet the DNA molecule, in an overall sense, may be viewed as relaxed because it may return to a relaxed state without the breaking of phosphodiester bonds. A rubber band, which in its normal unstrained form

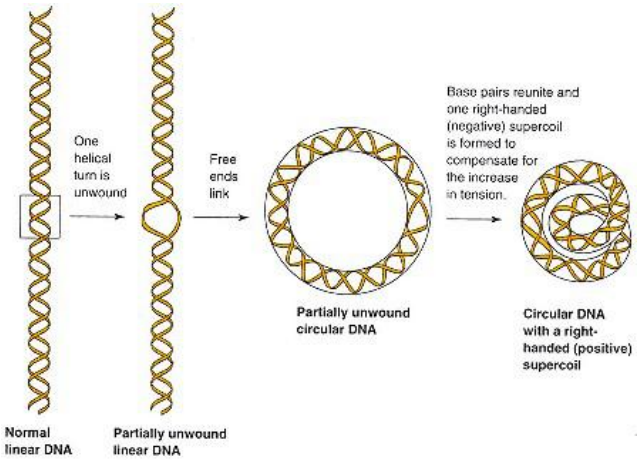


Figure 14.25
Right-handed (negative) DNA supercoiling.
Right-handed supercoils (negatively supercoiled DNA) are formed if relaxed DNA is partially unwound. Unwinding may lead to a disruption of hydrogen bonds or alternatively produce negative supercoils. The negative supercoils are formed to compensate for the increase in tension that is generated when disrupted base pairs are reformed.
Redrawn from Darnell, J., Lodish, H., and Baltimore, D. *Molecular Cell Biology*. New York: Freeman, W. H. 1986.

might be visualized as a circular relaxed structure (without supercoils), can be used as such a model (Figure 14.26). Grasping this band firmly at opposite sides and twisting one side of the band generates a structure characterized by two topological domains, with twisting of opposite handedness, that are clearly visible when the two sides are pulled apart. If the opposite sides are brought back close together, each domain becomes supertwisted; that is, each domain generates a supercoil. This requires an input of energy since the supertwisted state does not represent the low-energy state of the rubber band. When the band is released from the grasping that restrains rotation, it may return to its original relaxed configuration. During these manipulations, the physical structure of the band has remained intact. A difference between the rubber band model and cellular DNA is that the latter exists almost exclusively in supercoiled form. Cellular DNA can be described on the basis of the linking number of DNA, L, an integer number defined as the number of times one strand appears to cross over the other when the DNA structure is projected onto a flat surface (Figure 14.27). Examination of Figure 14.27a further indicates that the linking number of relaxed DNA (B-DNA), L_0, can be defined as

$$L_o = \frac{N}{10.5}$$

where N is the number of base pairs and 10.5 refers to the **average helical repeat,** that is, the number of base pairs per one complete turn of the helix.

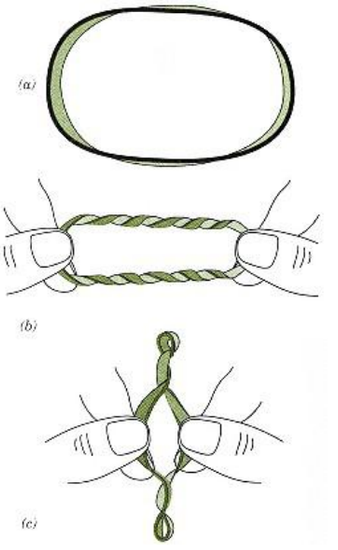


Figure
14.26
Superhelical model for DNA.
A rubber band represents
the topological properties of double-stranded circular
DNA. The relaxed form of the band, shown in (a), has
been twisted to generate two distinct domains, separated
by the pair of "thumb–forefinger anchors," as shown in
(b). Left-handed (counterclockwise) turns have been
introduced into the upper section of the band, with
compensating right-handed (clockwise) turns present
into the bottom section. When the "anchors" are
brought into close proximity with each other as
shown in (*c*), the upper section that contained the
left-handed turns forms a right-handed superhelix.
The bottom section produces a left-handed superhelix.
Clearly, superhelicity is not the property of a DNA
molecule as a whole but rather a property of specific
DNA domains.
Redrawn from Sinden, R. R., and Wells,
R. D. DNA structure, mutations, and human genetic
disease. *Curr. Opin. Biotech.* 3:612, 1992.

Geometric Description of Superhelical DNA

Conformations acquired by interlocking rings of a closed circular DNA can formally be characterized by three parameters: **linking number,** L, **number of helical turns or twist,** T, and **number of supercoils or writhing number,** W. These parameters are related by the equation $L = T + W$. The nature of T and W is self-explanatory.

Two important conclusions can be reached from consideration of these definitions and from examination of Figure 14.28. First, it is apparent that for

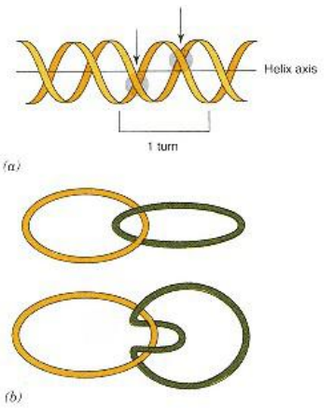


Figure 14.27
Determination of the linking number *L* in superhelical DNA.
(a) Side view of a schematic representation of the
double helix. Note that the strands cross twice
for each turn of the helix.
(b) DNA circles interwound
once and twice. Note that each pair of crossings is
equivalent to one interwind.

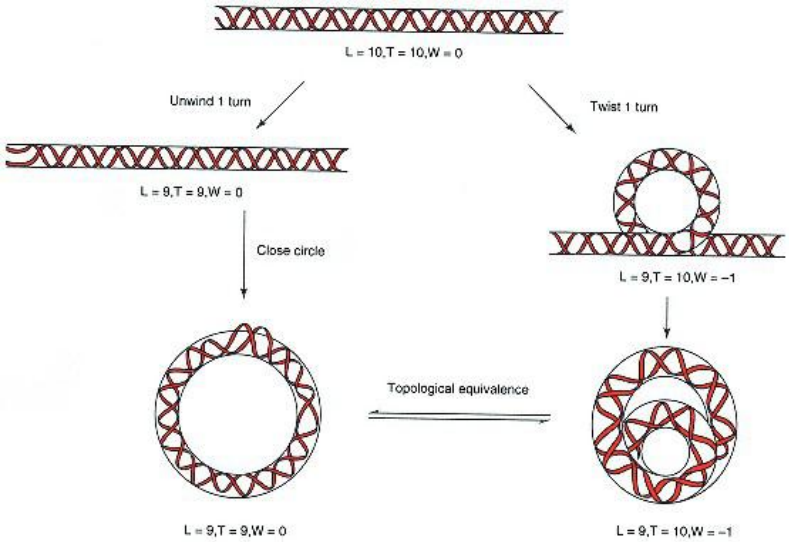


Figure 14.28
Various types of DNA superhelices.
An accurate representation of superhelical DNA structures can be made, using the number of helical turns or twists, *T*, and the number of supercoils or writhing number, *W*, along with a third parameter the linking number, *L*, as defined in the text. The figure shows ways of introducing one supercoil into a DNA segment of 10 duplex turns and the parameters of the resulting superhelices.
Redrawn with permission from Cantor, C. R., and Schimmel, P. R., *Biophysical Chemistry, Part III.* San Francisco: Freeman, 1980.

every relaxed DNA the linking number *L* and the number of helical turns *T* are identical. However, as will be apparent shortly, the reverse is not true. Second, DNAs with a specific linking number can acquire various different topological conformations. Different types of supertwists (*T*) may be formed. However, all conformations with the same linking number are interconvertible without breaking any covalent bonds. Therefore linking number is a constant for any covalently closed circular DNA.

Various forms of supercoiled DNAs can be described using *L*, *T*, and *W* numbers. The mental exercise shown in Figure 14.28 illustrates how these numbers apply. It should be recalled that the turns of the typical double helix are right handed. Therefore, if a hypothetical linear DNA duplex that is 10 turns long ($L = 10$ and $T = 10$) is unwound by, say, one turn, the resulting structure will have the following characteristics: $L = 9$ and $T = 9$. A potentially equivalent structure can be formed if instead ends of the same hypothetical DNA are secured so that they cannot rotate and the molecule is looped in a counterclockwise manner. Since in this case untwisting is not permitted to occur, the number of helical turns remains unchanged; that is, $T = 10$. However, as a result of "looping" operations, linking number is now reduced by 1; that is, $L = 9$. The structure resulting from this deliberate introduction of a loop is visibly superhelical. Furthermore, application of the equation that relates values of *L*, *T*, and *W* indicates that *W* must be equal to –1; that is, the structure is a negative superhelix with one supertwist.

The two structures described above—$L = 9$, $T = 9$, $W = 0$ and $L = 9$, $T = 10$, and $W = -1$—obviously have the same linking number and are therefore interconvertible without the disruption of any phosphodiester bonds. The potential equivalence of these two types of structure becomes more apparent when ends of polynucleotides in each structure are joined into a circle without strands being allowed to rotate. Circularization produces an **interwound** circular structure (a number 8-shaped structure referred to as a **plectonemic** coil) or a doughnut-shaped superhelical arrangement referred to as a **toroidal turn,** both of which are freely interconvertible. An interwound turn, shown in Figure 14.29, can be produced by unfolding a toroidal turn along an axis that is distinct from the supercoil axis.

In summary, if the termini of a linear DNA molecule are covalently attached, a "relaxed" covalent circle results. However, if one end of the double helix is maintained in a fixed and stationary position while the other end is rotated in either direction prior to closing the circle, the resulting structure will twist in the opposite direction so as to generate a supertwisted helical structure. For each additional complete turn of the helix, DNA will acquire one more superhelical twist in the opposite direction of rotation in order to relieve intensifying strain. As a result, topologically equivalent structures, such as those shown in Figure 14.28, will be created. A real superhelical DNA exists as an equilibrium among these forms and many other intermediate arrangements in space that have the same linking number but different numbers of helical turns and supertwists. Although linking number is a constant and an integer, the number of twists can change in positive and negative increments, which are compensated by negative and positive changes in the writhing number. DNA supercoils are distributed in part as mixtures of interwound (plectonemic) and toroidal coils and as decreases in twist angle of the double helix. The interwound form is by far the more predominant structure for supercoiled DNA. In solution about 70% of the deficiency in linking numbers may be distributed as writhe change and 30% as changes in twist.

Although the closed circular form of DNA is an ideal candidate for acquiring a superhelical structure, any segment of double-stranded DNA that is in some way immobilized at both of its terminals qualifies for superhelicity. This property therefore is not the exclusive province of circular DNA. Rather, any appropriately anchored linear DNA molecule can acquire a superhelical conformation. The DNA of animal cells, for instance, normally associated with nuclear proteins, falls into this category. Animal DNA can acquire a superhelical form because its association with nuclear proteins creates numerous closed topological domains. A topological domain is defined as a DNA segment contained in a manner that restrains rotation of the double helix. In addition, circular DNAs of most bacterial phages, animal viruses, bacterial plasmids, and cell organelles, such as mitochondria and chloroplasts, contain superhelical DNAs. Existence

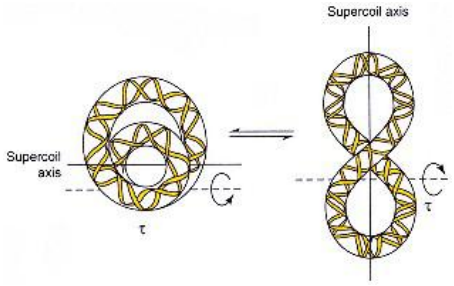


Figure 14.29
Equilibrium between two equivalent supercoiled forms of DNA.
The forms shown are freely interconvertible by unfolding the doughnut-shaped toroidal form along an axis parallel to the supercoil axis or by folding the number 8-shaped interwound form along an axis perpendicular to the supercoil axis. The two forms have the same *W, T*, and *L* numbers.
Redrawn with permission from Cantor, C. R., and Schimmel, P. R. *Biophysical Chemistry, Part III.* San Francisco: Freeman, 1980.

of negative superhelicity appears to be an important factor, promoting packaging of DNA within the confines of the cell because supercoils facilitate formation of compact structures. For instance, while the length of DNA in each human chromosome is of the order of centimeters, condensed mitotic chromosomes that contain this DNA are only a few nanometers long. Negative superhelicity may also be instrumental in facilitating the process of localized DNA strand separation during DNA repair, synthesis, and recombination.

Topoisomerases

Specific enzymes known as **topoisomerases** appear to regulate the formation of superhelices. These enzymes change the linking number, *L*, of DNA. Topoisomerases act by catalyzing the concerted breakage and rejoining of DNA strands, which produces a DNA that is more or less superhelical than the original DNA. Topoisomerases are classified into type I, which break only one strand, and type II, which break both strands of DNA simultaneously. **Topoisomerases I** act by making a transient single-strand break in a supercoiled DNA duplex, which changes the linking number by increments of 1 and results in relaxation of the supercoiled DNA (Figure 14.30). **Topoisomerases II** act by binding to a DNA molecule in a manner that generates two supercoiled loops, as shown in step 1 of Figure 14.31 Since one of these loops is positive and the other negative, and there is no disruption of phosphodiester bonds, the overall linking number of the DNA remains unchanged. In subsequent steps, however, the enzyme nicks both strands and passes one DNA segment through this break before resealing it. This manipulation inverts the sign of the positive supercoil, resulting in the introduction of two negative supercoils in each catalytic step and the changing of the linking number in increments of 2. This reaction occurs at the expense of ATP; that is, topoisomerases II are ATPases. Several well-studied topoisomerases are listed in Table 14.6.

Although all type II topisomerases can change the linking number of DNA, their individual properties vary considerably. A subset of type II topoisomerases (the **gyrases**—isolated from bacteria) are the only enzymes that introduce negative supercoils into relaxed DNA. Analogous eukaryotic topoisomerases have not been found. Apparently eukaryotes use alternative approaches for the

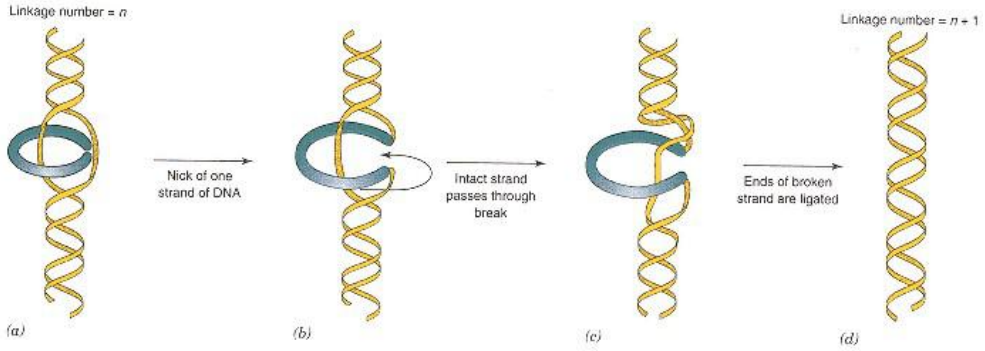


Figure 14.30
Mechanism of action of topoisomerases I.
Topoisomerases I can
(a) relax DNA by (a) first binding to it and locally separating the complementary polynucleotide strands; subsequently
(b) nick one of the strands;
(c) bind to the newly generated termini and prevent these termini from rotating freely; and
(d) ligate the intact strand through the gap generated
by the nick, close the gap by restoring the phosphodiester bond, and give rise to a relaxed structure.
Redrawn from Dean, F., et al. *Cold Spring Harbor Symp. Quant. Biol.* 47:773, 1982.

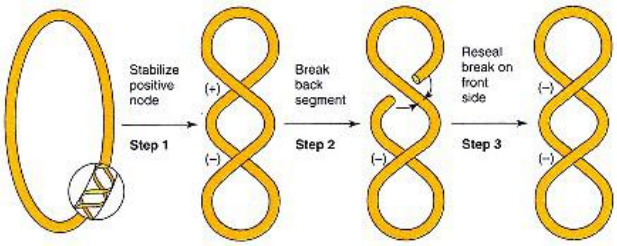


Figure 14.31
Mechanism of action of topoisomerases II.
Topoisomerases II (and gyrase) change the linking number of DNA by binding to a DNA molecule and passing one DNA segment through a reversible break formed at a different segment of the same DNA molecule. The mechanism of action of gyrase is illustrated above using as an example the conversion of a relaxed DNA molecule to a molecule that contains first two supercoils, one positive and one negative (step 1). Passage of a DNA segment through the positive supercoil shown on the right most part of the figure (step 3) changes the linking number, producing a molecule that contains two negative supercoils.
Redrawn with permission from Brown, P. O., and Cozzarelli, N. R. *Science* 206:1081, 1979.

introduction of negative supercoils into DNA. The wrapping of DNA around chromosomal proteins followed by the action of eukaryotic topoisomerases that relax DNA may be used by eukaryotes for the generation of negative supercoiling. Bacterial type III topoisomerases are a class of topoisomerases with type I topoisomerase properties; that is, they can relax supercoils without the requirement of an energy source, such as ATP hydrolysis. These topoisomerases may specialize in the resolution of circular DNA products (catenates) that are generated just prior to the completion of DNA replication. An unusual class of topoisomerases, **reverse gyrases,** have been isolated from various species of archaebacteria. Remarkably, these gyrases introduce **positive supercoils** into DNA. Positive supercoiling may protect DNA from the denaturing conditions of high temperature and acidity under which these bacteria "exist."

TABLE 14.6 Properties of DNA Topoisomerases

Enzyme	***Type***[a]	***ΔL***	***Activities***
E. coli topoisomerase I (*top A*)[b]	I	Increase *L* *L* = 1	Relaxes negatively supercoiled DNA
Eukaryotic topoisomerase I from yeast (*top 1*)	I	Increase or decrease *L* *L* = ±1	Relaxes either positively or negatively supercoiled DNA
E. coli topoisomerase II or DNA gyrase (*gyrA, gyrB*)	II	Increase or decrease *L* *L* = ±2	Introduces negative supercoiling to DNA; relaxes either positively or negatively supercoiled DNA
E. coli topoisomerase IV (*parC, parE*)	II	Increase *L* *L* = +2	DNA relaxing activity; it cannot introduce negative supercoils
Eukaryotic topoisomerase II from yeast (*top 2*)	II	Increase or decrease *L* *L* = ±2	Relaxes positively or negatively supercoiled DNA
E. coli topoisomerase III (*top B*)	I	Increase *L* *L* = +1	Relaxes negatively supercoiled DNA; decatenation activity
Eukaryotic topoisomerase III (*top 3*)	I	Increase *L* *L* = +1	Specific activity on DNA with single-stranded heteroduplex

[a] Type I topoisomerases use Mg^{2+} as cofactor but do not use ATP. Type II topoisomerases require Mg^{2+} plus ATP.

[b] The name of the gene coding for the topoisomerase is shown in parentheses.

Apparently, the energy released by ATP hydrolysis is used for restoring topoisomerase II conformation, after the enzyme has catalyzed the formation of 1 mol equiv of product. The reaction is inhibited by the antibiotics **nalidixic acid** and **novobiocin.** Derivatives of nalidixic acid are used clinically in the treatment of infections caused by bacteria resistant to other more commonly used antibiotics. Various compounds that inhibit topoisomerases are also effective antitumor agents (see Clin. Corr. 14.3).

During the reaction, topoisomerases remain bound to DNA by a covalent bond between a tyrosyl residue and a phosphoryl group at the incision site (a 5 -phosphotyrosine bond). This enzyme–polynucleotide bond conserves the energy of the interrupted phosphodiester bond for the subsequent repair of the nick. The cleavage sites do not consist of unique nucleotide sequences, although certain sequences are preferentially found at cleavage sites. Gyrase, isolated from *E. coli,* is a tetrameric protein consisting of two A subunits and two B subunits. It adds negative supercoils to DNA at a rate of about 100 per minute. Topoisomerases regulate the level of supercoiling. In *E. coli* DNA such regulation requires the involvement of both gyrase and topoisomerase I activities. The balance between these two opposing enzymic activities keeps DNA at a precisely regulated cellular level of superhelicity. The ATP to ADP ratio may play a role in this process, since this ratio influences the activity of gyrase.

Other biological reactions involving DNA require participation of topoisomerases. For example, topoisomerase IV, a type II topoisomerase, may be essential for separating two circular chromosomes that become entangled by **catenation** toward the end of replication. Also, topoisomerases are involved in relaxing

CLINICAL CORRELATION 14.3

Topoisomerases in Treatment of Cancer

Topoisomerases are emerging as important targets of both antimicrobial and antineoplastic agents including camptothecin, anthracycline, and amino-acridine. These agents share a common principal mechanism of action by interfering with the enzyme-catalyzed rejoining of DNA strands, in effect inhibiting only one of the two substeps in the mechanism of action of topoisomerases. Therefore topoisomerase drugs do not act by inhibiting the overall activity of the enzyme, as is the case with most enzyme-targeting drugs. Instead, they convert topoisomerases into "DNA-breaking agents." The DNA degradation that follows leads to cell death.

Both topoisomerases I and II can be targeted with therapeutic results. Camptothesin and its derivatives modify the function of topoisomerase I. An excellent correlation has been noted between antitumor activity of various camptothecin derivatives on murine leukemia and their interference with topoisomerase activity. Camptothecins may cause potentially lethal lesions in cells in the form of drug-stabilized covalent DNA cleavage complexes. Subsequent DNA replication may be a prerequisite for cell toxicity. Increased levels of topoisomerase I found in advanced stages of colon cancer and several other human malignancies may contribute to the therapeutic efficacy of 9-amino-20(*RS*) and 10,11-methylenedioxy-20(*RS*), two camptothecin derivatives. In clinical trials these camptothecins appear to induce long-term remissions from single-agent treatment of colon cancer xenografts.

Studies with two other potent antineoplastic agents—an acridine derivative, 4 (9-acridinylamino)methanesulfon-*m*-anisidide (m-AMSA), and epipodophyllotoxin toposide—that act selectively on topoisomerases II indicate that these clinically useful drugs stabilize covalent topoisomerase II–DNA cleavage complexes by interfering with the enzyme-mediated DNA religation reaction. Indirect evidence also suggests that these drugs may stimulate formation of these complexes. Contrary to observations regarding the importance of DNA replication in the expression of the cytotoxic effect of drugs that target topoisomerase I, topoisomerase II-mediated DNA breaks can exert their cytotoxic effect in the absence of ongoing DNA synthesis. Instead, the lethal lesions induced by topoisomerase II-targeted drugs may be dependent on recombinations and mutations at sites of formation of drug-induced topoisomerase II–DNA complexes. Many anticancer agents including anthracyclines (including adriamycin and doxorubicin), synthetic intercalators, ellipticines, and podophyllotoxins exert their therapeutic effects on topoisomerases II. Hematologic neoplasms, such as lymphoid and nonlymphoid leukemias, high-grade non-Hodgkin's lymphomas, and Hodgkin's disease, are treated mostly with combinations of one or more topoisomerase II inhibitors with or without additional cytotoxic agents.

Potmesil, M., and Kohn, K. W. (Eds.). *DNA Topoisomerases in Cancer*. New York: Oxford University Press, 1991; and Ellis, A. L., Nowak, B., Plunkett, W., and Zwelling, L. A. Quantification of topoisomerase–DNA complexes in leukemia cells from patients undergoing therapy with a topoisomerase directed agent. *Cancer Chemother. Pharmacol.* 34:249, 1994.

the superhelical tension generated by the separation of DNA strands during the process of transcription.

Separation of superhelical DNA from the relaxed or linear forms can be achieved by gel electrophoresis or by equilibrium centrifugation. With the latter method separation is achieved because the density of supercoiled DNA differs from that of the relaxed forms.

Alternative DNA Conformations

Conformational variants of DNA—that is, A-, B-, and Z-DNA—are associated mainly with variation in the conformation of the nucleotide constituents of DNA. It is now recognized that DNA is not a straight, stable, monotonous, and uniform structure. Instead, DNA forms unusual structures such as **cruciforms** or **triple-stranded** arrangements and bends as it interacts with certain proteins. Such variations in DNA conformation appear to be an important recurring theme in the process of molecular recognition of DNA by proteins and enzymes. Variations in DNA structure or conformation are favored by specific motifs in the sequence of DNA referred to as **defined, ordered sequence DNA** and are abbreviated as **dos DNA.** They include such DNA elements as **inverted repeats, mirror repeats, direct repeats, homopurine–homopyrimidine sequences, phased A tracts,** and **G-rich regions.** AT-rich sequences prone to easy strand separation exist near the origins of DNA replication. The human genome is rich in homopurine–homopyrimidine sequences and alternating purine–pyrimidine tracts. DNA bending, slipped DNA, cruciform formation, triplex DNA, and **quadruplex** arrangements are among the structures reviewed in this section.

DNA Bending

DNA sequences with runs of 4 to 6 A bases phased by 10-bp spacers produce bend conformations. **DNA bending** appears to be a fundamental element in the interaction between DNA sequences and proteins that catalyze central processes, such as replication, transcription, and site-specific recombination. Bending induced by interactions of DNA with enzymes and other proteins, such as histones, does not require the exacting nucleotide sequence conditions that are needed for bending of protein-free DNA. Bending also occurs because of photochemical damage and serves as a recognition signal for the initiation of DNA repair. Contrary to the bending effect generated by phased A tracts, the presence of poly A tracts without spacers or the presence of certain arrangements of polypurine–polypyrimidine tracts may generate a DNA, known as **anisomorphic DNA,** that is less flexible than usual.

Cruciform DNA

Dos DNA is generally present within noncoding DNA regions and it consists of various symmetry elements, including **inverted repeats,** completely **symmetrical inverted repeats,** known as *palindromes* (see p. 610), **mirror repeats,** and **direct repeats** as shown in Figure 14.32. Base pairing can be disrupted and conformational variants of DNA such as **junctions, cruciforms, triplex,** and **quadruplex DNA** and **slipped mispaired** structures can be formed within the **dos** sequences.

The biological function of cruciforms has not been generally established. Inverted repeats are quite widespread within the human genome and are often found near putative control regions of genes or at origins of DNA replication. It is therefore speculated that inverted repeats may function as molecular switches for replication and transcription. In fact, in a few instances there is evidence to support the involvement of cruciforms in the control of replication and transcription. The disruption of H bonds between the complementary strands and the formation of intrastrand H bonds within the region of the inverted

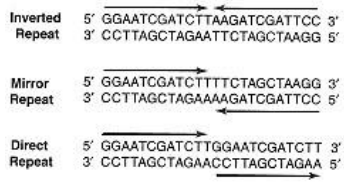


Figure 14.32
Symmetry elements of DNA sequences.
Three types of symmetry elements for double-stranded DNA sequences are shown. Arrows illustrate the special relationship of these elements in each one of these sequences. In inverted repeats, also referred to as palindromes, each single DNA strand is self-complementary within the inverted region that contains the symmetry elements. A mirror repeat is characterized by the presence of identical base pairs equidistant from a center of symmetry within the DNA segment. Direct repeats are regions of DNA in which a particular sequence is repeated. The repeats need not be adjacent to one another.

repeat produce a cruciform structure (Figure 14.33). The loops generated by cruciform formation require the unstacking of 3–4 unpaired bases at the end of the "hairpin" and therefore cruciform formation requires the expenditure of cellular energy.

Triple-Stranded DNA

Many sequences in the human genome, especially in regions involved in gene regulation, have the potential to form triple-stranded DNA structures. Such

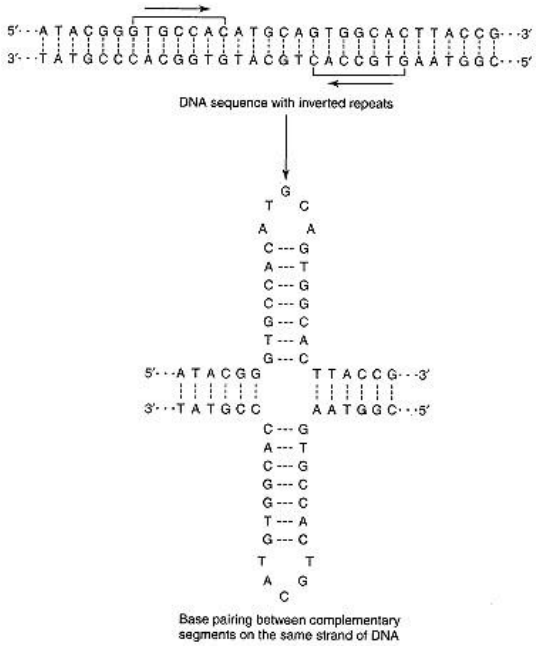


Figure 14.33
Formation of cruciform structures in DNA.
The existence of inverted repeats in double-stranded DNA is a necessary but not a sufficient condition for the formation of cruciform structures. In relaxed DNA, cruciforms are not likely to form because the linear DNA accommodates more hydrogen-bonded stacked base pairs than the cruciform structure, making the formation of the latter thermodynamically unfavored. Unwinding is followed by intrastrand hydrogen bond formation between the two symmetrical parts of the repeat to produce the cruciform structure. Formation of cruciform structures is not favored over DNA regions that consist of mirror repeats because such cruciforms would be constructed from parallel rather than antiparallel DNA strands. Instead, certain mirror repeats tend to form triple helices.

structures can be formed either within the same DNA structure (i.e., intramolecularly) or between DNA and a distinct or second polynucleotide (i.e., intermolecularly). In either case, **triple-stranded DNA** structures are formed, with few exceptions, in DNA regions characterized by the presence of a continuous string of purine bases, that is, **homopurine–homopyrimidine regions.** Such regions occur with frequencies much higher than expected from probability considerations alone. Polypurine tracts over 25 nucleotides long constitute as much as 0.5% of some eukaryotic genomes. Polypurine–polypyrimidine regions appear to have a multiplicity of potential biological roles, including possible effects in transcription control, in the initiation of replication, as replication terminators, as enhancers of stability at the ends of chromosomes (telomeres), and as initiators of genetic recombination.

Triple-stranded DNA is generated by the hydrogen bonding of a third strand into the major groove of B-DNA (Figure 14.34). Since base pairs are already formed in the B-DNA, the third strand forms hydrogen bonds with another surface of the double helix through so-called **Hoogsteen pairs.** The options available for the formation of a triple-stranded structure are limited to only four triplet bases—TAT, CGC, GGC, and AAT. The structure of two of these triplets is shown in Figure 14.35. Since pyrimidine does not have two H-bonding surfaces with more than one H bond, it follows that the central strand of the triplex must always be purine rich. Therefore, in practice, intermolecular triple-stranded DNA can only form within homopurine–homopyrimidine regions of DNA. Just as is the case for the Watson–Crick base pairs, formed between strands in double-stranded DNA, a polypurine–polypyrimidine region defines a unique third strand pairing sequence. Consequently, the sequence of a third strand can be designed so that it can form Hoogsteen base pairs with any specific polypurine–polypyrimidine region of DNA.

Figure 14.34
Structure of intermolecular triple helices.
Triple helices can form among
(a) two polypurine strands and one polypyrimidine
strand as exemplified by the polyG-polyG-polyC triplet or
(b) among two polypyrimidine and one polypurine
strand as in the case of the polyT-polyA-polyT triplet. In
(*a*), held together partially by Hoogsteen base pairing,
the polypurine strand is antiparallel to the polypurine strand of
the original DNA duplex. In (*b*), which is characterized by
reverse Hoogsteen base pairing, the polypyrimidine third strand is
parallel to the polypurine strand. Brackets enclose strands held
together by Watson–Crick hydrogen bonding.
Redrawn based on figure in Sinden, R. R. *DNA Structure and Function*. New York: Academic Press, 1994.

Intramolecular triple helices can be formed by disruption of H bonds, over regions of DNA characterized by the presence of polypurine strands, and refolding as illustrated in Figure 14.36 to generate a triple-stranded region and a single-stranded loop. This arrangement involves disruption of base stacking interaction in the unpaired region and therefore it is not the most thermodynamically stable structure that can be formed by the double-stranded polypurine–polypyrimidine DNA segment. Yet, intramolecular triple helices are detected in cellular DNA. Apparently DNA supercoiling provides the energy to drive the unwinding of DNA that is necessary for the formation of the triple helix. Triplestrand formation produces a relaxation of negative supercoils. In addition to the general requirement that a string of purines be present, structural considerations for the formation of hydrogen bonds dictate that the polypurine–polypyrimidine region must contain **mirror repeat symmetry** for the triplex to form. A mirror repeat is a region such as AGGGGA that has the same base sequence when read, from a central point, in either direction within one of the DNA strands. There are two possible pairs of alternative structures that can form from different foldings of the polypurine–polypyrimidine region in the triple helix. One of the pairs is characterized by a pyrimidine–purine–pyrimidine arrangement in which half of the pyrimidine strand is paired as the third strand and the complementary strand remains unpaired. The other pair of possible alternative structures is characterized by the less commonly occurring purine–purine–pyrimidine arrangement.

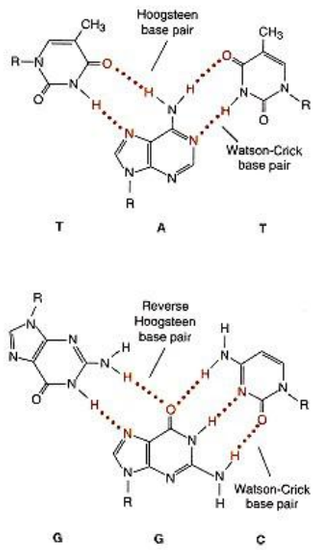


Figure 14.35
Base pairing in DNA triplexes.
Two examples of the type of hydrogen bonding involving the formation of triple-stranded DNA helices are shown, one for the polyG-polyG-polyC and one for the polyT-polyA-polyT triple helix. For the T-A-T triplex the purine (A) participates in a Watson–Crick base pairing to T and in an alternative type of base pairing (Hoogsteen base pairing) to a second T. In the G-G-C triplex, the purine (G) forms a Watson–Crick base pairing with C and a Hoogsteen base pairing with G. In this base pairing scheme the ribose groups of the two purines are in trans orientation, generating a so-called reverse Hoogsteen base pair. The relative orientation (polarity) of the three strands shown in Figure 14.36 depends on whether two of the participating polynucleotides form regular or reverse Hoogsteen base pairs.

A distinct type of intermolecular triple-stranded helix is formed by enzymatic catalysis, as an intermediate during general recombination. These intermediates are atypical triple helices in that they are not limited to polypurine–polypyrimidine regions but instead involve DNA strands of identical, or nearly identical, nucleotide sequences. These helices are unwound structures in which

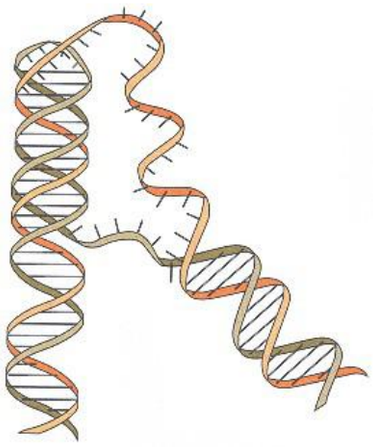


Figure 14.36
Intramolecular triple helices.
Polypurine–polypyrimidine regions of DNA with a mirror repeat symmetry can form an intramolecular triple helix in which the third strand lays in the major groove, whereas its complementary strand acquires a single-stranded conformation.
Redrawn based on figure in Sinden, R. R. *DNA Structure and Function*. New York: Academic Press, 1994.

the third strand binds on the major groove side of a double helix in a manner parallel to its identical strand.

Long polypurine–polypyrimidine sequences can form another variant DNA structure, the **nodule DNA,** that consists of a pair of two intermolecular triplex regions, as illustrated in Figure 14.37. Its biological significance has not been determined.

The role of DNA triplex formation in a hereditary affliction known as persistence of fetal hemoglobin is briefly reviewed in Clin. Corr. 14.4. The therapeutic potential of oligonucleotides capable of forming triplex DNA with segments of DNA having Hoogsteen base pairing potential is discussed in Clin. Corr. 14.5.

Four-Stranded DNA

Four-stranded DNA (quadruplex) can form as both parallel and antiparallel structures. Parallel structures may form during DNA recombination (see p. 661). A parallel four-stranded DNA may be found in an immunoglobulin heavy chain gene. The immunoglobulin genes undergo a type of recombination (specific recombination) that is responsible for the extensive diversity that characterizes antibody formation. The sequences that participate in this alternative type of DNA structure are repeated motifs high in guanine content such as GGGGAGCTGGG. A base pairing scheme for parallel four-stranded DNA, referred to as a **G-quartet DNA,** is shown in Figure 14.38. In this scheme all four DNA strands are arranged in a parallel orientation and are associated to one another through Hoogsteen base pairs. The glycosidic bonds in all nucleotides are in the **anti** configuration.

Parallel and antiparallel four-stranded DNA structures form at **telomeres.** These contain repetitive simple oligonucleotide sequences (such as G_4T_2) that are usually purine rich in one of the strands. This strand is longer and overhangs the complementary strand. The repetitive sequences make the formation of four-stranded DNA possible. One such four-stranded antiparallel structure forms when the single strand overhanging the telomere end is folded back into a hairpin structure with guanines binding to one another by Hoogsteen base pairing. Two folded double-helical regions can then interact to form four-

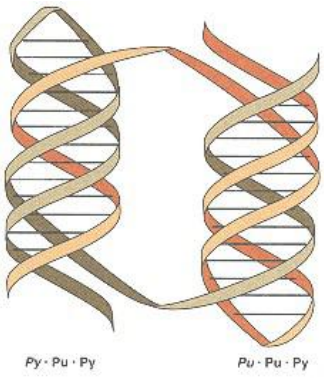


Figure 14.37
Nodule DNA. Nodule DNA consisting of a combination of a Py-Pu-Py triple helix and a Pu-Pu-Py triplex can be formed within a long polypurine–polypyrimidine tract.
The Py-Pu-Py structure can contribute its displaced single Pu strand to the other half of the Pu-Py region, forming the Pu-Pu-Py triplex structure.
Redrawn based on figure in Sinden, R. R. *DNA Structure and Function.* New York: Academic Press, 1994.

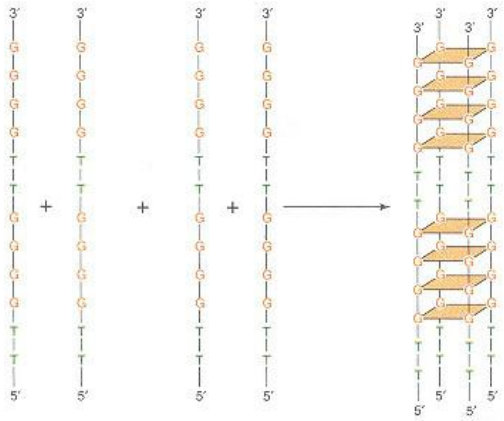


Figure 14.38
Parallel quadruplex DNA.
Quadruplex structures in which all four strands are parallel can form from four single-strand tracts of polyguanine. These quadruplexes, referred to as G-quartets, are associated by Hoogsteen base pairs.
Redrawn based on figure in Sinden, R. R. *DNA Structure and Function.* New York: Academic Press, 1994.

CLINICAL CORRELATION 14.4

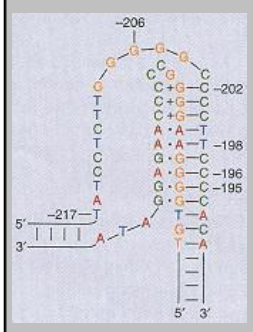
Hereditary Persistence of Fetal Hemoglobin

Hereditary persistence of fetal hemoglobin (HPFH) is a group of conditions in which fetal hemoglobin synthesis is not turned off with development but continues into adulthood. The homozygous form of the disease is extremely uncommon, being characterized by red blood cell changes similar to those found in heterozygous β-thalassemia. HPFH, in either the homozygous or heterozygous state, is associated with mild clinical or hematologic abnormalities. Mild musculoskeletal pains may occur infrequently but HPFH patients are generally asymptomatic.

The disease is the result of failure in control of transcription from human Gγ- and Aγ-globin genes. Affected chromosomes fail to switch from γ- to β-chain synthesis. Expression of these genes appears to be affected substantially by formation of an intramolecular DNA triplex structure located about 200 bp upstream from the initiation site for transcription of genes, specifically between positions –194 and –215.

Hemoglobin genes of patients contain mutations in positions –195, –196, –198, and –202. Mutations at –202 involve changes from C to G and C to T, at –198 from T to C, at –196 from C to T, and at –195 from C to G. These mutations influence the stability of the intramolecular DNA triple helix.

In general, the presence of polypurine–polypyrimidine sequences sufficiently long to form intramolecular triple helices tends to repress transcription, while short polypurine–polypyrimidine segments that are unable to induce triple helix formation have no effect on transcription. In the case of HPFH, a remarkable correspondence is noted between base changes that destabilize formation of the triple helix and presence of the genetic disease.

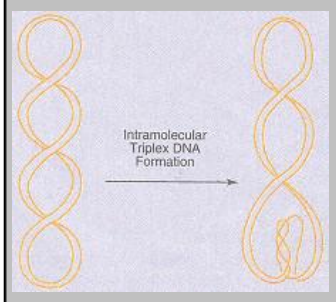


Ulrich, M. J., Gray, W. J., and Ley, T. J. An intramolecular DNA triplex is disrupted by point mutations associated with hereditary persistence of fetal hemoglobin. *J. Biol. Chem.* 267:18649, 1992; and Bacolla, A., Ulrich, M. J., Larson, J. E., Ley, T. J., and Wells, R. D. An intramolecular triplex in the human gamma-globin 5 -flanking region is altered by point mutations associated with hereditary persistence of fetal hemoglobin. *J. Biol. Chem.* 270:24556, 1995.

CLINICAL CORRELATION 14.5

Therapeutic Potential of Triplex DNA Formation

Control regions of genes often contain polypurine–polypyrimidine regions. Binding of a single third strand of DNA, complementary to the polypurine strand, may under certain conditions prevent binding of regulatory proteins, such as transcription factors, and thus affect gene expression. Alternatively, triplex formation may influence regulation of gene expression by affecting the level of DNA supercoiling in the topological domain in which the triple helix forms, as shown in the figure below.



For instance, a polypurine–polypyrimidine region, which can form a triplex, is present upstream of the human *c-myc* oncogene. This region, which interacts with transcription factors, can form an intermolecular triplex with an oligonucleotide designed to provide base complementary with the polypurine–polypyrimidine region. Formation of triplex DNA results in inhibition of *c-myc* transcription *in vitro*. The above example suggests that formation of intermolecular complexes has the potential to regulate expression of specific proteins that may play important roles in health and development of disease. The great individuality inherent in the sequence of unique oligonucleotide segments provides the potential to design specific therapeutic oligonucleotides for turning certain genes off and on.

The specificity of DNA triplexes also provides another approach for the potential control of expression of certain genes. Oligonucleotide sequences that are targeted to specific regions of a eukaryotic or viral genome can be coupled with artificial nucleases or covalent modifiers of DNA. Such targeting produces endonucleolytic cutting or covalent modification of the DNA at specific sites. This approach has therapeutic potential for gene regulation and killing of virus-infected cells or other abnormal cells.

Kinniburg, A. J. A cis-acting transcription element of the *c-myc* gene can assume an H-DNA conformation. *Nucleic Acids Res.* 17:7771, 1989; and Pei, D., Corey, D. R., and Schultz, P. G. Site specific cleavage of duplex DNA by a semisynthetic nuclease via triple-helix formation. *Proc. Natl. Acad. Sci. USA* 87:9858, 1990.

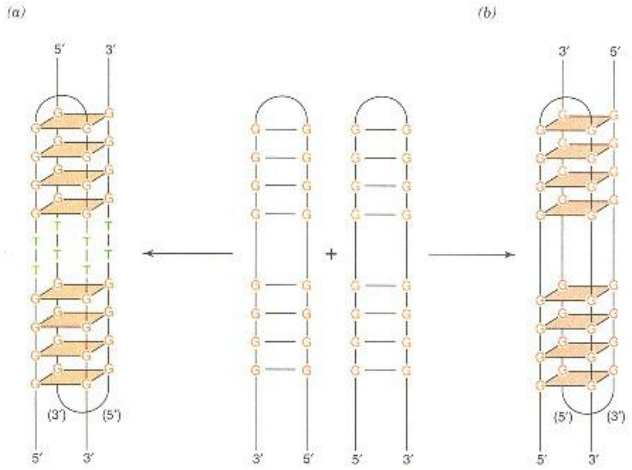


Figure 14.39
Antiparallel quadruplex DNA.
Several quadruplexes both of the antiparallel
(a) and parallel
(b) type can form at telomeres as these terminal regions are guanine-rich.
Redrawn based on figure in Sinden,
R. R. *DNA Structure and Function*. New York: Academic Press, 1994.

stranded structures held together by Hoogsteen base pairs between guanines. A number of alternative four-stranded structures can form and their existence has been confirmed by X-ray diffraction and NMR spectroscopy. An example of an antiparallel quadruplex DNA is shown in Figure 14.39.

Slipped DNA

DNA regions with direct repeat symmetry can form structures known as **slipped, mispaired DNA** (SMP-DNA). Their formation involves the unwinding of the double helix and realignment and subsequent pairing of one copy of the direct repeat with an adjacent copy on the other strand. This realignment generates a single-stranded loop (Figure 14.40). Two isomeric structures of a SMP-DNA are possible. One generates a loop consisting of the 5 direct repeat in both strands and the other produces loops of the 3 direct repeat. Although SMP-DNA has not yet been identified, genetic evidence suggests that this type of DNA is undoubtedly involved in spontaneous frameshift mutagenesis that is manifested as base addition or deletion occurring within runs of single bases. A mechanism that explains these mutations is shown in Figure 14.41. First, a homopolymeric sequence in one strand (template strand) unpairs from a newly synthesized complementary strand and reforms hydrogen bonds with a different set of bases, resulting in the formation of an extrahelical base on either the template strand or progeny strand. Continued replication produces a deletion when the progeny strand slips forward or a duplication when the strand slips backward. Deletions and duplications of DNA segments, longer than a single base, occur during DNA replication between direct repeats, which can form slipped-looped structures. Duplication of certain simple triplet repeats that are implicated as the basis of several human genetic diseases (see Clin. Corr. 14.6) may also occur by this mechanism.

Nucleoproteins of Eukaryotes Contain Histones and Nonhistone Proteins

DNA in eukaryotic cells is associated with various types of protein to form **chromatin.** In resting (nondividing) cells, chromatin is amorphous and dispersed within the nucleus. Just prior to cell division (mitosis), chromatin becomes organized into compact structures (fibers) called **chromosomes.** The

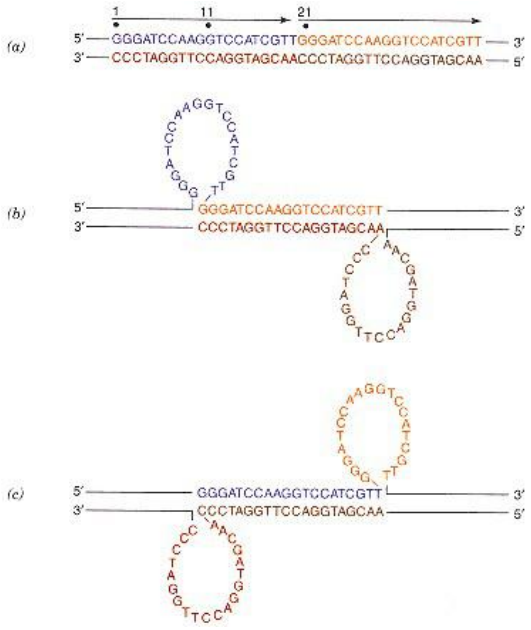


Figure 14.40
Slipped, mispaired DNA.
The presence of two adjacent tandem repeats (a) can give rise to either one of two isomers of slipped, mispaired DNA. In one of these isomers (b) the second copy of the direct repeat in the top strand pairs with the first copy of the repeat on the bottom strand. Pairing of the first copy of the direct repeat in the top strand with the second copy of the direct repeat in the bottom strand produces the second isomer (c). A pair of single-stranded loops is generated in both isomers.

CLINICAL CORRELATION 14.6

Expansion of DNA Triple Repeats and Human Disease

The presence of reiterated DNA sequences, consisting of three base pairs, has been noted in a number of human genetic diseases including fragile X syndrome, myotonic dystrophy, X-linked spinal and bulbar muscular atrophy (Kennedy syndrome), spinocerebellar ataxia, colon cancer, and more recently Huntington's disease. These diseases are associated with expansion of certain triplet nucleotide repeats that appear to be overrepresented in the human genome. For example, fragile X syndrome is characterized by expansion of a GCC triplet and spinocerebellar ataxia type I with expansion of a CAG triplet. Diseases associated with expansion of triplets are characterized by an increase in severity of the disease with successive generation, which is known as anticipation. For example, anticipation in fragile X syndrome, a leading cause of mental retardation, is associated with a major expansion of the CGG triplet. Normally, about 30 copies of this triplet are present on the 5 side of a gene associated with the disease, the *FMR-1* gene. The site of the repeat is expanded to as many as 300 copies in males that carry fragile X gene mutations but have no symptoms of the disease. Offspring of male carriers who express the disease can have a remarkable expansion of the triplet repeat, up to thousands of copies.

The disease develops when normal expression of *FMR-1* gene is turned off. Methylation of CpG dinucleotides present in CGG triplets appears to be associated with shutting off of the *FMR-1* gene. It appears that triplet expansion is the result of slipped mispairing during DNA synthesis. Because of the massive amplification that characterizes the diseases associated with triplet expansion, repeated or multiple slippage would have to be involved to explain the high degree of expansion. What promotes repeated slippage is not known but it may be that expansion is associated with a repeated dissociation of the enzyme DNA polymerase from the DNA template. This may allow DNA breathing and repeated slippage of DNA strands that are obviously required for the observed extensive expansion of the triplets. For slippage to occur, a single-stranded break needs to be generated within the tandem repeat during replication, which can lead to addition (or deletion) of a few copies of the tandem repeat. For modest size repeats, that is, repeats of less than about 80 copies, at least one such break is expected to be generated. When a larger number of repeats are present, it is possible that two single-stranded breaks are generated during replications. The strand segment flanked by these single-stranded breaks is not anchored by a unique sequence at either end and therefore it is free to slide during synthesis, leading to triplet amplification.

Behn-Krappa, A., and Doerfler, W. Enzymatic amplification of synthetic oligodeoxyribonucleotides: implications for triplet repeat expansions in the human genome. *Hum. Mutat.* 23:19, 1994.

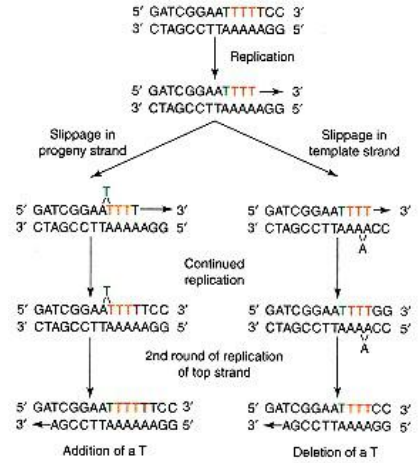


Figure 14.41
Frameshift mutagenesis by DNA slippage.
DNA replication within a run of a single base can produce a single base frameshift. In the example shown here, a run of five A's is replicated and, depending on whether a slippage occurs in the progeny strand or the template strand, a T may be added or deleted from the DNA.

division of genetic information into numerous independent domains, that is, chromosomes, may be necessitated by the enormous length of the genome of most eukaryotes. Each chromosome is characterized by the presence of a **centromere,** which functions as a site for attachment to proteins that link the chromosome to the mitotic spindle. Sister chromatids are connected at the centromere. **Telomeres** define the termini of linear chromosomes. A third element that characterizes chromosomes is the presence of a sequence required for the initiation of DNA replication **(origin of replication).** The number of chromosomes observed is species specific with human cells containing 46 chromosomes (chromatids) organized into 23 pairs. The average DNA length of each one of these chromosomes is 1.3×10^8 nucleotide pairs or approximately 5 cm. It is believed that each human chromosome consists of a single intact DNA molecule varying in size from 263×10^6 base pairs for chromosome 1 to less than 50×10^6 bp for chromosome 23. If the DNA of all 46 chromosomes were lined up in the B-DNA conformation, it would be more than 2 m long.

The chromosomal organization that makes it possible for DNA to fit within a cell nucleus with a diameter of approximately 5 μm requires a "condensation ratio" of more than five orders of magnitude. During metaphase the DNA molecule is very tightly wound. For example, human chromosome 16 is 2.5 μm long, whereas the DNA molecule is 3.7 cm in each of the two chromatids, giving a condensation ratio of 1.5×10^4:1. The parceling of DNA in 46 chromosomes provides for a further increase in the condensation ratio to 10^5:1. This remarkable degree of condensation of cellular DNA is shown in Figure 14.42. The early stages of DNA packing that lead to formation of 30-nm fibers have been extensively studied. The latter stages, in which looped domains of the 30-nm fiber are organized into scaffolds and chromatid coils, are based on indirect evidence and are more speculative. At each stage of packing, shown in this model, DNA is condensed severalfold. The cumulative effect of the successive folding stages provides the large condensation ratio necessary for the **packing of DNA** within the nucleus. The first stage of organization is the formation of a **"beads-on-a-string"** structure consisting of DNA associated with a class of highly basic proteins known as **histones.** These bind tightly to DNA, forming very stable complexes. The "beads-on-a-string" arrangement is seen in chromatin treated under conditions of low ionic strength and examined

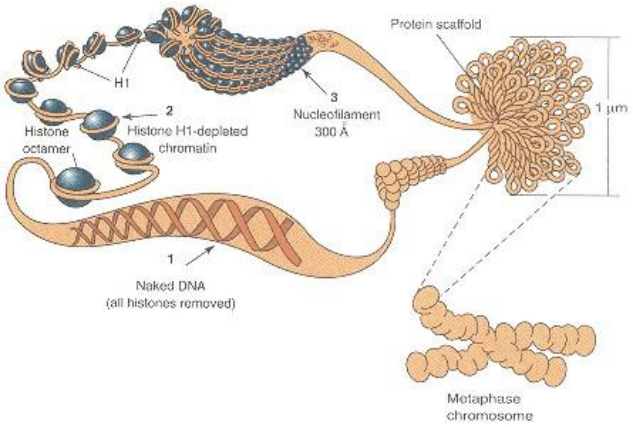


Figure 14.42
Organization of polynucleosomes into chromosomes. A speculative drawing showing the condensation of polynucleosomes into the 30-nm fiber and the subsequent packaging of this fiber into a twisted, looped structure attached to a protein scaffold within the chromosome.

under the electron microscope. The "string" is free DNA and the "beads" are coiled around histones.

Histones, regardless of their source, consist of five types of polypeptides of different size and composition (Table 14.7). The most "conserved" histones are H4 and H3, which differ very little even between extremely diverse species; histones H4 from peas and cows are very similar, differing by only two amino acids, although these species diverged more than a billion years ago. The H2A and H2B histones are less highly conserved but still exhibit substantial evolutionary stability, especially within their nonbasic portions. H1 histones are quite distinct from the inner histones. They are larger, more basic, and by far the most tissue-specific and species-specific histones. Vertebrates contain an additional histone, H5, which has a function similar to H1. As a result of their unusually high content of the basic amino acids lysine and arginine, histones are highly polycationic and interact with the polyanionic phosphate backbone of DNA so as to produce uncharged nucleoproteins. All five histones are characterized by a central nonpolar domain, which forms a globular structure, and N-terminal and C-terminal regions that contain most of the basic amino acids. The basic N-terminal regions of H2A, H2B, H3, and H4 comprising 20–25% of the histone octomer are the major, but not the exclusive, sites of interaction with DNA. Nonpolar domains and C-terminal regions of histones H1, H2A, and H2B are involved in subunit and DNA and histone interactions.

A heterogeneous group of proteins with high species, and even organ, specificity is also present in chromatin. These proteins, grouped together as

Table 14.7 Structure of the Five Types of Histones[a]

Name	*Structure*[b]	*Residues*	*Molecular Weight*
H4	N⊢●	102	11,300
H3	N⊢●⊣C	135	15,300
H2A	N⊢●⊣C	129	14,000
H2B	N⊢●⊣C	125	13,800
H1	-N⊢●⊣C	~216	~21,000

nonhistone proteins, consist of several hundred members, most of which are present in trace amounts. Many nonhistone proteins are associated with various chromosome functions, such as replication, gene expression, and chromosome organization.

Nucleosomes and Polynucleosomes

Histones interacting with DNA form the periodic "beads-on-a-string" structure, called a **polynucleosome,** in which an elementary unit, a nucleosome, is regularly repeated. Each **nucleosome** is a disk-shaped structure about 11 nm in diameter and 6 nm in height that consists of a DNA segment and a histone cluster composed of two molecules each of H2A, H2B, H3, and H4 histones. The clusters are organized as tetramers consisting of $(H3)_2$-$(H4)_2$ with an H2A-H2B dimer stacked on each face in the disk. The DNA is wrapped around the octamer as a negative toroidal superhelix at a pitch of about 30 Å with the central $(H3)_2$-$(H4)_2$ core interacting with the central 70–80 bp of the DNA wrap. Histones are in contact with the minor groove of DNA and leave the major groove available for interaction with the proteins that regulate gene expression and other DNA functions. Two distinct structures of nucleosomes can be distinguished: the **nucleosome core** and the **chromatosome**, as presented in Figure 14.43. The chromatosome constitutes the most elementary structural unit of nucleoproteins. These two structures are obtained by the digestion of polynucleosomes with nucleases (DNases) that, depending on conditions, can re-

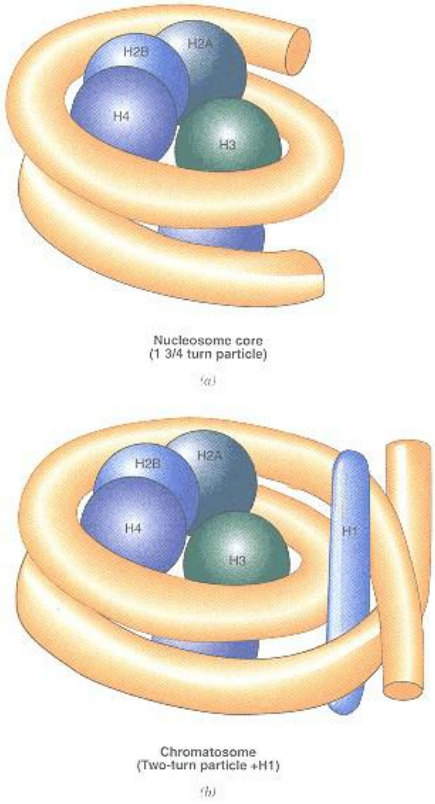


Figure 14.43
Postulated structures for the nucleosome and chromatosome.
The nucleosome consists of approximately 146 bp of DNA corresponding to 1 3/4 superhelical turns wound around a histone octamer. The chromatosome (two-turn particle) consists of about 166 bp of DNA (two superhelical turns). The H1 subunit is retained by this particle and may be associated with it, as shown. Chromatosomes containing less than 166 bp do not bind the H1 subunit.

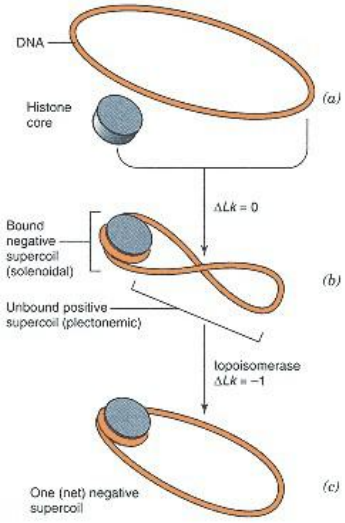


Figure 14.44
Generation of negative supercoiling in eukaryotic DNA.
The binding of a histone octamer to a relaxed, closed-domain DNA forces the DNA to wrap around the octamer, generating a negative supercoil. In the absence of any strand breaks, the domain remains intact and a compensating positive supercoil must be generated elsewhere within the domain. The action of a eukaryotic type I topoisomerase subsequently relaxes the positive supercoil, leaving the closed domain with one net negative supercoil.

move most or all DNA that is not tightly bound with histones. Nucleosomes obtained by nuclease digestion can be crystallized and studied by X-ray diffraction.

The structure of nucleosomes explains the puzzling finding that eukaryotic cells lack topoisomerases that can underwind DNA. It appears that negative superhelicity is, instead, introduced into eukaryotic cells as a result of DNA forming a toroidal wrapping around the histone core of nucleosomes (Figure 14.44). Such wrapping requires the removal of approximately one helical turn in DNA. Initially relaxed DNA subjected to such wrapping will generate a negative toroidal supercoil within the region bound around the histone core and a compensating positive supercoil elsewhere in the molecule, so as to maintain a constant linking number. Subsequent relaxation of the positive supercoil by eukaryotic topoisomerases leaves one net negative supercoil within the nucleosomal region.

Polynucleosomes consist of numerous nucleosomes joined by **"linker" DNA,** the size of which differs among cell types. Usually the nucleosome core is used as the elementary unit for describing the polynucleosome, in which case linker DNA size varies anywhere from about 20 to 90 bp. (Linker sequences would of course be proportionally smaller if the chromatosome were to be used as the elementary unit for the polynucleosome.) Since in addition to the linker sequence approximately 146 ± 1 bp are wrapped around the nucleosome core, the polynucleosome has a minimum nucleosome repeat frequency of about 168 ± 2 bp. Repeat frequencies for nucleosomes are found to depend on both the organism and the organ from which the cell is isolated and, as a rule, they appear to be relatively long in transcriptionally inactive cells. For example, chick erythrocytes have a repeat frequency of 212 bp. Active cells, such as yeast cells that have a frequency of 165 bp, generally have shorter linker sequences.

Periodicity of distribution of nucleosomes along the polynucleosome structure has been determined by controlled digestion with a nuclease that preferentially attacks linker DNA. The digestion pattern suggests the presence of nucleo-protein segments, which on the average contain about 200 bp of DNA or multiples of 200 that result from incomplete digestion. The relationship between size of segments and expected number of nucleosomes associated with them has been confirmed by electron microscopy. With the exception of a small amount of eukaryotic DNA, which is located in mitochondria and chloroplasts and which occurs in the form of small superhelices generally free of protein, all eukaryotic DNA is associated with histones.

Although nucleosomes are periodically positioned along the polynucleosome, their distribution is not random with respect to the base sequence of DNA. DNA does not bend uniformly but rather bends gently and then more sharply around the histone octamers. This suggests that DNA binding is sequence dependent and that **nucleosome positioning** may be influenced by the nucleotide sequence of DNA. In fact, nucleosomes tend to associate preferentially with certain DNA regions. DNA tracts that resist binding, such as long A tracts or G-C repeats, are not usually associated with nucleosomes. In contrast, certain bend DNA regions, for instance, periodically phased A tracts, associate strongly with histones. The majority of nucleosome core particles can relocate over a cluster of positions along the DNA separated by about 10 bp. The resulting mobility of these coil particles probably allows DNA polymerases and other enzymes to gain access to specific DNA sequences. The organization of DNA into nucleosomes appears to have fundamental consequences for transcription and DNA repair.

Polynucleosome Packing into Higher Structures

The wrapping of DNA around histones to form nucleosomes results in a tenfold reduction in the apparent lengths of DNA and the formation of the so-called

10-nm fiber (which is actually 11 nm wide), corresponding to the diameter of the nucleosomes. In chromosomes isolated by very gentle methods, both 10-nm fibers and thicker 30-nm fibers (in fact, 34 nm wide) can be seen in electron micrographs. The relationship between 30-nm fibers and 10-nm fibers has been further confirmed experimentally by the observation that 30-nm fibers can be dissociated into 10-nm fibers by treatment at low ionic strength. The 30-nm fibers appear to form by condensation of 10-nm fibers into a **solenoid arrangement** involving six to seven chromatosomes per solenoid turn (Figure 14.45). **Chromatosomes** are nucleosomes that contain a molecule of H1 histone. This histone is a protein consisting of three different domains that may bind DNA at the ends of the turn and at the point where DNA enters and exits the nucleosome at a ratio of one H1 per nucleosome. Adjacent H1 molecules may also bind to one another cooperatively, bringing the nucleosomes closer together in 30-nm fibers. The formation of the polynucleosome and its subsequent condensation into the 30-nm fibers provides for DNA a compaction ratio that may be as high as two orders of magnitude. The 30-nm fibers form only over selected regions of DNA that are characterized by the absence of binding with other sequence-specific (nonhistone) DNA-binding proteins. The presence of DNA-binding proteins and the effects on formation of 30-nm fibers may depend on the transcriptional status of the regions of DNA involved.

How polynucleosomes are organized into higher structures is not fully understood. Models as to the higher levels of packing of 30-nm fibers are based on indirect evidence obtained from studies of two specialized types of chromosomes—the **lampbrush chromosomes** of vertebrate oocytes and **the polytene chromosomes** of fruitfly giant secretory cells. These chromosomes are exceptional in that they maintain precisely defined higher-order structures in interphase, that is, when cells are in a resting (nondividing) state. The structural features of interphase lampbrush chromosomes have led, by extrapolation, to the proposal that chromosomes in general are organized as a series of looped, condensed domains of 30-nm fibers of variable size for different organisms. It

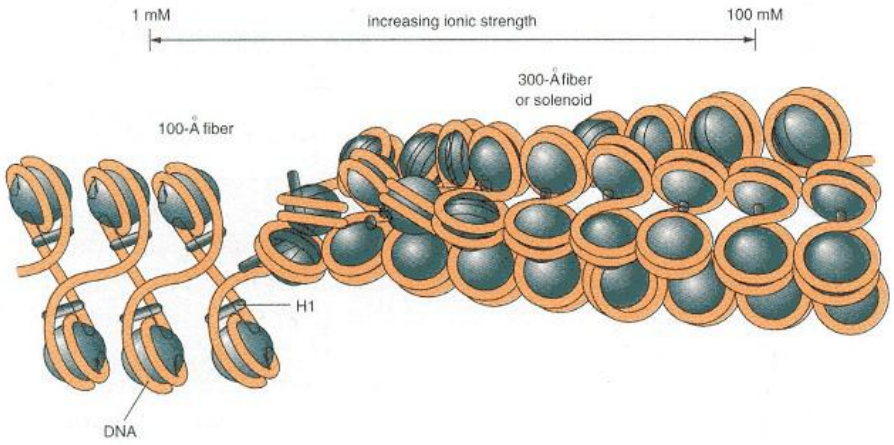


Figure 14.45
Nucleofilament structure.
Nucleofilament has the "string of beads" appearance, which corresponds to an extended polynucleosome chain. H1 histone is attached to the "linker" regions between nucleosomes, but in the resulting structure H1 molecules, associated to adjacent nucleosomes, are located close to one another. Furthermore, at higher salt concentrations, polynucleosomes can be transformed into the higher order structure of the 300-A fiber. It has been proposed that at higher ionic strengths the nucleofilament forms a very compact helical structure or a helical solenoid, as illustrated in the upper part of the figure. H1 histones appear to interact strongly with one another in this structure. In fact, the organization of the 10-nm (100-A) nucleofilament into the 30-nm (300-A) coil or solenoid requires, and may be dependent on, the presence of H1.
Adapted from Kornberg, R. D., and Klug, A. *The Nucleosome.* San Diego, CA: Academic Press, 1989.

is estimated that these loops may contain anywhere from 5000 to 120,000 bp with an average of about 20,000. Thus the haploid human genome of 3×10^9 bp would correspond to about 60,000 loops, which is close to the estimated number of genes of 70,000 to 100,000. It appears likely that each loop contains one or a few linked genes. The domains are bound to a nuclear scaffold consisting of H1 histone and several nonhistone proteins, including two major **scaffold proteins Sc1** (a topoisomerase II) and **Sc2.** The loops are fixed at their bases and therefore they can accumulate supercoils. Specific AT-rich regions of DNA known as **SARs (scaffold attachment regions)** are preferentially associated with the scaffold. SARs also contain topoisomerase II binding sites. The presence of type II topoisomerase at the base of closed topological domains, which define the scaffold loops, suggests that supercoiling and supercoiling changes within these domains are biologically important functions. Formation of looped domains may account for as much as an additional 200-fold condensation in the length of DNA and an overall packing ratio of more than four orders of magnitude. Each loop can be coiled and then supercoiled into 0.4 μm of a 30-nm fiber. Since the thickness of a sister chromatid is about 1 μm in diameter, packing of the 20-nm fiber into a chromatid would require just one more order of folding.

The next level of chromosomal organization may therefore involve the packing of loops as suggested in Figure 14.45. The packing may be achieved by arranging the loops of the 30-nm fiber in the form of tightly stacked helical coils. It is speculated that chromatids of metaphase chromosomes consist of helically packed loops of 30-nm fibers. Packing changes, and therefore the transition between the various forms of chromatin, appear to be partially controlled by the covalent modification of core histones. Histones H3 and H4 can undergo cell-cycle-dependent reversible acetylation on the ε-amino group of lysine by two different enzymes, a **histone acetylase** and a **histone deacylase.** Acetylation appears to affect the negative superhelical tension within domains and, in certain instances, the binding of transcription factors. The hydroxyl group of the N-terminal serine residue in histone H4 is subject to phosphorylation catalyzed by a kinase. Acetylation and phosphorylation change the charge of the N-terminal region of histone H4 from +5 to –2. The overall negative charge of the core histones causes histones to bind less tightly to DNA and promotes the unraveling of 30-nm fibers and the decondensation of chromatin. Finally, phosphorylation of terminal H1 correlates with chromosome condensation into metaphase chromosome. This may result from a modulation of affinity between phosphorylated–dephosphorylated H1 with the histone octamer. The change from compact to decondensed chromatin is also promoted by the binding of proteins, known as **HMG proteins (high-mobility-group proteins),** which interact preferentially with the transcriptionally active decondensed form of chromatin, that is, the 10-nm fiber.

Control of eukaryotic transcription and replication apparently involves both histone and nonhistone protein. While dissociation of histones from chromosomal DNA may be a prerequisite for transcription, nonhistone proteins provide more finely tuned transcription controls. Whatever the details of control may be, chromosomal regions actively synthesizing RNA are least condensed, in distinction from the more compacted, inactive regions. **Active genes** must be packaged in a way that makes them accessible to regulatory proteins. At the same time permanently **repressed genes** must remain inaccessible. Packaging may also determine the accessibility of DNA to DNA-damaging agents. Finally, nonhistone proteins control gene expression during differentiation and development and may serve as sites for the binding of hormones and other regulatory molecules.

Viral DNA is almost always complexed with protein, where the function of the protein is generally one of "packaging." In essence the protein protects the DNA from mechanical damage or digestion by endonucleases.

Nucleoproteins of Prokaryotes Are Similar to Those of Eukaryotes

In prokaryotic cells DNA is generally organized as a single chromosome that is a double-stranded circular supercoil. Some bacteria contain more than one chromosome and, in some, chromosomes may have linear structures. Prokaryotes lack histones. Instead, an abundant histone-like protein, the **HU protein,** is apparently responsible for the formation of a "beaded" structure seen in prokaryotes. HU (molecular mass 18 kDa) exists as a heterodimer of two nearly identical subunits (HU-1 and HU-2). Upon binding to DNA, HU changes the shape and the supercoiling of the double helix. The binding of HU to DNA *in vitro*, compacts DNA and restrains supercoils in a concentration-dependent manner and up to an equimolar ratio. This means that the interaction of DNA with HU at an equimolar ratio prevents topoisomerases from relaxing negatively supercoiled DNA in the DNA–HU complex. It also means that HU can introduce restrained supercoils in relaxed DNA. Higher concentrations of HU do not result in the restraining of additional supercoils. From the effects of HU on DNA supercoiling and other evidence, it appears that HU bends DNA sharply into a tight circle. In addition, another abundant small histone-like protein, referred to as **H-NS,** may be involved in chromosomal organization either directly or indirectly through interaction with the HU proteins.

Bacterial chromosomes are organized into compacted structures, called **nucleoids,** by interaction of HU and H-NS proteins and participation of various cations, **polyamines** (such as spermine, spermidine, putrescine, and cadaverine), RNA, and nonhistone proteins. In the case of *E. coli* the nucleoid consists of a single supercoiled DNA molecule organized into about 40 loops, each consisting of approximately 100 kb of DNA, that merge into a scaffold rich in protein and RNA (Figure 14.46). In prokaryotic scaffolds, the loops are maintained by interactions between DNA and RNA rather than DNA–protein interactions only, as is the case with eukaryotes. The genome of *E. coli* consists of about 4.5×10^6 bp, which, if they were straightened as a linear B-DNA, would be 1.5 mm long and therefore 80 times larger than the diameter of the *E. coli* cell. As a result of a nucleoid formation, which has a diameter of only 2 μm, the *E. coli* genome can easily be fitted within the constraints of the cell. Although the nucleoid, in analogy with the chromatin of eukaryotes, is organized in the form of looped domains, the organization of domains within larger compacted structures (chromosomes) that characterize eukaryotes is absent from prokaryotes. Bacterial chromosomes are dynamic structures formed with histone-like proteins, which bind and dissociate fairly rapidly. This may reflect the need for rapid DNA synthesis, cell division, and transcription that characterize bacterial cells. In contrast, histones bind much more stably with eukaryotic DNA and may dissociate only over areas of the genome that are engaged in DNA synthesis, repair, recombination, or transcription.

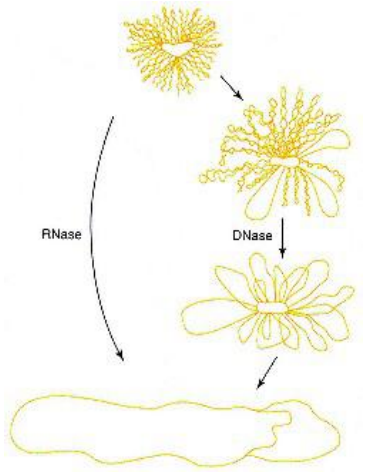


Figure 14.46
Schematic depiction of the folded chromosome of *E. coli*.
This chromosome contains about 50 loops of supercoiled DNA organized by a central RNA scaffold. DNase relaxes the structure progressively by opening individual loops, one at a time. RNase completely unfolds the chromosome in a single step.
Redrawn from Worcel, A., and Burgi, E. *J. Mol. Biol.* 71:127, 1972.

14.4— DNA Structure and Function

Overall base composition characterizes DNA only in a very general manner. A more specific property, which characterizes any DNA in a unique way, is the nucleotide sequence. Direct determination of nucleotide sequences in DNA remained an intimidating undertaking until the discovery of the restriction endonucleases.

Restriction Endonucleases and Palindromes

Restriction endonucleases cleave DNA chains at a specific sequence, making possible the sectioning of large DNA molecules into small segments. These highly specific bacterial enzymes act by making two cuts, one in each strand

TABLE 14.8 Examples of Sites of Cleavage of DNA by Restriction Enzymes of Various Specificities[a]

			Number of Cleavage Sites for Two Commonly Used Substrates	
Enzyme	***Microorganism***	***Specific Sequence***	***ϕX174***	***pBR 322***
EcoRI	*E. coli*	–G AATT–C– –C–TTAA G–	25	9
HaeIII	*Haemophilus aegyptius*	-GG ↓ CC- -CC ↑ GG-	11	22
HpaII	*Haemophilus parainfluenzae*	–C CG–G– –G–GC C–	5	26
HindIII	*Haemophilus influenzae* Rd.	–A AGCT–T– –T–TCGA A–	0	1

[a] Cleavage takes place within palindromes. The cleavage sites are indicated by arrows.

of double-stranded DNA of an invading phage, generating 3 -OH and 5 -P termini. This fragmentation exposes phage DNA to eventual degradation by bacterial exonucleases. The terminology for these endonucleases originates from the bacterial sources from which they are isolated. The first three letters of the name is an abbreviation of the species from which the enzyme is isolated. The next letter (or letters) designates the strain of the source and the Roman numeral simply refers to the order in which the enzyme was discovered from the strain. Many hundreds of restriction endonucleases have been isolated in pure form and the list of new restriction enzymes is growing daily. With few exceptions, these enzymes have been found to recognize sequences four to six nucleotides long. These sequences are completely **symmetrical inverted repeats**, known as **palindromes,** as illustrated by the examples listed in Table 14.8. The order of the bases is the same when the two strands of the palindrome are read in opposite directions. For example, in the case of the restriction enzyme EcoR1, isolated from *E. coli*, the order of the bases is GAATTC when read from the 5 terminus of either of the strands.

Restriction endonucleases are classified into three categories. Types I and III make cuts in the vicinity of the recognition site in a unpredictable manner. Type II specifically cleaves DNA within the recognition sequence. The cuts made by type II enzymes are indicated in Table 14.8 by arrows. Examples of products generated are shown in Figure 14.47.

These enzymes recognize specific sequences that occur along large DNAs with relatively low frequencies and fragment DNA very selectively. For example, a typical bacterial DNA, which may contain about 3×10^6 bp, will be cleaved into a few hundred fragments. A small virus or plasmid may have few or

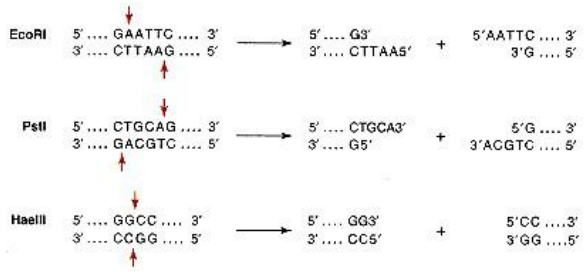


Figure 14.47
Types of products generated by type II restriction endonucleases.
Enzymes exemplified by EcoRI and PstI nick on both sides of the center of symmetry of the palindrome, generating single-stranded stubs. Commonly used enzymes generate 5 ends, although some produce stubs with 3 ends as shown for PstI. Other restriction nucleases cut across the center of symmetry of the recognition sequence, producing flush or blunt ends, as exemplified by HaeIII.

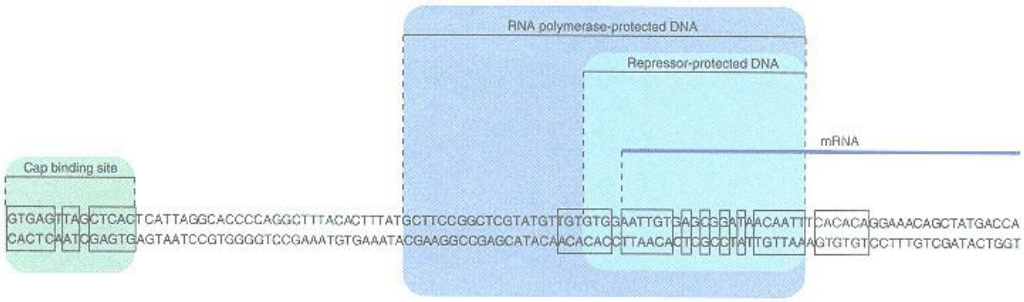


Figure 14.48
Nucleotide sequence of part of the DNA segment that controls synthesis of the enzyme β-galactosidaser in *E. coli* (the *lac* operon).
The binding regions of the cap protein, which acts as an activator of transcription, and of the lac repressor protein, an inhibitor of transcription, are indicated. Also shown is the region of RNA polymerase interaction.
Two palindromic sequences are indicated by boxes.
Redrawn from Cantor, C. R., and Schimmel, P. R. *Biophysical Chemistry, Part I.* San Francisco: Freeman, 1980.

no cutting sites at all for a particular restriction endonuclease. The practical significance of this selectivity of restriction enzymes is that a particular enzyme generates a unique family of fragments for any given DNA molecule. This unique fragmentation pattern is called a **restriction digest.**

The availability of restriction enzymes for sectioning large DNA sequences and the development of new gel electrophoresis techniques for separating DNA segments have made the determination of sequences a simple matter. These sequencing techniques are described in Chapter 18.

Early attempts to determine DNA sequences were limited to small DNA fragments that could easily be separated from the remaining DNA. Sequences that bind selectively with various functional proteins, for example, RNA polymerase and the repressor proteins, were among the first to be determined. The binding protein protects the DNA section over which it is bound from the action of a nuclease and the protected DNA is recovered after digestion and removal of the protein. These studies indicated that many functional proteins and enzymes interact with DNA over regions of palindromic sequence (Figure 14.48).

Palindromes in DNA also serve as recognition sites for **methylases** that modify the host DNA by introducing methyl groups into two bases of the palindrome. Once methylated, these palindromes cannot be recognized by the corresponding restriction enzymes, and the DNA of the host is protected from cleavage.

Contemporary sequencing methods have made possible determination of the complete nucleotide sequences of the DNA of viruses and small bacteria and the partial sequence of many eukaryotic genomes. An ambitious current goal of DNA sequencing is the determination of the sequence of the entire human genome, which consists of almost 3×10^9 bp, and that of several other mammalian organisms.

Most Prokaryotic DNA Codes for Specific Proteins

In prokaryotes a large percentage of total chromosomal DNA codes for specific proteins. Bacterial genomes vary from about 500 kb to over 10,000 kb. More than one-half of the *E. coli* genome has been sequenced. This genome consists of about 4600 kb of DNA and contains as many as 3000 genes. The products

of about one-half of *E. coli* genes have already been identified. It is possible that some of the remaining "genes" do not code for expressible functional proteins. Eighty genes code for tRNA molecules.

In an overall sense, *E. coli* DNA is densely packed with sequence information; there is little repetition of information in the genome. As much as 1% of the *E. coli* genome is composed of multiple copies of short repetitive sequences known as **repeated extragenic palindromic elements (REP elements).** REP elements are present at sites of DNA interaction with functional proteins as exemplified by the presence of such elements in the region of initiation of DNA synthesis (referred to as **OriC**). At **OriC, REP** elements with a consensus sequence of 34 nucleotides serve as sites for the binding of topoisomerase II, and REP elements with the sequence GCTGGTGG (**Chi** sites) bind the enzyme **RecBCD,** initiating DNA recombination. **Chi** sites are regularly spaced at intervals separated by about 4 kb.

Genetic information is even more densely organized in smaller organisms, such as bacteriophages, where the primary sequence of DNA reveals that structural genes—nucleotide sequences coding for protein—do not always have distinct physical locations. Rather, they frequently overlap with one another, as illustrated by the partial sequence of bacteriophage ϕX174 shown in Figure 14.49. It is believed that this type of overlap provides for the efficient and

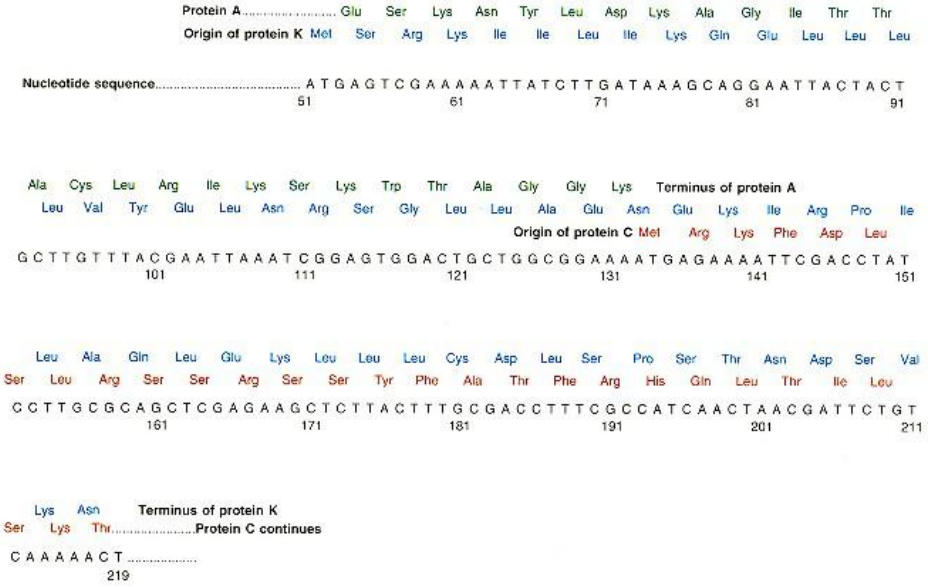


Figure 14.49
Partial nucleotide sequences of contiguous and overlapping genes of bacteriophage ϕX174.
The complete nucleotide sequence of X174 is known. Only the sequence starting with nucleotide 51 and continuing to nucleotide 219 is shown in this figure. This sequence codes for the complete sequence of one of the proteins of X174, protein K. A part of the same sequence, nucleotide 51 to nucleotide 133, codes for part of the nucleotide sequence of another protein, protein A. The sequence coding for protein K, which starts with nucleotide 133, also codes for part of a third protein, protein C. Similar overlaps are noted between other genes of X174. Adapted with permission from Smith, M. Am. Sci. 67:61, 1979. *Journal of Sigma Xi*, The Scientific Research Society.

economic utilization of the limited DNA present in these organisms. This arrangement of genes may also be a factor in controlling the sequence in which genes are expressed.

Only a Small Percentage of Eukaryotic DNA Codes for Structural Genes

Eukaryotes have a much larger genome than prokaryotes, from about 1.5×10^7 bp for yeast to about 3.5×10^9 bp for the haploid human genome. The latter contains sufficient DNA to code for nearly 3×10^6 genes. It is estimated, however, that the human genome codes for no more than $70–100 \times 10^3$ genes. As a result, genetic information in the form of genes need not be as densely packed in eukaryotes as in bacteria. A typical mammalian DNA, with only 20 times as many genes as that of *E. coli*, contains 500 times more DNA than *E. coli*. Clearly then, structural genes—that is, genes coding for specific proteins—and sequences used to control gene expression cannot account for the entire DNA content of eukaryotic cells. In fact, only 10% of DNA present in a mammalian cell may suffice for all of its genes that are present. Some of the remaining DNA, such as DNA found in centromeres and telomeres, has well-defined function, but the majority of this uncoding DNA has been referred to as "junk" because no specific function could be assigned to it. However, there is increasing evidence that junk DNA may have a vital role in the regulation of gene expression during development.

Nucleotide sequences indicate that eukaryotic genes not only do not overlap but are instead spaced on the average 40 kb apart. However, some eukaryotic genes may be closer together in regions containing genes that are expressed in a tightly coordinated manner (gene families). As a rule eukaryotic genes are, in addition, interrupted by **intervening nucleotide sequences (IVSs),** called **introns,** as shown in Figure 14.50. The nucleotide sequences in the gene that are expressed, either in the final RNA product (mature RNA) or as a protein, are termed **exons** (see p. 703). The **intervening genomic sequences** (the **introns**), which are expressed in the initial RNA transcript and are considered part of the gene, are removed during the processing of the transcript. The remainder of the message, namely, the **exons,** is then ligated. This tailoring of the original transcript is referred to as **splicing.** The sequence and size of introns vary greatly among species, but generally these intervening segments are very large and, cumulatively, they may be five to ten times the length of the parts of the structural genes they separate. Most genes are interrupted by introns at least once, whereas others are interrupted repeatedly. Some genes, however, such as the gene for human interferon-α, contain no introns.

Introns are common in genes of vertebrates and flowering plants but occur infrequently in the genes of other species. The biological role of introns is not clear. Their presence in eukaryotes may represent a stage in the evolution of the gene, in that introns are rare in prokaryotes and much less common in lower eukaryotes, such as **yeasts.** It has been speculated that introns in eukaryo-

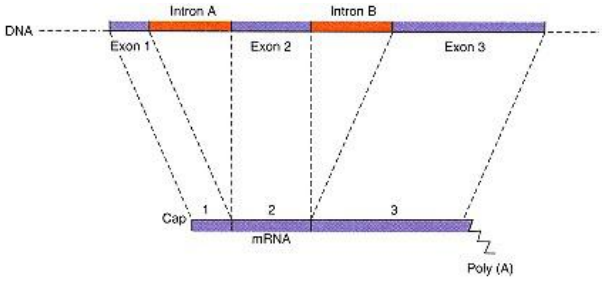


Figure 14.50
Schematic presentation of a eukaryotic gene.
The top horizontal line represents a part of the DNA genome of a eukaryote: the bottom line represents the mRNA produced by it. In this hypothetical example the DNA consists of two introns and three exons. The intron sequences are transcribed as hnRNA (precursor mRNA) but are not present in mature mRNA.
Redrawn from Crick, F. *Science* 204:264, 1979.

tic genes have arisen relatively recently in evolution as a result of migration of certain **mobile DNA elements (transposons)** from other parts of the genome and their insertion into protein-coding genes. These inserts subsequently lost, by mutation, their transposon-like character and therefore their mobility. Some repetitive DNA, such as the DNA found near centromeres and telomeres, may have well-defined structural and/or functional roles. Other repetitive DNA may simply be characterized as a leftover of evolutionary change with no identifiable function.

Repeated Sequences

As distinct from prokaryotes, where repetition of particular DNA sequences is very limited, the DNA of eukaryotes contains nucleotide sequences that are repeated anywhere from a few times, for certain coding genes, to millions of times per genome for certain simple, relatively short, sequences. Repetition of certain types of DNA sequences can be observed directly by electron microscopy, as in rRNA genes undergoing transcription. Depending on the species, **repetitive DNA** may constitute between 3% and 80% of the total DNA. In mammalian genomes, including the human, 25–35% of the DNA is repetitive.

Sequences are classified as **single copy, moderately reiterated,** and **highly reiterated.** The content of single-copy DNA varies among eukaryotes, increasing initially with genome size but reaching a plateau. Repetition classes are defined experimentally from their **rates of reassociation.** Reassociation rates also define a fourth class of DNA, **inverted repeats.**

A distinction between the terms "reiterated" and "repetitive" in describing a DNA sequence needs to be made. The term reiterated is used to describe a unique DNA sequence, usually several hundred nucleotides long, present in multiple copies in a genome. An individual DNA sequence is termed repetitive if a certain, usually short, nucleotide sequence is repeated many times over the DNA sequence.

The genome size of prokaryotic DNA can be determined by fragmenting the DNA, denaturing the fragments, and allowing them to reassociate and form double-stranded molecules. The kinetics of reassociation obey a second-order equation, indicating that essentially all the sequences in the prokaryotic genomes occur as single copies. When a mouse DNA was first studied by this method, unexpected results were obtained, which led to the realization that eukaryotic DNAs contain reiterated sequences. It was assumed that since the mammalian genome is about three orders of magnitude larger than the *E. coli* genome, the rates of reassociation of denatured mammalian DNA would be exceedingly slow. Instead, it turned out that a fraction of the mouse DNA, the highly repetitive fraction, reassociated far more rapidly than DNAs of small viruses. This is reasonable, since the probability that a fragment will encounter a complementary fragment leading to reassociation is proportional to the number of similar sequences repeated in the original DNA. The more reiterated the sequence, the more rapid the reassociation. Consequently, the reassociation kinetics of eukaryotic DNAs provided the first evidence for four classes or sequences. Inverted repeats and the highly repetitive sequences reassociate extremely rapidly. The unique sequences reassociate slowly, and the moderately reiterated at intermediate rates.

Most highly reiterated sequences have a characteristic base composition different from that of the remaining DNA. These sequences can be isolated by shearing the DNA into segments of a few hundred nucleotides each and separating the fragments by density gradient centrifugation. These fragments are termed **satellite DNA** because after centrifugation they appear as satellites of the band of bulk DNA. Other highly reiterated sequences, which cannot be isolated by centrifugation, can be identified by their property of rapid reannealing. Some of the highly reiterated sequences can also be isolated by digestion of total

DNA with restriction endonucleases that cleave at specific sites within the reiterated sequence. The exact boundaries separating the various types of reiterated DNAs do not appear to have been strictly defined.

Single-Copy DNA

About one-half of the human genome is made up of unique nucleotide sequences but, as indicated previously, only a small fraction of these sequences code for specific proteins. A part of the remaining DNA contains **pseudogenes**—that is, tracts of DNA that have significant nucleotide homology to a functional gene but contain mutations that prevent gene expression. These genes, which may be present in a frequency as high as one pseudogene for every four functional genes, significantly increase the size of eukaryotic genomes without contributing to their expressible genetic content. Additional DNA sequences are committed to serve as introns and as regions that are flanking genes.

Moderately Reiterated DNA

This class of DNA includes copies of identical or closely related sequences that are reiterated from a few to a thousand times. These sequences are relatively long, varying between a hundred to many thousand nucleotides before the same polynucleotide sequence is repeated. About 20% of mouse DNA occurs in lengths up to a few hundred base pairs that are repeated more than a thousand times. About 15% of the human genome consists of moderately reiterated DNA. Normally, single-copy and moderately reiterated sequences are present on the chromosome in an orderly pattern known as the **interspersion pattern,** which consists of alternating blocks of single-copy DNA and moderately reiterated DNA. Moderately repetitive sequences are further classified as **short interspersed repeats** that are families of related, but distinct, sequences typically 100–500 bp long and **long interspersed repeats** anywhere from about 100 bp up to several thousand base pairs long. Both short and long repeats are present at 1000 or up to 100,000 copies or more per genome. Long interspersed repeats consist of sequences several thousand nucleotides long that are present at up to 1000 copies per genome. These repeats are flanked on either side of the sequence by DNA sequences that are direct repeats (Figure 14.51). One

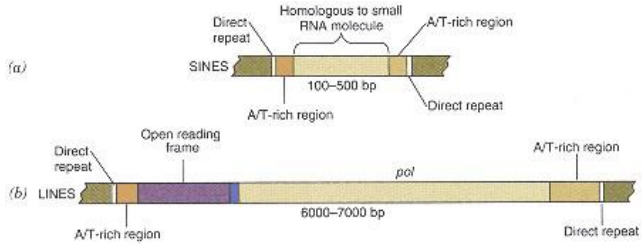


Figure 14.51
Short and long interspersed repeats in DNA.
Two types of interspersed repeats, short and long repeats, are found within eukaryotic DNA. (a) Short interspersed repeats are sequences 100–500 bp long that are homologous to small RNA molecules such as tRNA, 5SRNA or 7SLRNA (signal recognition particle). The human version of 7SLRNA is referred to as the AU sequence and accounts for approximately 10% of human DNA. (b) The long interspersed repeats that are present in hundreds of copies are homologous to tRNA genes and contain open reading frames with additional protein encoding sequences that resemble retroviral genes, such as the *Pol* gene. Both types of interspersed repeats contain short AT-rich sequences at the 3 terminals which are flanked by short direct repeat DNA.

example of a short interspersed repeat is the **Alu** family that constitutes a substantial portion (about 5%) of the human genome. **Alu** sequences consist of approximately 300 bp and are repeated over one-half million times. The structures of the short interspersed repeats, including the **Alu** family, are reminiscent of mobile DNA elements (transposons). The function of the **Alu** family remains to be established.

Interspersed repeats may have arisen during evolution from viruses or other transportable DNA elements that have been duplicated repeatedly and inserted into various locations within the chromosome. If this is the case, then short interspersed repeats would be nothing more than an evolutionary relic that performs no useful function for the host cell. On the assumption that this premise is correct, short interspersed repeats have been called "selfish DNA."

The interspersion pattern implicates the moderately reiterated sequences in control of transcription of structural genes since the large majority of structural genes are adjacent to reiterated sequences. A different type of moderately reiterated sequence occurs in the form of segregated tandem arrays. The two distinct types of arrangements of the moderately reiterated sequences appear to relate to different functions for these sequences. **Tandem arrays** are used for synthesis of products that must be rapidly generated in numerous copies, such as ribosomal RNA and certain proteins of specialized function. For example, in sea urchin oocyte histone, genes are amplified so that sufficient amounts of histone are available during the rapid cycles of DNA replication that follow fertilization. The genes for the five histones are arranged in tandemly repeated clusters, with each histone gene separated from its neighbor in the cluster by spacers about 400–900 nucleotides long. These spacers are AT-rich and can be separated as satellite DNA from the GC-rich DNA of the histone genes.

Single-copy and moderately repetitive sequences together normally account for more than 80% of the total nucleotide content of the eukaryotic genomes.

Highly Reiterated DNA

The remaining DNA consists of sequences constructed by the repetition, many thousand or even a million times, of a nucleotide sequence that is typically shorter than 20 nucleotides. About 10% of mouse DNA consists of 10-bp repeats that are reiterated millions of times in each cell. Because of the manner in which they are constructed, **highly reiterated DNAs** are also referred to as simple sequence DNA. Simple sequences are typically present in the DNA of most, if not all, eukaryotes. In some only one major type of simple sequence may be present. Thus in the rat the sequence 5 -GCACAC-3 is repeated every six bases. In other eukaryotes several simple sequences are repeated up to one million times. Some considerably longer repeat units for simple sequence DNA have also been identified. For instance, in the genome of the African green monkey a 172-bp segment is highly repeated and there are few sequence repetitions within the segment. Because of its characteristic composition, simple sequence DNA can often be isolated as satellite DNA. Satellite DNA found in the centromeres of higher eukaryotes consists of thousands of tandem copies of one or a few short sequences. **Satellite sequences** have been found to be only 5–10 bp long. Simple sequence (satellite) DNA is also a constituent of telomeres where it has a well-defined role in DNA replication.

Inverted Repeat DNA

Inverted repeats are a structural motif of **dos DNA.** Short inverted repeats, consisting of up to six nucleotides, such as the palindromic sequence GAATTC, occur by chance about once for every 3000 nucleotides. Such short repeats cannot form a stable "hairpin" structure formed by longer palindromic sequences. Inverted repeat sequences that are long enough to form stable "**hair-**

CLINICAL CORRELATION 14.7

Mutations of Mitochondrial DNA: Aging and Degenerative Diseases

Somatic mutations, such as deletions of bases or oligonucleotide segments from mtDNA, are generated by oxygen damage during the life span of an individual. Somatic mutations in mtDNA are acquired at a much higher rate than in nuclear DNA. They are responsible for disorders associated with the process of oxidative phosphorylation and they may also be involved in aging and the development of degenerative diseases.

MtDNA mutations are the cause of Leber hereditary optic neuropathy (LHON). This disease, which is maternally inherited, is characterized by loss of vision in early adulthood, as a result of optic nerve degeneration. One mutation, an Arg to His substitution that leads to this disease, has been traced to a gene coding for NADH dehydrogenase (Complex I). The mutation results in mitochondria that are partially defective in electron transfer from NADH to ubiquinone and have a reduced capacity of ATP synthesis needed to support the active metabolic needs of neurons. LHON can also result from a single base change in the mitochondrial gene coding for cytochrome *b*. A mutation of the mitochondrial gene coding for a tRNA is responsible for myoclonic epilepsy and ragged-red-fiber disease (MERRF). This genetic disease, which is characterized by uncontrollable muscular jerking, is apparently caused by inadequate production of proteins that depend on mitochondrial transfer RNAs for their synthesis.

Deletions and rearrangements in mtDNA are noted with aging in both humans and mice. Five different mtDNA deletions have been noted with aged mice but these deletions are absent from young mice. The deletions involve a small portion (less than 0.01%) of total mtDNA. The deletion of a large portion of mtDNA (a 4977-bp segment), which is the most frequently noted DNA abnormality in patients with mitochondrial myopathies, is also noted, although to a much lesser degree, in tissues of healthy aging individuals.

The observations that mtDNA is easily mutated and poorly repaired have led to speculation that aging may be correlated with accumulation of somatic mutations in mtDNA. However, both environmental and genetic factors probably affect the aging process and aging is not likely to be explained solely as the result of defective mtDNA function.

Tanhauser, S. M., and Laipis, P. J. Multiple deletions are detectable in mtDNA of aging mice. *J. Biol. Chem.* 270:24769, 1995.

pins" are not likely to occur by chance, and therefore they should be classified as a separate class of eukaryotic sequences. Short repeats can easily be detected and quantitated on the basis of their extremely rapid rates of reassociation. In human DNA, about two million inverted repeats are present, with an average length of about 200 bp, although inverted sequences longer than 1000 bp have been detected. Some of these repeats may be separated by a spacer sequence that is not part of the inverted repeat. Most inverted repeat sequences are repeated 1000 or more times per cell.

Mitochondrial DNA

The **DNA of mitochondria (mtDNA)** is a small double-stranded circular structure of approximately 16,500 bp. In mammals, mtDNA makes up about 1% of total cellular DNA. Mitochondria contain multiple copies of DNA, usually distributed within several clusters. It is not known how this DNA is packaged but its structure probably resembles that of a bacterial chromosome rather than eukaryotic chromatin. The sequence of human mtDNA consists of 16,569 bp and contains 37 genes. Thirteen genes code for proteins that are subunits for factors essential for the maintenance of mitochondrial ATP synthesis. The remaining 24 genes code for mitochondria-specific RNAs, two ribosomal and 22 transfer RNAs.

The rate of mutation is one order of magnitude greater in the mitochondrial genome as compared to the nuclear genome. These high rates of mutation probably reflect a low fidelity of DNA replication, DNA repair, or both. Mitochondrial genes are maternally inherited because mitochondria from the sperm cells do not enter the fertilized egg. The effects of mtDNA mutations are discussed in Clin. Corr. 14.7.

Bibliography

Bates, A. D. *DNA Topology*. Oxford, England: IRL Press, 1993.

Binden, R. R. *DNA Structure and Function*. San Diego, CA: Academic Press, 1994.

Borisova, O. F., Golova, Y. B., Gottikh, B. P., et al. Parallel double stranded helices and the tertiary structure of nucleic acids. *J. Biomol. Struct. Dyn.* 8:1187, 1991.

Bouffler, S., Silver, A., and Cox, R. The role of DNA repeats and associated secondary structures in genomic instability and neoplasia. *Bioessays* 15:409, 1993.

Bray, A. B., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. *Molecular Biology of the Cell*, New York: Garland, 1994.

Calladine, C., and Drew, H. *Understanding DNA—The Molecule and How It Works*. San Diego, CA: Academic Press, 1992.

Clark, D. J., and Felsenfeld, G. Formation of nucleosomes on positively supercoiled DNA. *EMBO J.* 10:387, 1991.

Clayton, D. A. Replication and transcription of vertebrate mitochondrial DNA. *Annu.-Rev. Cell Biol.* 7:453, 1991.

Cortopassi, G., and Liu, Y. Genotypic selection of mitochondrial and oncogenic mutations in human tissue suggests mechanisms of age-related pathophysiology. *Mutat. Res*. 338:151, 1995.

Cozzarelli, N. R., Boles, T. C., and White, J. H. Primer on the topology and geometry of DNA supercoiling. In: N. R. Cozzarelli and J. C. Wang (Eds.), *DNA Topology and Its Biological Effects*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1990.

De Vries, J. E., and de Waal Malefyt, R. *Therapeutic Applications of Oligonucleotides.* New York: Demos Vermande, 1995.

Drake, J. W. Spontaneous mutation. *Annu. Rev. Genet.* 25:125, 1991.

Drlica, K., and Riley, M. *The Bacterial Chromosome*. Washington, DC: American Society for Microbiology, 1990.

Drlica, K., Bacterial topoisomerases and the control of DNA supercoiling. *Trends Genet.* 6:433, 1990.

Drlica, K. Control of bacterial DNA supercoiling. *Mol. Microbiol.* 6:425, 1992.

Getzenberg, R. H., Pienta, K. J., Ward, W. S., and Coffey, D. S. Nuclear structure and the three-dimensional organization of DNA. *J. Cell. Biochem.* 47:289, 1991.

Han, J., Hsu, C., Zhu, Z., Longshore, J. W., and Finley, W. H. Over-representation of the disease associated (CAG) and (CGG) repeats in the human genome. *Nucleic Acids Res.* 22:1735, 1994.

Haran, T. E., Kahn, J.D., and Crothers, D. M. Sequence elements responsible for DNA curvature. *J. Mol. Biol.* 244:135, 1994.

Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates. *Biochemistry* 30:4460, 1991.

Kameneteskii, F. *Unraveling DNA*. New York: VCH, 1993.

Kim, M. G., Zhurkin, V. B., Jernigan, R. L. and Camerini-Otero, R. D. Probing the structure of a putative intermediate in homologous recombination: the third strand in the parallel DNA triplex is in contact with the major groove of the duplex. *J. Mol. Biol.* 247:874, 1995.

Leach, D. R. Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* 16:893, 1994.

Lebowitz, J. Through the looking glass: the discovery of supercoiled DNA. *Trends Biochem. Sci.* 15:202, 1990.

Lee, M. S., and Garrard, W. T. Positive DNA supercoiling generates a chromatin conformation characteristic of highly active genes. *Proc. Natl. Acad. Sci. USA* 88:9675, 1991.

Lewin, B. *Genes V*. New York: Oxford University Press, 1994.

Lindsten, J., and Pettersson, U. *Etiology of Human Disease at the DNA Level.* New York: Raven Press, 1991.

Luzio, J. P., and Thompson, R. J. In: *Molecular Medical Biochemistry*, New York: Cambridge University Press, 1990.

Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., Haworth, I. S. Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: $(CTG)_{15}$. *Nucleic Acids Res.* 23:1050, 1995.

Neidle, S. *DNA Structure and Recognition*, 1st ed. Oxford, England: IRI. Press, 1994.

Orphanides, G., and Maxwell, A. Topoisomerases. In one gate, out the other. *Curr. Biol.* 4:1006, 1994.

Owen-Hughes, T., and Workman, J. L. Experimental analysis of chromatin function in transcription control. *Crit. Rev. Eukaryot. Gene Expr.* 4:403, 1994.

Ozawa, T. Mitochondrial DNA mutations associated with aging and degenerative diseases. *Exp. Gerontol.* 30:269, 1995.

Saluz, H. P., and Wiebauer, K. *DNA and Nucleoprotein Structure in Vivo.* New York: Demos Vermande, 1995.

Soyfer, V. N., and Potaman, V. N. *Triple-Helical Nucleic Acids*. New York: Springer-Verlag, 1995.

Therman, E., and Susman, M. In: *Human Chromosomes, Structure, Behavior and Function*, 3rd ed. New York: Springer-Verlag, 1992.

Trinh, T. Q., and Sinden, R. R. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli. Nature* 352:544, 1991.

Trinh, T. Q., and Sinden, R. R. The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli. Genetics* 134:409, 1993.

Turner, B. M. Histone acetylation and control of gene expression. *J. Cell. Sci.* 99:13, 1991.

Ussery, D. W., and Sinden, R. R. Environmental influences on the in vivo level of intramolecular triplex DNA in *Escherichia coli. Biochemistry* 32:6206, 1993.

Van Holde, K. E. *Chromatin*. New York: Springer-Verlag, 1989.

Vologodskii, A. V., and Cozzarelli, N. R. Conformational and thermodynamic properties of supercoiled DNA. *Annu. Rev. Biophys. Biomol. Struct.* 23:609, 1994.

Wang, C. I., and Taylor, J. S. Site-specific effect of thymine dimer formation on $dA_n \cdot dT_n$ tract bending and its biological implications. *Proc. Natl. Acad. Sci. USA* 88:9072, 1991.

Wolfe, S. L. *Molecular and Cellular Biology*, Belmont, CA: Wadsworth, 1993.

Wolffe, A. *Chromatin Structure and Function*. San Diego, CA: Academic Press, 1992.

Yu, H., Etitja, R., Bloom, L. B., and Goodman, M. F. Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. *J. Biol. Chem.* 268:15935, 1993.

Questions

C. N. Angstadt and J. Baggott

1. A. polynucleotide is a polymer in which:

   A. the two ends are structurally equivalent.

   B. the monomeric units are joined together by phosphodiester bonds.

   C. there are at least 20 different kinds of monomers that can be used.

   D. the monomeric units are not subject to hydrolysis.

   E. purine and pyrimidine bases are the repeating units.

2. The best definition of an endonuclease is an enzyme that hydrolyzes:

   A. a nucleotide from only the 3′ end of an oligonucleotide.

   B. a nucleotide from either terminal of an oligonucleotide.

   C. a phosphodiester bond located in the interior of a polynucleotide.

   D. a bond only in a specific sequence of nucleotides.

   E. a bond that is distal (d) to the base that occupies the 5′ position of the bond.

3. All of the following tend to favor a helical conformation of a single polynucleotide chain EXCEPT:

A. hydrophobic interactions of the rings of the purine and pyrimidine bases that exclude water.

B. interchange of electrons in the π orbitals of the purine and pyrimidine bases.

C. charge–charge repulsion of phosphate residues of the polynucleotide backbone.

D. hydrogen bonding between appropriate purine–pyrimidine pairs.

E. spacing of bases in the helical conformation that excludes water.

4. In a DNA double helix:

A. the individual strands are not helical.

B. hydrogen bonds form between a purine and a pyrimidine base on the same strand.

C. adenine on one strand is hydrogen-bonded to thymine on the opposite strand.

D. phosphodiester bonds are oriented toward the interior of the helix.

E. the outside of the helix is neutral.

5. The A helix of DNA differs from the B helix in all of the following EXCEPT:

A. appearance of the major and minor grooves.

B. pitch of the base pairs relative to the helix axis.

C. thickness of the helix.

D. tilt of the bases.

E. polarity of the strands.

6. The Z DNA helix:

A. has fewer base pairs per turn than the B DNA.

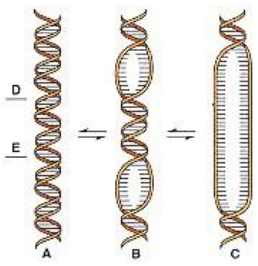
B. is favored by an alternating GC sequence.

C. tends to be found at the 3 end of genes.

D. is inhibited by methylation of the bases.

E. is a permanent conformation of DNA.

Use the accompanying figure to answer Questions 7 and 8.



7. A, B, and C represent conformations at different temperatures. Which one represents the highest temperature?

8. Which section, D or E, has the higher content of guanine and cytosine?

Refer to the following for Questions 9–11.

A. annealing

B. electrophoresis

C. equilibrium centrifugation

D. C_0t curves

9. A technique for determining the molecular weight of large (10^5 to 10^9 Da) DNA.

10. A technique involved in locating a specific gene on DNA with a probe.

11. A technique for assessing genome complexity.

12. The superhelices that form in double-stranded circular DNA:

A. may have fewer turns of the helix per unit length than does a linear double helix.

B. are associated with a restricted topological domain.

C. may exist in multiple conformations that are interconvertible without breaking covalent bonds.

D. may be either formed or relaxed by enzymes called topoisomerases.

E. all of the above are correct.

13. Triple-stranded DNA:

A. generally occurs in DNA in regions that play no role in transcription.

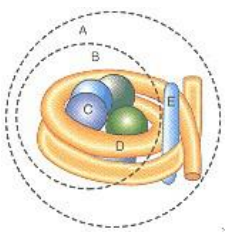
B. involves the formation of Hoogsteen pairs.

C. is characterized by the presence of a string of alternating purine–pyrimidine bases.

D. forms only intermolecularly.

E. assumes a cruciform conformation.

Using the accompanying figure to answer Questions 14–16.



14. A chromatosome.

15. DNA.

16. H1 class of histones.

17. A palindrome is a sequence of nucleotides in DNA that:

A. is highly reiterated.

B. is part of the introns of eukaryotic genes.

C. is a structural gene.

D. has local symmetry and may serve as a recognition site for various proteins.

E. has the information necessary to confer antibiotic resistance in bacteria.

18. An interspersion pattern in DNA consists of:

A. highly repetitive DNA sequences.

B. the portion of DNA composed of single copy DNA.

C. Alu sequences.

D. alternating blocks of single copy DNA and moderately repetitive DNA.

E. alternating blocks of short interspersed repeats and long interspersed repeats.

Answers

1. B A: The structure of a polynucleotide possesses an intrinsic sense of direction that does not depend on whether a 3 -OH or 5 -OH terminal is esterified C and E: There are only four different monomers, and the repeating unit is the base mono-phosphate (p. 566).

2. C Both A and B describe exonucleases. D does refer to an endonuclease but only to a specific type, a restriction endonuclease, and is therefore not a definition of the general type. E: Both endo- and exonucleases show specificity toward the bond hydrolyzed and so this is not a definition of an endonuclease (p. 568).

3. D This is very important in holding two different polynucleotide chains together, but it is unlikely that the proper positioning would occur within a single chain. A, B, C, and E: The exclusion of water by stacking of the bases is a strong stabilizing force that is enhanced by the interaction of π orbital electrons. The repulsive forces of the phosphate groups confer a certain rigidity to the structure (pp. 568–569).

4. C This results in complementarity of the strands. A: Single strands are right-handed helices. B: Bases in a single stand interact through the hydrophobic faces of the rings. D and E: The phosphate groups are negative and on the exterior of the helix (p. 580).

5. E The two strands are always antiparallel. A: The A helix has a narrower and deeper major groove and a wider and shallower minor groove than B. B and D: Bases in the B helix are almost perpendicular to the helix axis while those in A are tilted. C: The B helix is thinner than the A (pp. 570–573).

6. B The alternating purine-pyrimidine sequence is important. A: Z DNA is longer and thinner than the B form because it has 12 bp per turn instead of 10. C: It is more likely to be found at the 5 end, consistent with one of its proposed roles in transcriptional regulation D: Methylation favors the Z form in which the methyl is protected from water. E: B Z transition is influenced by such things as methylation and rotation of the G to the syn conformation (p. 574).

7. C; 8. E. The figure represents the process of denaturation with the extent of disruption increasing as temperature increases. Since a guanine–cytosine pair has three hydrogen bonds and an adenine–thymine pair only has two, higher temperatures are required to disrupt regions high in G-C (Figure 14.14).

9. C Equilibrium centrifugation is a method for determining molecular weight, but it is limited to a molecular weight of 10^9 Da or less because of the effects of shear forces on large molecules (p. 586).

10. A The probe is a labeled polynucleotide with a sequence complementary to the gene of interest. Annealing of the two permits location of the gene (pp. 580 and 582).

11. D Frequency of reassociation is dependent on amounts of highly reiterated versus unique sequences (p. 580).

12. E A describes a negative superhelix. There may also be more turns per unit length in a positive superhelix. B: Once a closed system is interrupted, a superhelix can unwind. C: All conformations with the same linking number *L* are interconvertible without breaking covalent bonds. D: Topoisomerase I (omega protein) from *E. coli* relaxes and gyrase (topoisomerase II) can introduce or remove superhelices (depending on the conditions) (pp. 587–592).

13. B Hoogsteen pairs, like TAT or GGC, are responsible for holding the third strand in the major groove. A: They are found frequently in regions involved in gene regulation. C: The required sequence is a homopurine string. D: They can also form intramolecularly by unfolding and refolding of the DNA. E: A cruciform is an alternate conformation of DNA but does not involve a third strand (pp. 595 and 597).

14. A; 15. D; 16. E. The chromatosome, the basic structural element of nucleoprotein, contains the nucleosome core with associated H1 histones. B: The nucleosome core is a discrete particle consisting of an octamer of specific histones with a segment of DNA wrapped around it. D: The strand depicted represents DNA; the circles, histones. E: The H1 class of histones is bound to the spacer regions between nucleosomes. C: Represents one of the histones (H2A, H2B, H3, or H4), which are part of the nucleosomes (Figure 14.43).

17. D A palindrome, by definition, reads the same forward and backward. Short palindromic segments of DNA are recognized by a variety of proteins such as restriction endonucleases and CAP-binding protein. A is not likely since it would be incompatible with specific recognition B is possible but has not been shown. C is not correct since genes are thousands of base pairs in length, whereas palindromes are short segments. E also would not be likely because palindromes are too short (p. 610).

18. D A and B: These are two of several kinds of DNA but do not constitute patterns. C and E: Alu is a type of short interspersed repeat. Short and long interspersed repeats are the two classes of moderately repetitive DNA (p. 615).