Chapter 3
The Families of Biological Molecules

Diversity is essential to the sustainability of living systems. This is true for species in ecosystems as it is for molecules in cells, tissues, and organisms. Yet, the same way different species are linked by common ancestors and may be grouped in taxonomic classes according to common characteristics they share, molecules may be grouped in classes and classified according to common chemical and physical characteristics. One of such characteristics is solubility in water (in other words: how polar atoms are distributed in the 3D structure of molecules). One class of biological molecules, the lipids, includes only low water solubility ("hydrophobic") molecules, this being the characteristic that defines this class. Other classes include molecules that are mostly moderately or highly soluble in water and can be recognized for the dominant presence of specific chemical groups: OH in saccharides (also referred to as "carbohydrates") and a combination of amino and carboxyl groups in amino acids. Lipids, saccharides, and amino acids may combine with molecules of their own class to form either polymers (molecules formed by successively covalently attaching smaller molecules), such as polysaccharides and proteins, or supramolecular assemblies (organized arrangements of molecules that are in contact but are not covalently attached), such as the lipid bilayer of cell membranes. It is common to find molecules and supramolecular assemblies that combine elements from different classes, such as nucleotides, which contain saccharides. Proteins are extremely versatile in this regard because protein interactions with saccharides, lipids, and nucleic acids (nucleotide polymers) are ubiquitous in virtually all cells.

Figure 3.1 depicts the basic principles that support the organization of biological molecules in different classes.

A.T. Da Poian, M.A.R.B. Castanho, *Integrative Human Biochemistry*,
DOI 10.1007/978-1-4939-3058-6_3

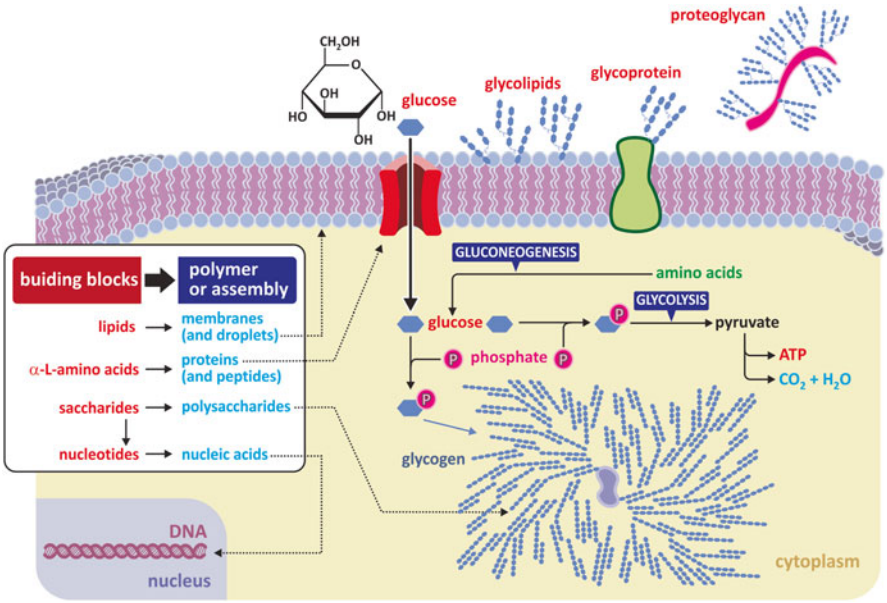


Fig. 3.1 Three classes of fundamental molecules ("building blocks") are enough to organize in different families most biological macromolecules (polymers) and large supramolecular assemblies (such as lipid bilayers and lipid droplets). Nucleic acids are polymers of nucleotides, which have saccharide residues in their composition. Proteins are polymers that are commonly combined with molecular residues of other classes such as saccharides (glycoproteins or proteoglycans when the saccharidic content is high). Glycolipids (glycosylated lipids—lipids with saccharide groups attached) are also common in membranes. Both building blocks and polymers/assemblies are important in the structure and functioning of cells, constituting the vast majority of matter in a cell (excluding water). Glycogen is a branched saccharide polymer (polysaccharide). Branching occurs at every 10 glucose monomer residues approximatly (branching unrealisticly highlighted in the figure for illustrative purposes)

3.1 Lipids and the Organization of Their Supramolecular Assemblies

Lipids are highly hydrophobic molecules that nonetheless may have polar chemical groups in their composition. Because one portion of the structure of the molecule is polar and the other is non-polar, the molecule is referred to as amphiphilic, which stresses its dual nature: the polar part will tend to interact with water and other polar molecules, and the other will tend to minimize its interaction with water and other polar molecules. Nevertheless, it is important to bear in mind that lipids are molecules in which hydrophobicity predominates, even if they are amphiphilic. This is a qualitative definition with no clear boundaries in terms of molecular structure, which is nonetheless a useful working definition because hydrophobicity grants lipids the ability to organize in supramolecular assemblies that are very distinctive from polar molecules. Take lipid bilayers as example: they are very extensively

organized supramolecular assemblies that are very stable and yet do not involve covalent bonds between lipid molecules. Lipids spontaneously self-associate in aqueous environments, and amphiphilic lipids in particular may self-associate in a very organized way. This results from the so-called hydrophobic effect, although the most appropriate term would be "entropic effect."

The entropic effect is a corollary of the second law of thermodynamics, which in one of its possible statements implies that all physical and chemical events tend to evolve in a way so that total entropy ("disorder") increases. Consider Fig. 3.2; strongly amphiphilic molecules of generic cylindrical or rectangular cuboid shape will spontaneously form a bilayer to minimize the contact of nonpolar regions with water. The driving force for this event may be counterintuitive at first glance: the bilayer is the arrangement that corresponds to the most disordered system. This may seem absurd because we tend to focus our attention in the solute (the lipids, in this case) and forget the solvent (water); yet, the gain in entropy refers to both. The lipids become ordered relative to each other, but the contact of hydrophobic groups with water molecules imposes restrictions to the orientational freedom of water, which is very costly in terms of entropy.

Broadly speaking, diacyl lipids (i.e., lipids containing two acyl—aliphatic—chains) with polar "heads", such as most of the lipids found in cell membranes, have a generic shape with the characteristics depicted in Fig. 3.2, and the entropic effect

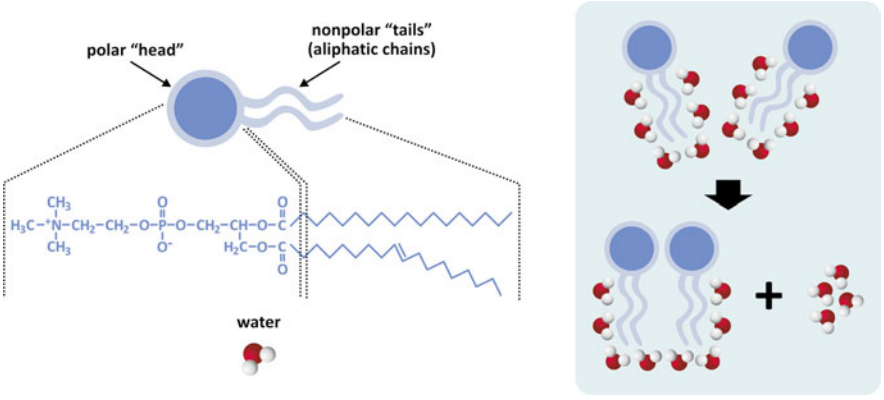


Fig. 3.2 Simplistic representation of two molecules of lipids having two aliphatic (hydrocarbon) chains and a polar "head." Polar heads usually contain phosphate (phospholipids) and other polar groups (*left panel*). Nonpolar regions of lipids tend to associate to each other because fewer molecules of water get exposed to the aliphatic chains (*right panel*). Exposed aliphatic chains force molecules to orient their oxygen atom away from the hydrocarbon constituents, i.e., they force a certain degree of order to the solvent. When two lipid molecules associate, less water molecules are forced to order and although the lipid molecules become more ordered relative to each other, the whole molecular systems (water included) becomes more disordered, in agreement with the entropic formulation of the second law of thermodynamics. Thus, this is named the entropic effect (occasionally imprecisely referred to as "hydrophobic effect"). The same principle applies when 3, 4, 5, …, *n* molecules form large assemblies of lipids, such as lipid bilayers (see Sect. 3.1.1)

compels these molecules to an ordered and parallel self-association. This is the physical process that sustains the stability of cell membranes as we shall see in Sect. 3.1.1. Now we will look more closely into the chemical structure of diacyl lipids: although these lipids tend to assemble as bilayers, the differences in their chemical nature determine which chemical processes they are able to participate in (e.g., cell signaling) and their tendency to associate to membrane proteins and/or form specific domains in the membrane. Most triacyl lipids occurring in nature, such as triacylglycerols (often referred to as triglycerides), lack a bulk polar "head" group and therefore do not display the propensity to form ordered supramolecular structures such as membranes (Fig. 3.3). Instead, they self-associate disorderly, forming aggregates, such as the lipid droplets found in cells at the site these lipids are synthesized (hepatocytes) or stored (adipocytes). Figure 3.3 shows histological images of the adipose tissue in which the presence of lipid aggregates occupying almost all cellular volume is detected.

Fatty acids can be considered as the basic unit common to most lipids in human cells. They consist of a linear unbranched aliphatic chain ("tail") with a carboxyl group (Fig. 3.4a). The aliphatic chain can be saturated (i.e., having all the carbon–car-

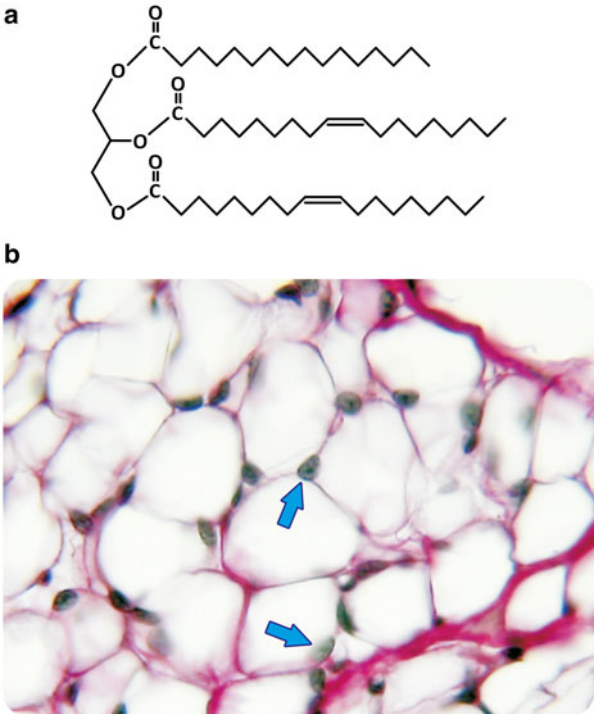


Fig. 3.3 Triacylglycerols lack bulk polar "head" groups, which keeps them from self-assembly into bilayers (**a**). Triacylglycerols aggregate in disordered agglomerated structures that can be detected in histological preparations of adipocytes (**b**). In panel (**b**) the lipids were extracted leaving white empty areas; arrows point to the nucleus. Figure reprinted with the permission of Instituto de Histologia e Biologia do Desenvolvimento, Faculdade de Medicina, Universidade de Lisboa, FMUL

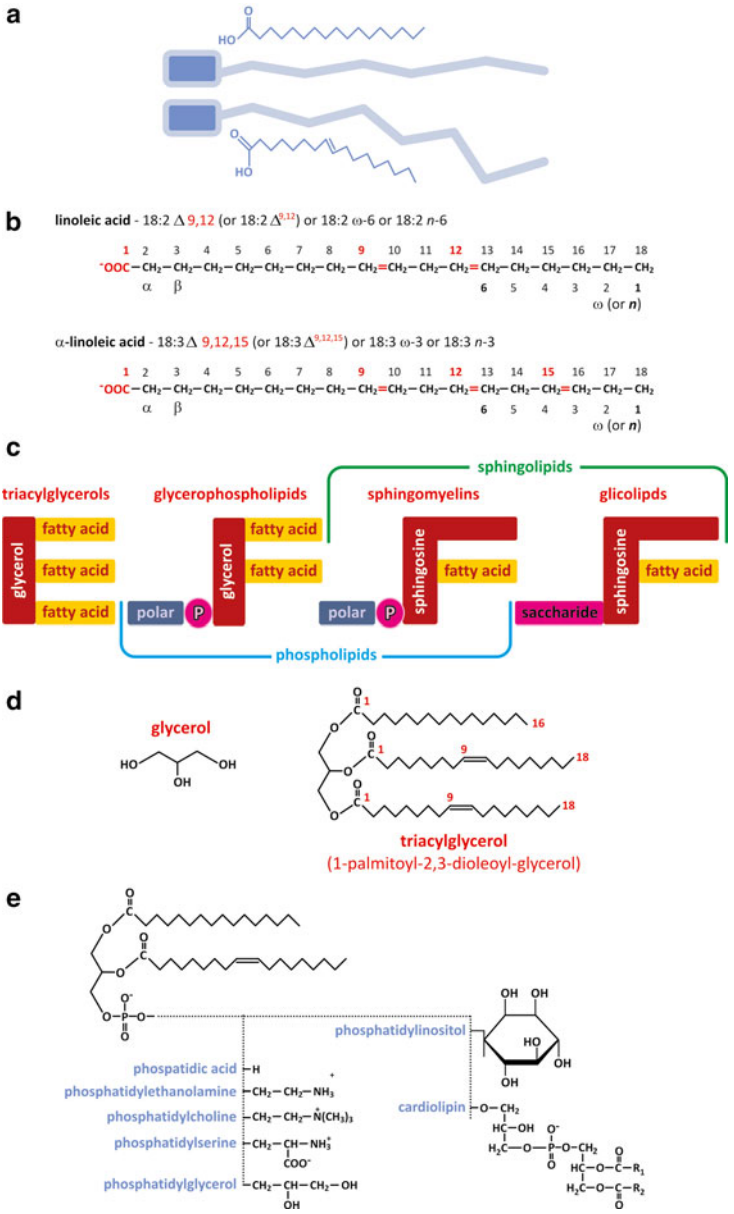


Fig. 3.4 (**a**) Pictoric representation of free fatty acids, highlighting the polar carboxylic group in *blue* and the nonpolar aliphatic chain in *gray*; double bonds impose constraints in the conformation of the chain, preventing stretching and rotation around the carbon skeleton, for instance. (**b**) Illustrative examples of the application of different nomenclature systems. Recent systems (numbering carbons from the carboxylic group) are more precise and descriptive, but older systems (Greek lettering and *ω*—numbering from the non-carboxylic ending carbon) are still widely used in some disciplines as nutrition. Trivial (nonsystematic) names (Table 3.1) were never abandoned for their "user-friendly" nature. (**c**) Fatty acids form ester bonds with other molecular structures to form more complex lipids such as triacylglycerols [detailed example in panel (**d**)], glycerophospholipids, or sphingolipids. Fatty acid residues are present in all. In sphingomyelins the polar group attached to phosphate is either choline or ethanolamine [panel (**e**)]. The example of specific triacylglycerol in (**d**) is 1-palmitoyl-2,3-dioleyl-glycerol. (**e**) Glycerophospholipids form different classes depending on the chemical nature of the "polar head" (fatty acid chains in this example are illustrative)

bon bonds reduced to simple bonds, without double or triple bonds) or unsaturated. Most fatty acids in humans have an even number of carbons, ranging from 10 to 28. Fatty acids are important molecules in the energetic metabolism of the organism, and they can be found either in "free" form (unattached to other molecules) or attached to other molecular structures through ester bonds (Fig. 3.4d, e). Free fatty acids differ from each other in the number and position of double bonds they have and in the length (in practice, the number of carbons) of the chains. The diversity attainable by changing these characteristics is virtually infinite, but, in practice, not all combinations are detected in nature (Table 3.1 displays examples of the most frequent fatty acids found in the human cells). Naturally, the nomenclature of fatty acids highlights their characteristic differences in terms of number and position of double bonds and chain length (Fig. 3.4b). The most recent nomenclature systems are very descriptive of these characteristics, although older, traditional nomenclature systems are widely used. The recent nomenclature identifies the carbon atoms in a chain by number, starting consecutively from the carbon of the carboxyl group (see Fig. 3.4b). Double bonds, if existent, are identified by the number of the carbon of lowest order forming the bond. The ω, or n, nomenclature numbers the carbons in reverse order (see Fig. 3.4b). Another nomenclature, still popular among many biochemists, identifies carbons by Greek letters in alphabetical order (α, β, γ, δ…) starting from the carbon adjacent to the carboxyl group (see Fig. 3.4b). A process of metabolic degradation of lipids is titled "β-oxidation" because of the importance of carbon β.

Glycerolipids such as diacylglycerols and triacylglycerols result from the chemical conjugation of a glycerol molecule with fatty acids through sterification (Fig. 3.4). Yet, glycerolipids may also conjugate other groups besides fatty acids. Usually a phosphate group is used to bridge glycerol to a polar group, forming glyc-

Table 3.1 Nomenclature of the most common fatty acids found in humans, both saturated (only simple carbon–carbon bonds) and unsaturated

Number of *C* (chain length)	Saturated	Unsaturated	Number double bonds	Nomenclature system		
				Δ	ω	n
14	Myristic					
16	Palmitic	Palmitoleic	1	16:1 Δ9	16:1 $\omega-7$	16:1 $n-7$
18	Stearic	Oleic	1	18:1 Δ9	18:1 $\omega-9$	18:1 $n-9$
		Linoleic	2	18:2 Δ9,12	18:2 $\omega-6$	18:2 $n-6$
		α-Linoleic	3	18:3 Δ9,12,15	18:3 $\omega-3$	18:3 $n-3$
		γ-Linoleic	3	18:3 Δ6,9,12	18:3 $\omega-6$	18:3 $n-6$
20	Arachidic	Arachidonic	4	18:4 Δ5,8,11,14	18:4 $\omega-6$	18:4 $n-6$
24	Lignoceric					

There are trivial names and descriptive abbreviations. Trivial names usually refer to the neutral (protonated) form of the acid, but one should bear in mind that at most physiological pHs (including plasma pH, 7.2–7.4), the acidic groups are unprotonated (pH above 4–5). Abbreviated nomenclatures highlight the number of carbon atoms in the aliphatic chain and number and position of unsaturated carbon–carbon bonds. Carbon numbering varies depending on adopting modern or older (ω, n) systems. Older systems are still widely used in some fields such as nutrition

erophospholipids (Fig. 3.4). The phosphate binds both the glycerol and the polar group through phosphodiester bonds, which are ubiquitous in biochemical processes for their importance in molecular structure and energetics.

Glycerophospholipids are named according to the fatty acids attached to glycerol and the chemical nature of the polar group (Fig. 3.4e).

Sphingolipids constitute another class of lipids present in many human cells, being responsible for rigid domains in the membranes. Figure 3.5 shows that although these lipids are similar in structure to glycerophospholipids, they are chemically different. These differences are responsible for the formation of rigid bilayers when mixed with phospholipids and cholesterol.

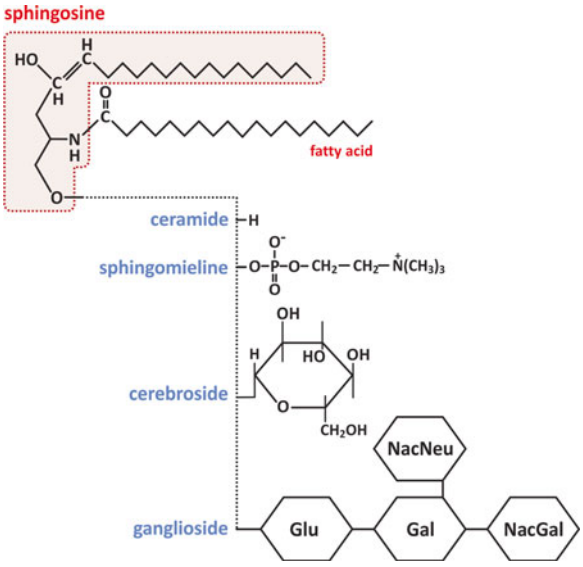


Fig. 3.5 Molecular structure of sphingolipids. As with glycerophospholipids, sphingolipids form different classes depending on the chemical nature of the "polar head" (the fatty acid chain in this example is illustrative). Glu - glucose, Gal - galactose, NacGal - N-acetylgalactosamine, NacNeu - N-acetylneuraminic acid

Sterols, of which cholesterol in an example (Fig. 3.6), are extremely hydrophobic alcohols. These molecules are characterized by a flat rigid system of carbon rings, cyclopentaphenanthrene, and have some remarkable properties:

1. Only one small polar group present (OH in cholesterol; C=O in the equivalent ketones).
2. A very flat molecule formed of four coplanar rings.
3. When placed in lipid bilayers, it acts as a fluidity regulator (not too rigid to compromise dynamics and not too fluid to compromise barrier integrity).

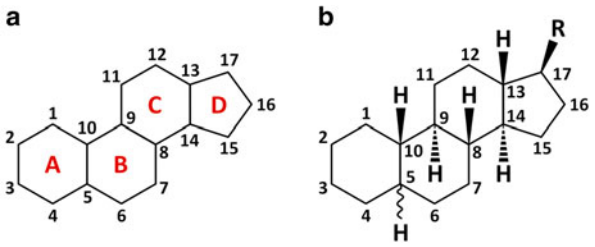


Fig. 3.6 Cyclopentaphenanthrene [panel (**a**)] is the basic structure of steroids [panel (**b**)], in which R is a generic group. The numbering system of carbons is highlighted. Cyclopentaphenanthrene is extremely rigid, planar, and hydrophobic. Sterols are a subfamily of steroids having an alcohol group in carbon 3 (see Table 3.2). Sterols can be found in the lipid membranes of animals, fungi, and plants but not bacteria. Sterols are the targets of some anti-infectious drugs in humans (this issue is readdressed in Fig. 3.9)

Cholesterol is a vital molecule to humans (Box 3.1). In addition to its properties in membranes, it takes part in other biochemical phenomena such as the synthesis of some hormones. Nevertheless, the seriousness of problems associated to its ingestion in excessive amounts in unhealthy diets, associated to its very low solubility in aqueous media, turns this important molecule into a potential killer. Curiously, although cholesterol is essential to many species in nature, including humans (Table 3.2), its synthesis is relatively recent in evolutionary terms as it uses molecular oxygen, which was only present in atmosphere after the advent of photosynthesis.

Box 3.1: Cholesterol: A Hero with Bad Reputation
High concentration ("level") of cholesterol circulating in the blood is a major risk factor for cardiovascular disease, which affects a significant fraction of the whole population in many countries. The campaigns to prevent cardiovascular disease are often centered in reducing dietary cholesterol, which gives the impression that cholesterol is some kind of poison or toxin that should be banned. The persistent unhealthy doses of cholesterol in diet are indeed harmful, but this must not create the illusion that cholesterol is by itself harmful to cells. Virtually all molecules in excessive doses are harmful and cholesterol is not an exception. Moreover, human body is so dependent on cholesterol that it synthesizes its own cholesterol and has homeostatic mechanisms to regulate its production in connection with several metabolic processes.

(continued)

Box 3.1 (continued)

The way the human body works at biochemical and physiological levels is crucially dependent on cholesterol. This extremely hydrophobic molecule intervenes in three major processes in humans:

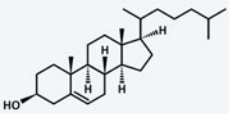
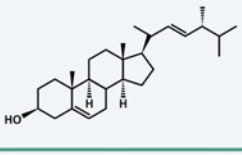
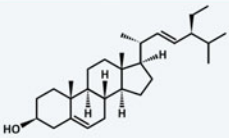
(a) It contributes to the balance of physical–chemical properties of lipid membrane dynamics in cells; this includes the plasma membrane and the intracellular membranes.
(b) It is a part of the bile acid synthesis; bile acids are important to absorb the lipids existing in food after ingestion.
(c) It is also a part of the synthesis of vitamin D and hormones as important as estrogen in women and testosterone in men.

Because of its hydrophobicity, the solubility in aqueous media such as blood is very low. There are specialized structures, lipoproteins, which incorporate cholesterol and form stable emulsions in aqueous medium. When cholesterol is not contained in these structures and/or lipoproteins deteriorate due to oxidation, depots are formed in the endothelial tissue that forms blood vessels. These depots are typical of atherosclerosis. In the most extreme cases, the blood vessels may be blocked and neighboring tissues are not irrigated. Lack of nutrients and oxygen (ischemia) may cause severe lesions in tissues. Vulnerable atherosclerotic plaques may also detach and clog other vessels. Any of these conditions is called infarction. Myocardial infarction, commonly known as a heart attack, is one example. Cerebral infarction is another example, commonly known as stroke.

Different lipoproteins transport cholesterol to different destinations (see Sect. 3.1.2). Cholesterol associated to HDL (high-density lipoproteins) can be disposed to form bile acids, whereas the cholesterol associated to other lipoproteins such as LDL (low-density lipoproteins) cannot. This has been referred to as good ("HDL-associated") and bad ("non-HDL-associated") cholesterol. The distinction simplifies medical communication to lay audiences but is absolutely nonsense from the biochemical point of view. Cholesterol synthesis, distribution, and transformation can be considered a whole and interconnected process in the human body.

The most eloquent demonstration of how cholesterol is indispensable is the Smith–Lemli–Opitz syndrome. This is a rare disease characterized by failure to thrive; mental retardation; visual problems; physical defects in hands, feet, and/or internal organs; increased susceptibility to infection; and digestive problems, among others. Smith–Lemli–Opitz syndrome is a genetic disease caused by a defect in cholesterol synthesis, namely, deficiency of the enzyme 3β-hydroxysterol-Δ7-reductase, the final enzyme in the sterol synthetic pathway that converts 7-dehydrocholesterol (7DHC) to cholesterol (the complete pathway of cholesterol synthesis is described in Box 8.8). This results in low plasma cholesterol levels and elevated levels of cholesterol precursors, including 7DHC.

Table 3.2 Naturally occurring sterols

name	structure	occurrence
cholesterol		animals
ergosterol		fungi
sitosterol		plants

3.1.1 The Structure of Biological Membranes

The concept illustrated in Fig. 3.2 can be extended to the association of many lipid molecules. The entropic effect will cause the molecules to self-associate orderly, ultimately forming bilayers of lipids with a hydrophobic core (Fig. 3.7). If these bilayers are extensive enough for the overall structure to bend, then the bilayer

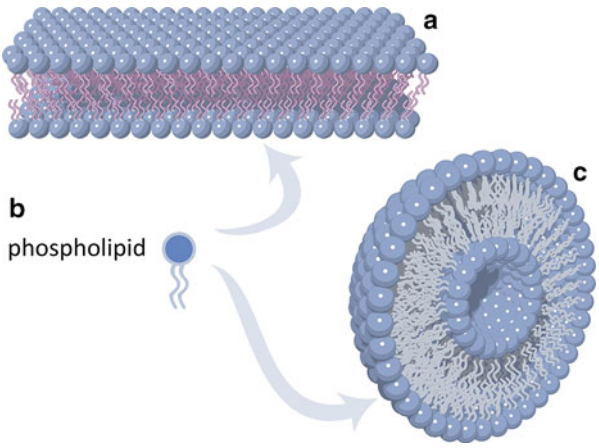


Fig. 3.7 Lipid molecules with a nearly cylindrical shape as a whole, such as phospholipids (**b**), tend to self-associate orderly due to the entropic effect (Fig. 3.2). Extensive self-association forms lipid bilayers (**a**) that ultimately bend and curve to self-seal into vesicles (**c**). Vesicles are the simplest prototype of cell membranes

curves allowing the sealing of the hydrophobic borders. All hydrophobic aliphatic chains become protected from water because the only areas in direct contact with the surrounding aqueous environment are the external and internal ("luminal") surfaces formed by polar head groups of the lipids. This is depicted in Fig. 3.7.

Lipid vesicles form spontaneously and they are the simplest models of biological membranes. Amazingly, there is not a single cell known in nature that does not have their membranes formed with lipid bilayers to some extent. Diversity arises from the different kinds of lipids used and from their combination with other molecules, but the structural arrangement of the membrane itself is configured by lipid bilayers. Some cells (e.g., bacteria, fungi, and plants) may also have a cell wall in addition to the cellular membrane, which serves for structural and/or protection purposes.

In fact, lipid membranes are relatively malleable and fragile. Yet, such malleability and apparent fragility are extremely important characteristics from a dynamic point of view: membrane division and fusion, for instance, are favored as these processes do not imply covalent bonds to be broken or formed among lipids, and lipid bilayers are flexible. Membrane fusion is advantageous in many biological circumstances (Fig. 3.8). In biotechnology, mainly in pharma and cosmetics, lipid vesicles are valuable tools due to their properties (Box 3.2).

Because lipid bilayers are the basic structure of cell membranes, several drugs directed to bacteria and fungi target the organization of lipid bilayers of these pathogens. Several antimicrobial peptides, for instance, are cationic so they bind and disrupt the lipid membranes of bacteria, which are highly anionic. Polyene antibiotics such as nystatin B and amphotericin B bind specifically to ergosterol disrupting the selective membrane permeability and ultimately causing the lysis of fungi (Fig. 3.9) because ergosterol only exists in the membrane of fungi.

While lipid vesicles depicted in Fig. 3.7 are formed only by lipids (a single pure lipid or a mixture of lipids), biological membranes are often composed of lipids, proteins, and saccharides. How these components organize in the membrane has been the subject of intensive scientific research over the years (Fig. 3.10). Nowadays, a biological membrane of a human cell is regarded as a lipid bilayer having on heterogeneous distribution of lipids both in each layer and among layers. This heterogeneity leads to the formation of specific domains of lipids having defined functions. Rigid platforms, for instance, may serve to anchor proteins on the membrane (Fig. 3.11). The cell membrane is directly connected to the cytoskeleton through an array of proteins, the so-called cytoskeleton anchors. The outer surface of cell membranes may have lipids and proteins that are glycosylated (i.e., covalently attached to saccharides) contributing to a rich chemical diversity on the surface of cells (Fig. 3.10).

The lipid composition of cell membranes varies a lot from species to species, from organelle to organelle in the same cell, and from the inner leaflet to the outer leaflet in the same membrane (Fig. 3.12). This variety of compositions grants the necessary diversity to membranes so that they are specific for certain functions, in spite of the common feature of all membranes of all cells: in the end they are all constructions based on lipid bilayers.

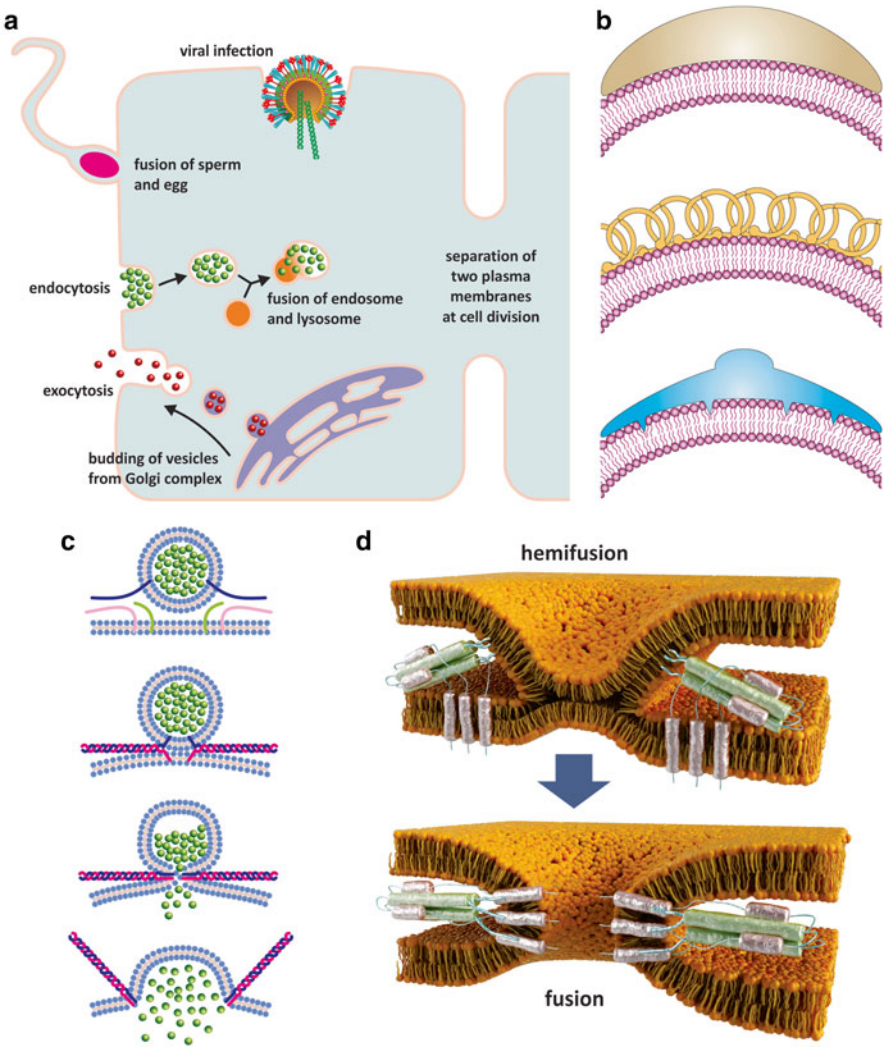
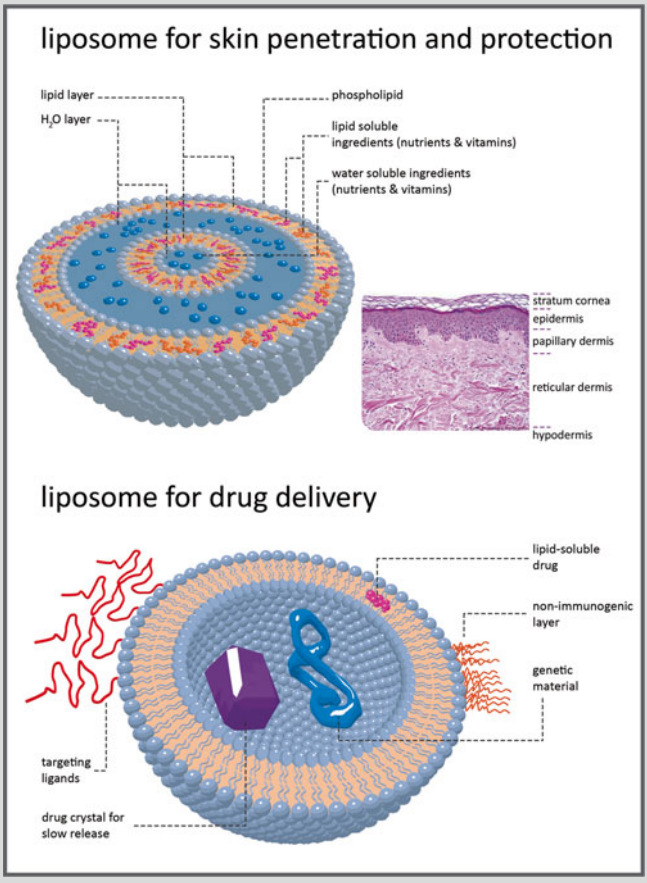


Fig. 3.8 Lipid membrane fusion occurs in nature associated with a plethora of cellular events (**a**). Some of these events such as endocytosis or budding of vesicles from the Golgi complex require inducing a curvature in membranes so that vesicles are formed. This induction is accomplished by the action of specialized proteins (**b**) that are curved and adhere to the surface of lipid bilayers (**b**, *top*), stabilize spontaneous curvatures (**b**, *middle*), or insert into one leaflet of the bilayer only thus forcing the membrane to bend (**b**, *bottom*) (reproduced with permission from Zimmerberg & Kazlov, Nature Rev Mol Cell Biol 7: 9–19, 2006). Membrane fusion implies that two different bilayers come together (**c**, **d**), which is due to the action of proteins that insert the two bilayers and promote conformational changes that lead to the contact of the lipid bilayers. First, a common bilayer is formed from the mix of the two membranes (hemifusion state), and then total fusion merges the two entities that were initially enclosed in their own membranes. Panel (**c**) shows the example of membrane fusion during neurotransmission, namely, a neurotransmitter being released at a synapse. The proteins responsible for fusion are named SNAREs (represented by *green*, *pink*, and *blue filaments*). Panel (**d**) shows the details of hemifusion and fusion of an envelope virus such as influenza virus, HIV, or dengue virus with the target cell. In this case viral proteins at the surface of the virus insert the membrane of the target cell and undergo conformational changes ending in fusion and consequent release of viral contents in the cytosol

Box 3.2: Lipid Vesicles in Pharma and Cosmetics

For many decades, lipids were considered relatively inert biologically, with functions of storage of energy in the adipose tissue or constitution of a matrix for cell membranes. Thus, in general, there was little interest in research to discover the properties, structures, biosynthetic pathways, biological utilization, and other functions of lipids. In the present, the situation is completely opposite. Lipids are regarded as important biological molecules that in addition to being energy stores and membrane components, participate in the regulation of many biochemical processes in cells and endocrine physiological regulation in the human body. Moreover, lipids are now important tools in pharma and cosmetic industries because they can be used in formulations that distribute and deliver drugs or other biologically active molecules in the human body. In most of these formulations, the lipids self-assemble in bilayers that form extensive vesicular systems, liposomes (see figure), able to encapsulate molecules having desirable functions. Liposomes are very versatile systems as hydrophobic molecules may be accumulated in the lipid areas, and hydrophilic molecules may remain solubilized in the aqueous spaces inside vesicles.

In cosmetic applications, the liposomes may be part of formulations to be applied topically in skin. The lipids help in the diffusion of the formulation through the outer layers of skin. Moreover, the simple fact that lipids and water are forming an emulsion will help the formulation to hydrate the desired areas of the skin.



(continued)

Box 3.2 (continued)

Liposomal encapsulation can substantially improve the action of a drug, such as the decreased toxicity observed with amphotericin B (see Fig. 3.9). Conventional amphotericin B has been generally considered the drug of choice for many types of systemic fungal infections. These infections are a major threat to those whose immune systems are compromised, such as patients undergoing chemotherapy for cancer, bone marrow transplant recipients and AIDS patients. However, amphotericin B is very toxic, thus limiting its utility. For these patients, who have a high rate of morbidity and mortality, there is a dosage form distinct from conventional amphotericin B, which consists of amphotericin B complexed with two phospholipids in approximately a 1:1 drug-to-lipid molar ratio: L-α-dimyristoylphosphatidylcholine (DMPC) and L-α-dimyristoylphosphatidylglycerol (DMPG), present in a 7:3 molar ratio. Doxorubicin is another example. Liposomal doxorubicin is designed to target to tumor cells and spare healthy tissue, maintaining efficacy while reducing toxicity. Conventional doxorubicin, drug commonly used to treat cancer, is limited by its potential for causing a variety of severe side effects, particularly irreversible heart damage.

Researchers are developing innovative liposomes with refined drug delivery properties to be part of future medicines. Some have their surface modified with proteins and other selected polymers to target selected cells. Synthetic phospholipids are suitable for specific applications in liposome targeting and gene therapy. Gene therapy is based on the efficient delivery of genes to their intended targets. Researchers have successfully put DNA into liposomes and have achieved fusion of these liposomes to cells. Scientists have also succeeded in protecting these liposomes from degradation and are able to modulate their circulation time.

Fig. 3.9 (continued) sides of the polyene rings face the lipids and the polar sides face each other forming a hydrophilic pore (**a**, *right*). Filipin is not so selective for ergosterol when compared to cholesterol. Amphotericin B, therefore, it is more toxic to human cells. Panel (**b**) (*right*) shows atomic force microscopy images of pores created by filipin in cholesterol-containing bilayers (*arrows*). Large structures in the surface of the lipid bilayer (**b**, *center*) are filipin aggregates. Panel (**c**) *left* also shows atomic force microscopy images but of an individual bacterium (*E. coli*). Exposure of the bacterium to the cationic amphipathic peptide BP100 caused the collapse of the bacterial membrane. The bacterial membrane is anionic, thus attracting electrostatically the peptides, which then aggregate on the lipid bilayer causing perturbation and increasing permeability. This perturbation may be caused by formation of pores or by unspecific destruction of the lipid organization (**c**, *right*, shows the action of $rBPI_{21}$, a peptide derived from bactericidal/permeability-increasing protein potentially useful against meningitis). (Figures in panel (**b**) *right* are reprinted with permission from Santos et al., Biophys J. 75:1869–1873, 1998. Figures in panel (**c**) *left* are reprinted with permission from Alves et al., J. Biol. Chem. 285:27536–27544, 2010. Figures in panel (**c**, *right*) are reprinted with permission from Domingues et al., PLoS ONE, 4:12, e8385, 2009

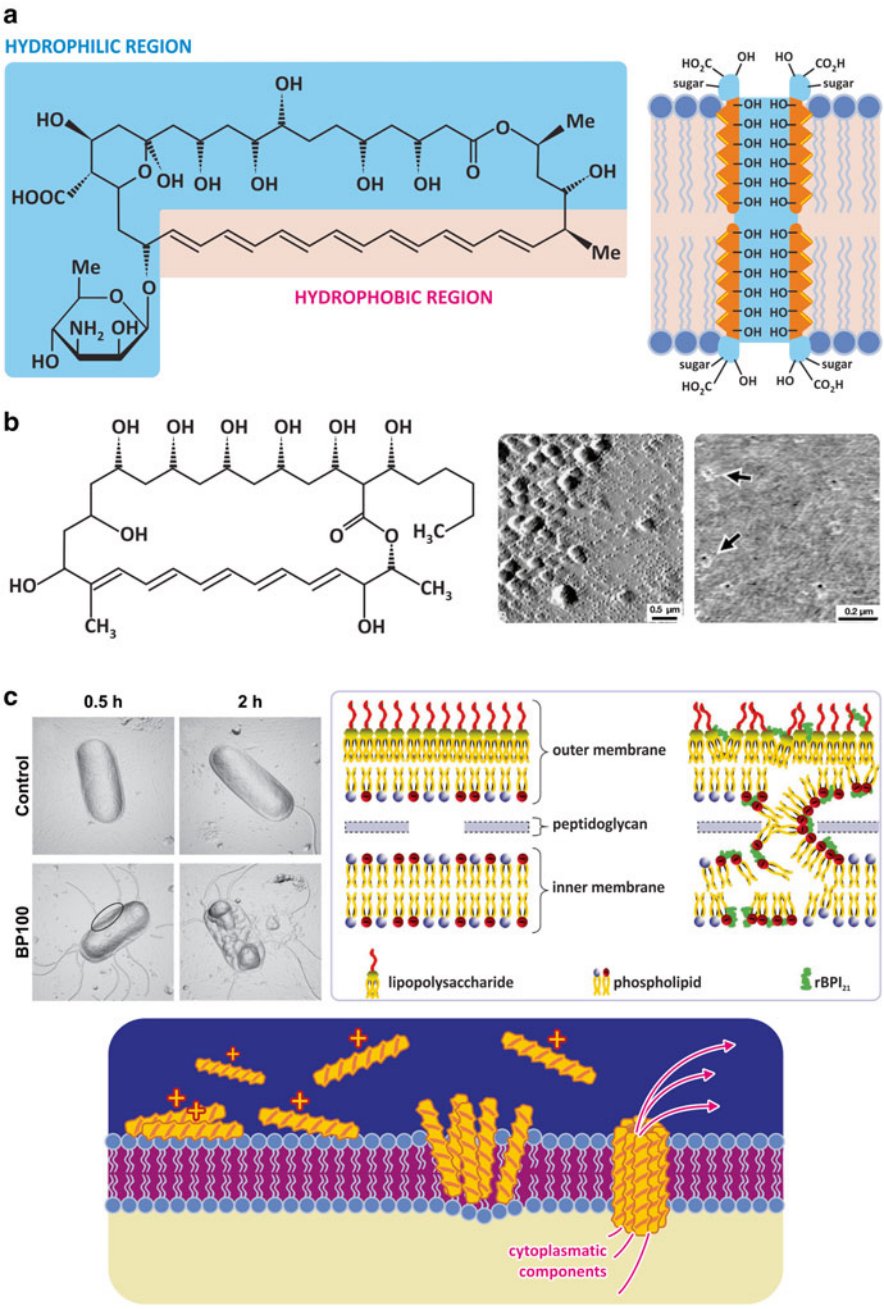


Fig. 3.9 Some drugs such as amphotericin B (**a**), a fungicide used in the treatment of infections with *Candida* sp. among others, or filipin (**b**), a fungicide also toxic to human cells and therefore not used in therapies, and cationic amphipathic peptides with antibacterial properties (**c**) target cell membranes. Amphotericin B binds to ergosterol forming ordered complexes in which the hydrophobic

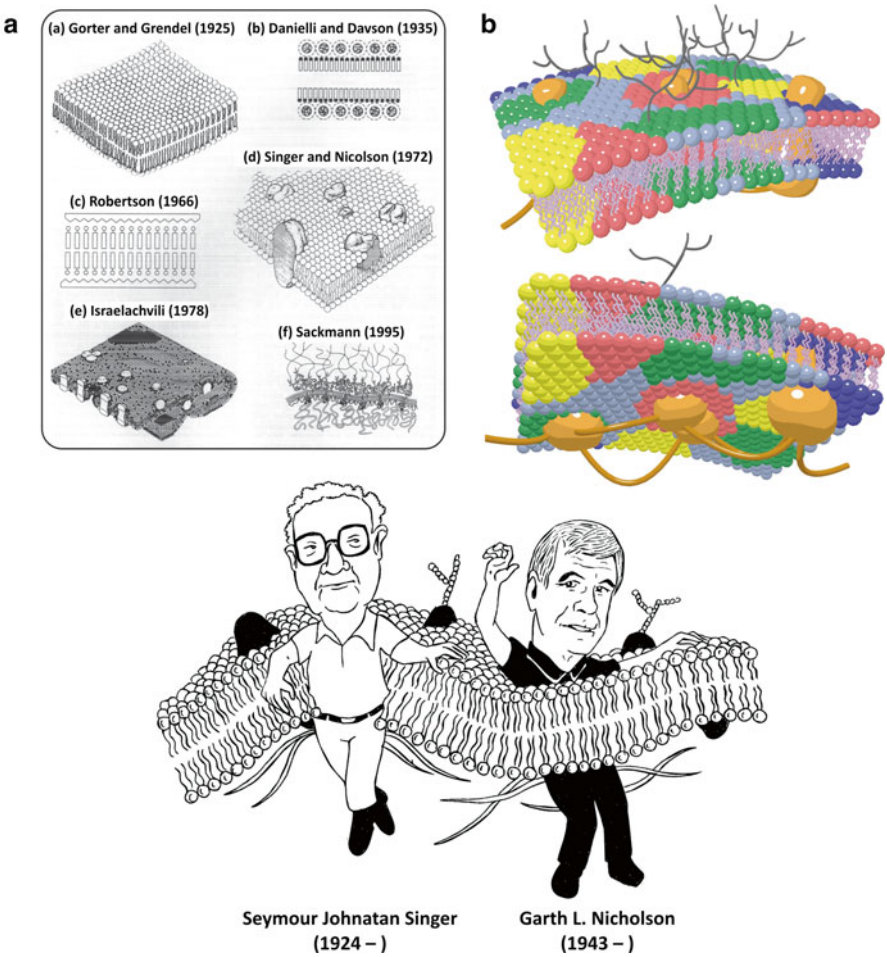


Fig. 3.10 (**a**) Historical evolution of the concept of cell membranes (according to Ole Moritsen; reprinted with permission from Biol. Skr. Vid. Selsk. 49:7–12, 1998). The lipid bilayer was first identified, followed by the discovery that proteins interact with lipids. The concept that there are integral proteins embedded in the lipid bilayer forming a dynamic structure came with the fluid mosaic model by Singer and Nicolson in 1972, which is still accepted as the basic framework of membrane structure. Nevertheless, from then on, the organization of cell membranes has been continuously unraveled. Lipids are now known to self-associate in lateral domains of different composition, and some membrane proteins are bound ("anchored") to the cytoskeleton (A(e) and A(f)). The modern view on cell membranes is schematized in (**b**), in which lipid colors represent heterogeneous lipid compositions. In the outer surface of cell membranes (**b**, *top*), glycosylated (i.e., saccharide-containing, *black*) lipids and proteins are present with different functionalities (see Sect. 3.2). Cytoskeleton and cytoskeleton-binding proteins are represented in *brown*

Fig. 3.11 (continued) segregated areas are clear, and the height profile along the line seen in the *top image* (*bottom graph*) shows that the segregated areas are higher, therefore corresponding to more rigid areas of the membrane. Epifluorescence images of lipid bilayers having one of the lipid tagged with a fluorescent dye confirm segregation of both lipids (**b**, *right*). Figure reprinted with permission from Franquelim et al., J. Am. Chem. Soc. 130:6215–6223, 2008; and Franquelim et al., Biochim. Biophys. Acta. 1828:1777–1785, 2013. These more rigid areas of the membrane are more adapted to anchor proteins and membrane receptors (**c**). Sphingolipids and cholesterol, for instance, further enhance these characteristics. Proteins covalently attached to lipids are typically found in these rigid platforms

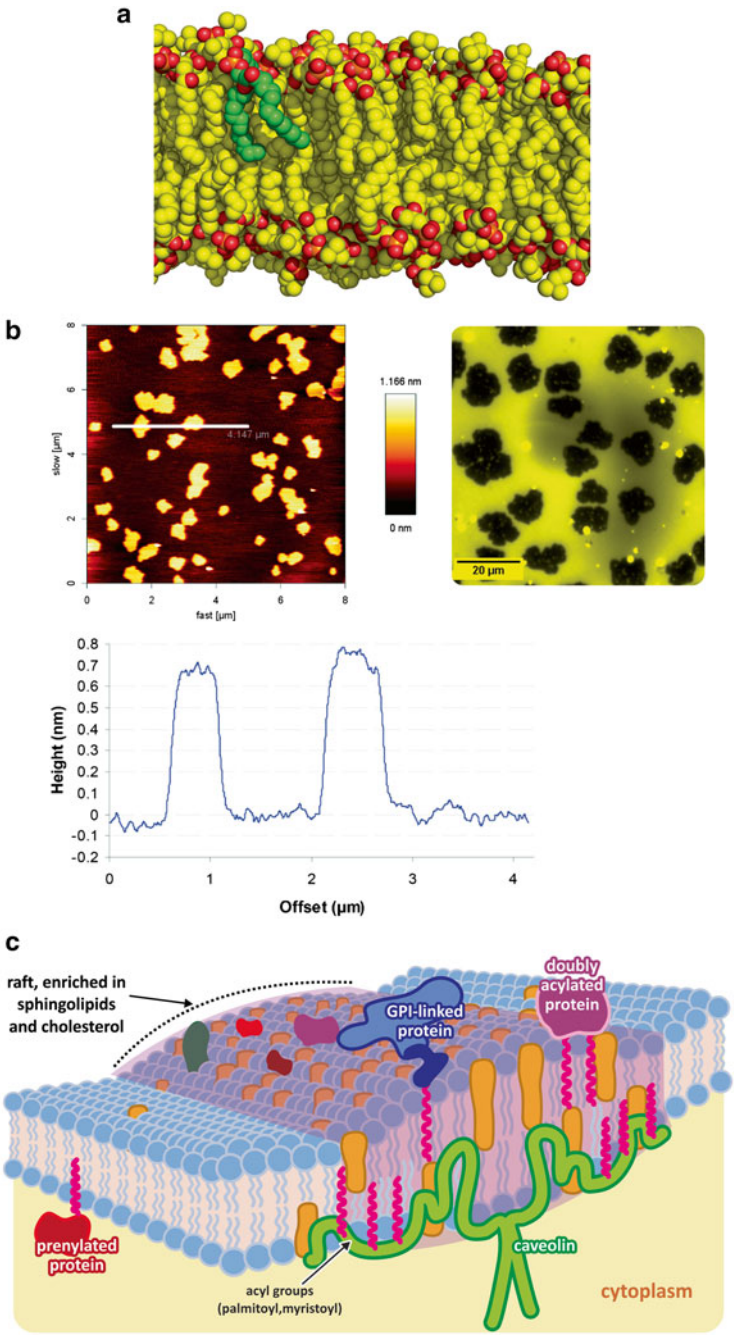


Fig. 3.11 Heterogeneity in lipid membranes. Although phospholipids are frequently depicted in an oversimplified form with ordered stretched aliphatic chains, as in Fig. 3.2, in reality most molecules in lipid bilayers in physiological conditions in cells have very flexible and dynamic acyl chains. Panel (**a**) shows molecular dynamics simulation of lipids with one single lipid in *green* to highlight the bent conformation (courtesy of Dr. Claudio Soares, ITQB-UNL, Portugal). Lipids with longer and saturated chains adopt stiffer and linear conformations as they interact more tightly with each other. Mixing stiff and fluid lipids results in partial segregation of the lipids. The atomic force microscopy image of a mixture of a fluid unsaturated lipid (palmitoyl-oleyl phosphatidylcholine, POPC—50 % molar) with a saturated rigid lipid (dipalmitoyl-phosphatidyl choline, DPPC) is shown in panel (**b**, *left*);

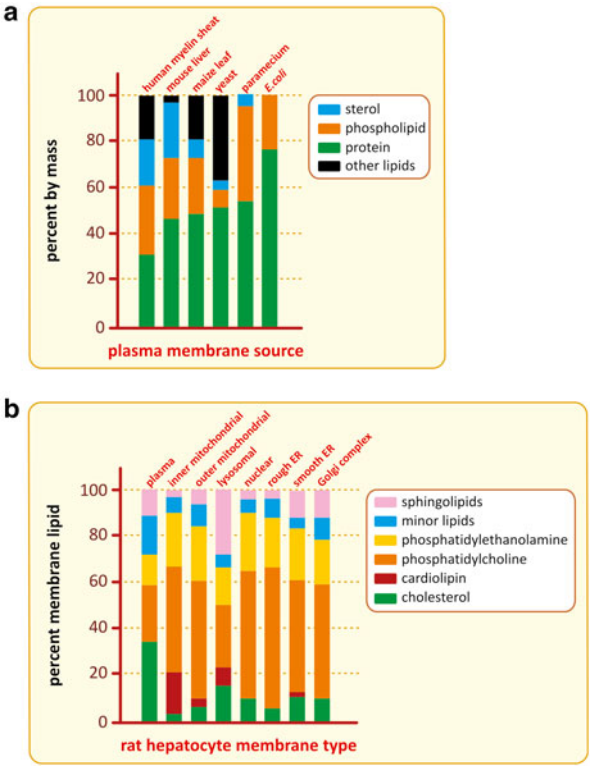


Fig. 3.12 Although all membranes in cells in nature are constructions based on lipid bilayers, the membrane composition varies a lot among species (*top*), among organelles of the same cell (*bottom*), and even between the two leaflets of the same membrane (not shown)

Because lipid bilayers are hydrophobic barriers, hydrophilic molecules such as glucose cannot freely transverse them, but channels and transporters in their membranes overcome the limitation. Channels and transporters are proteins specific for certain molecules or ions that facilitate or enable the translocation of such molecules or ions across the membranes. This subject will be revisited in Sect. 5.3.1, after careful consideration of protein structure.

3.1.2 The Structure of Lipoproteins

Lipoproteins are organized assemblies of lipids and proteins covering a wide range of sizes and densities (Fig. 3.13). These assemblies have a lipidic core formed mainly by triacylglycerols and cholesterol esters, covered by a monolayer of phospholipids. The global arrangement is largely determined by the entropic effect as the monolayer of phospholipids minimizes the contact between apolar components and water molecules in blood (Box 3.3). Lipoproteins transport lipids between different tissues. There are proteins inserted in the lipid layer, exposed at the surface of the lipoprotein, as shown in Fig. 3.14.

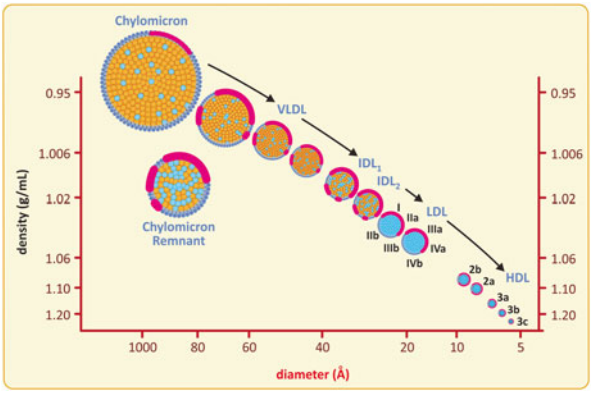
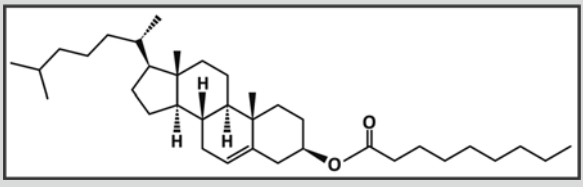


Fig. 3.13 Lipoproteins are grouped according to their densities and sizes, although the most common nomenclature refers to density (*HDL* high-density lipoprotein, *LDL* low-density lipoprotein, *VLDL* very-low-density lipoprotein). Chylomicra are the structures formed with dietary lipids in the enterocyte (intestines) and released in plasma

Box 3.3: Lipoproteins: The Burden of Lipid Transport

Lipids have extremely low solubility in aqueous media. Therefore, as a consequence of the entropic effect (see Sect. 3.1), when placed in aqueous medium, they tend to self-associate. Most phospholipids, having a hydrophilic "head" and two hydrophobic acyl chains, tend to pack side by side and form bilayers. Cholesterol is not prone to form very organized supramolecular assemblies itself but is able to insert in the lipid bilayers and contribute to its stability. Triacylglycerols ("triglycerides") and esters of cholesterol (cholesteryl esters; see figure) do not have the amphipathic properties and structural requirements to form lipid bilayers. Instead, triacylglycerols and cholesteryl esters amalgamate in an aggregate having no polar surface. These aggregates tend to be spherical, the geometry that minimizes the surface area exposed to the solvent. The lipid aggregates tend to grow until free lipid is nearly absent unless a phospholipid monolayer covers the surface of these aggregates, forming an entropically favorable interface. The phospholipid monolayers stabilize the lipid aggregates and an emulsion is formed. Emulsion means the lipids are heterogeneously distributed in microscopic scale because the lipids are clustered in aggregates but homogeneously distributed in macroscopic scale since the aggregates are evenly disseminated in the solvent.



Example of a cholesteryl ester: cholesteryl nonanoate. The molecule is composed of a cholesterol moiety and a fatty acid moiety (nonanoic acid in this case)

Box 3.3 (continued)

In human body, very large lipid aggregates are found in the cells of the adipose tissue (adipocytes), occupying almost all cytoplasmic space. It is a storage place. Smaller aggregates are found emulsified in blood, in association with specific proteins. These smaller aggregates are dragged by blood and serve as lipid transporters. In both cases, the aggregates are covered by monolayers of phospholipids having the polar groups exposed to aqueous environment and the acyl "tails" in contact with the lipids. The ensemble formed by the emulsified lipid aggregate covered with a phospholipid monolayer and associated to specific proteins is named lipoprotein (see Fig. 3.14 in the main text). The proteins themselves are named apolipoproteins. The lipoprotein as a whole is the lipid carrier entity; apolipoproteins bind to specific receptors so that lipids are delivered to target cells only.

Lipoproteins vary among them in the proportion of triglycerides/cholesteryl esters/protein, which directly interferes in their compactness and density. Early studies on the properties of lipoproteins achieved separation of several classes of lipoproteins based on their different densities, so the density-based nomenclature was naturally adopted, from high-density lipoproteins (HDL) to very-low-density lipoproteins (VLDL) and chylomicra (see below table). There is a concomitant change in volume but it is not given importance regarding lipoprotein classification. It is also worth highlighting that apolipoproteins are also divided in classes (A, B, C, etc.) and HDLs are the only lipoproteins not bearing apolipoproteins B, which is a distinctive feature.

Properties of plasma lipoproteins

Plasma lipoproteins	Density (g ml^{-1})	Diameter (nm)	Apolipoprotein	Physiological role
Chylomicron	<0.95	75–1200	B48, C, E	Dietary fat transport
Very-low-density lipoprotein	0.95–1.006	30–80	B100, C, E	Endogenous fat transport
Intermediate-density lipoprotein	1.006–1.019	15–35	B100, E	LDL precursor
Low-density lipoprotein	1.019–1.063	18–25	B100	Cholesterol transport
High-density lipoprotein	1.063–1.21	7.5–20	A	Reverse cholesterol transport

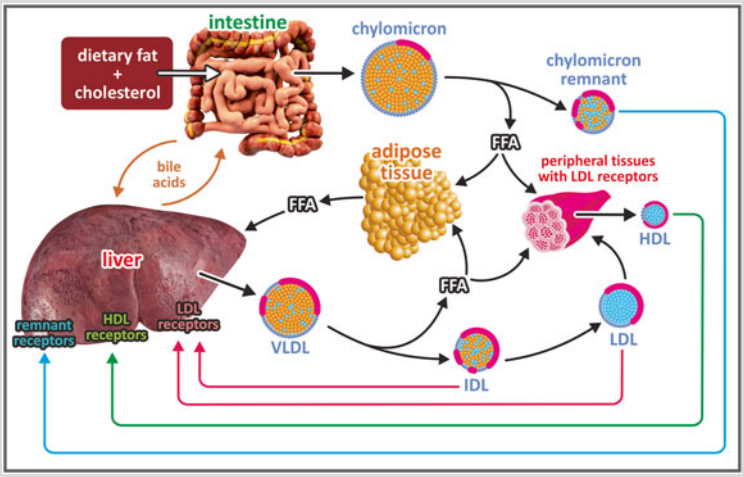
TAG triacylglycerol, *CE* cholesterol ester, *C* free cholesterol, *PL* phospholipid, *P* protein

During digestion, the lipids are partially degraded and emulsified in the intestinal lumen by bile acids, molecules similar to cholesterol but having several polar groups. Lipids and other nutrients are taken up by the intestinal

(continued)

Box 3.3 (continued)

cells, enterocytes (see figure). Chylomicra are formed in these cells and released in the blood. Eighty to ninety percent of the lipids in chylomicra are triglycerides, which account for their low density. The remaining cargo is free cholesterol (1–3 %), cholesteryl esters (3–6 %), and phospholipids (7–9 %). Chylomicra circulate in the blood, where degradation of their triacylglycerides into free fatty acids occurs. Part of these fatty acids is delivered to adipose tissue and peripheral tissues. The chylomicra remnants bind to liver cells having specific receptors. The excess of nutrient incorporated after digestion is converted to lipids in the liver, where they form VLDLs in a process similar to the assembly of chylomicra in the intestine. VLDLs are released from the liver cells. In the bloodstream, they are depleted of free fatty acids from triacylglycerols, resulting in the intermediary density lipoproteins (IDLs) and low-density lipoproteins (LDLs), which transfer lipids to peripheral tissues bearing LDL receptors. Liver cells having LDL receptors also bind LDL. In contrast, HDLs transport cholesterol from peripheral tissues to the liver, where there are cells having specific receptors for HDL. The cholesterol is then used to synthesize bile acids. HDLs are the only lipoproteins that dispose of cholesterol. This characteristic renders the name "good cholesterol" to HDL-associated cholesterol in public health campaigns for lay audiences. This name makes no sense on biochemical grounds but helps to spread the message that in cardiovascular risk evaluation, it is important to differentiate between cholesterol that is being removed and cholesterol that is being incorporated. Interestingly from the biochemical point of view, cholesterol to be disposed is associated to lipoproteins having no apolipoprotein B and cholesterol to be incorporated is associated to apolipoprotein B-containing lipoproteins.



Origin and fates of plasma lipoproteins

(continued)

Box 3.3 (continued)

LDLs are degraded inside the cells after being taken up by endocytosis. The LDL receptor is segregated in the endocytic vesicle, which then divides in two: one empty vesicle having the receptors in the membrane returns to the surface of the cell, and the other vesicle has the proteins and lipids of the lipoproteins and fuses with the lysosome. LDL-derived cholesterol may then be used either in cell membranes, or as a precursor to synthesize steroid hormones or bile acids, or simply be stored as cholesteryl esters. The exact destination of cholesterol in the cell depends on the type of cell and its metabolic state. Dietary cholesterol suppresses the synthesis of cholesterol by the body, and high free cholesterol levels inhibit the synthesis of LDL receptors. Cellular uptake is thus inhibited in the presence of excess cholesterol, and the level of LDL in the blood increases. Moreover, the LDLs take more time to be uptaken and circulate in blood for longer periods. This increases the chances of having the LDL exposed to oxidative agents such as NO, hydrogen peroxide, or the superoxide ion. Oxidized LDLs are then removed from circulation by macrophages, but the macrophages get their properties severely altered, becoming the so-called foam cells. Foam cells accumulate in the walls of endothelia, releasing growth factors and cytokines that stimulate the migration of smooth muscle cells that proliferate in the site of accumulation of foam cells and form collagen matrices. This consists of the deposition of atherosclerotic plates, which poses a severe cardiovascular risk.

Interestingly, the proteins responsible for triacylglycerol conversion into fatty acids in the heart (enzymes heart lipoprotein lipases) have a much higher affinity for triacylglycerols than the corresponding proteins in the adipose tissue. The affinity parameter, $1/K_M$, which will be addressed in Sect. 4.2.1 is about tenfold higher in the heart. During starvation the levels of plasma triacylglycerol drop, but delivery of fatty acids from triacylglycerols is kept in the heart even when suppressed to the adipose tissue.

Different classes of lipoproteins differ in density (due to differences in the relative amounts of proteins, phospholipids, cholesterol, triacylglycerols, and cholesterol esters in their composition), size, and specific proteins associated (see Table in Box 3.3). Nevertheless, these classes are named after the differences in density only, which relates to the most practical property that can be used for their separation in different fractions (see Fig. 3.13). Lipoproteins formed in the intestine with dietary lipids are known as chylomicra, and the remaining classes range from very-low-density lipoproteins (VLDLs) to high-density lipoproteins (HDLs). Intermediate-density lipoproteins (IDLs) and low-density lipoproteins (LDLs) are in between.

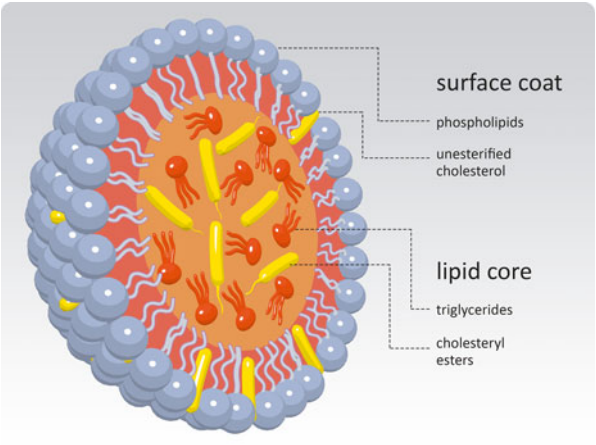


Fig. 3.14 Global generic structure shared by all classes of lipoproteins. A monolayer of phospholipids and cholesterol covers a lipid droplet of triacylglycerols and cholesterol esters. There are proteins at the surface (not shown), the apolipoproteins, which are specific of each class of lipoproteins and serve for cell recognition, i.e., interaction with specific receptors in cells

The different classes of lipoproteins have different functions and different target tissues for their action (see Box 3.3). Target recognition depends on the specificity of the proteins present on the lipoprotein surface (referred to as apolipoproteins to highlight that only the proteic part is being addressed) for well-defined receptors.

Other lipidic supramolecular assemblies, such as lipid droplets in the liver and lipid depots in adipose tissue, exist. Lipid depots may occupy most of the volume of an adipocyte, the adipose tissue cells (see Fig. 3.3), confining the cytoplasm to a very small fraction of the cell. These lipid depots are also covered by phospholipids that are stabilized as a consequence of the entropic effect.

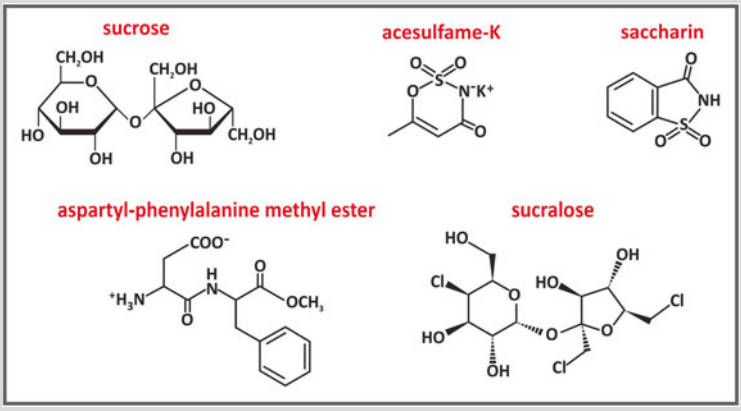
3.2 Saccharides and Their Polymers and Derivatives

Saccharides, different from lipids are extremely polar, therefore, hydrophilic, molecules. They are linear aldehydes or ketones with hydroxyl groups bound to the carbons that do not form the carbonyls (C=O). Many of these molecules have only C, H, and O in their composition and fit the formula $(CH_2O)_n$. This spurious characteristic consecrated the designation "carbon hydrate", which is still widely used to identify saccharides despite its total inadequacy in chemical terms: not all saccharides obey to $(CH_2O)_n$ and this does not reflect an hydration of carbon, only a specific molar proportion between C, H, and O atoms. Referring to saccharides as "sugars" is equally inadequate and misleading. "Sugar" is related to a property, sweetness, which not all saccharides possess and extends to molecules other than saccharides, such as peptide sweeteners (Box 3.4). Saccharides

Box 3.4: Sweeteners and Sugar Substitutes

The problem of popularization of high caloric diets stimulated the search for sugar substitutes. Sucrose, the most commonly used sugar in cooking, is a natural sweet molecule from which a certain amount of energy can be used by the human body after metabolization. However, there are molecules known as high-intensity sweeteners that have manyfold the sweetening power of sucrose. Saccharin, for instance, is approximately 300-fold sweeter than sucrose when equal quantities are compared. Aspartame and acesulfame-K are approximately 200-fold sweeter than sucrose. For sucralose the ratio raises to an impressive 600-fold. A specific chemical modification in aspartame, advantame, grants an impressive 20,000-fold increase in sweetness relative to sucrose. Therefore, much less mass of sweetener is needed to achieve the sweet taste of a food or beverage. Even though the "caloric content" of a unit mass of the molecule may be equivalent to sucrose in some cases, the total amount used is several orders of magnitude less and the total calories in the diet drops drastically.

The chemical structure of sucrose and most artificial sweeteners are very different. Sucrose is a disaccharide composed of the residues of the monosaccharides glucose and fructose. Sucralose is prepared from sucrose via the substitution of three hydroxyl groups for three chlorides. Saccharin and acesulfame-K have much different structures. Aspartame is the methyl ester of the dipeptide L-aspartyl-L-phenylalanine.



The molecular structure of sweeteners must be such that they bind to a specific receptor molecule at the surface of the tongue. The receptor is coupled to a G protein (see Sect. 5.4), which dissociates when the sweetener binds to the receptor. This dissociation leads to the activation of an enzyme

(continued)

Box 3.4 (continued)

that triggers a sequence of events resulting in signals that are carried to and interpreted by the brain. The sweetness perception depends on fine details of the interaction between receptor and sweetener. The importance of molecular shape to sweetness is illustrated by the case of aspartame, as its stereo isomer, L-aspartyl-D-phenylalanine methyl ester, has a bitter, not a sweet, taste.

There has been a long and continuous controversy about the impact of artificial sweeteners on health, which has driven a lot of research on the possible toxicity of their metabolic products. Saccharin has been very controversial and banned in some countries. In the body, aspartame is broken down into/absorbed as products that include aspartate, phenylalanine, and methanol. Phenylalanine is toxic to individuals with phenylketonuria, a genetic disease wherein individuals cannot process phenylalanine. Products containing aspartame must therefore be labeled for phenylalanine. Aspartame's breakdown products—phenylalanine and aspartate, as well as methanol and its breakdown products formaldehyde and formate—are a matter of debate. Regardless of the controversies and limitations in their use, artificial sweeteners have an important role in the improvement of the quality of life of diabetics, who are limited in the consumption of sucrose and other saccharides.

or oses are therefore the preferred nomenclatures for biochemists although "carbon hydrates", or carbohydrates, in short, and "sugars" are commonly used in closely related areas such as nutrition, for instance.

Saccharides are the most abundant biomolecules and owe this ubiquity to their reactivity and structural plasticity, which enable a great variety of functionalities, including energetic storage, cell communication, and cell protection against mechanical aggressions and dehydration. In order to understand such structural plasticity and the functionalities arising from it, one has to start with the basic chemistry and reactivity of saccharides. Although this is a wide and complex area in the realm of biochemistry, we will devote ourselves to the understanding of the most important saccharides in human biochemistry only. One will stick to the basics of this fascinating world for the sake of clarity and focus on processes that are foundational for other medical disciplines such as histology and physiology.

The simplest conceivable saccharides have three-carbon chains, i.e., they are trioses ("tri" for three carbons, "ose" for saccharide). Depending on the position of the carbonyl group, C=O, which may be terminal (aldehyde) or not (ketone), the saccharide is an aldose or a ketose. In the specific case of trioses, only glyceraldehyde and dihydroxyketone are possible (Fig. 3.15). But even in these cases, two kinds of common chemical reactions in nature are possible: esterification and reduction (Fig. 3.16).

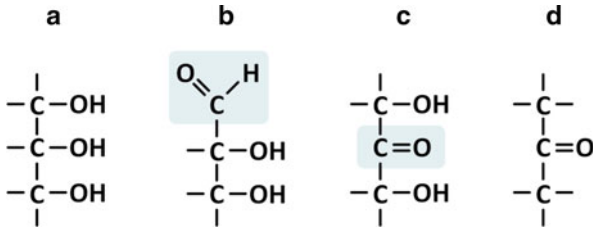


Fig. 3.15 Glycerol [(**a**); also in Fig. 3.4d] is related to an aldehyde, glyceraldehyde (**b**), which has a ketone isomer, dihydroxyketone (**c**), which in turn is related to a ketone (**d**), dimethyl ketone. The aldehyde and ketone groups of glyceraldehyde and dihydroxyketone are highlighted in the chemical structures (**b**) and (**c**), respectively, by a *shadowed box*

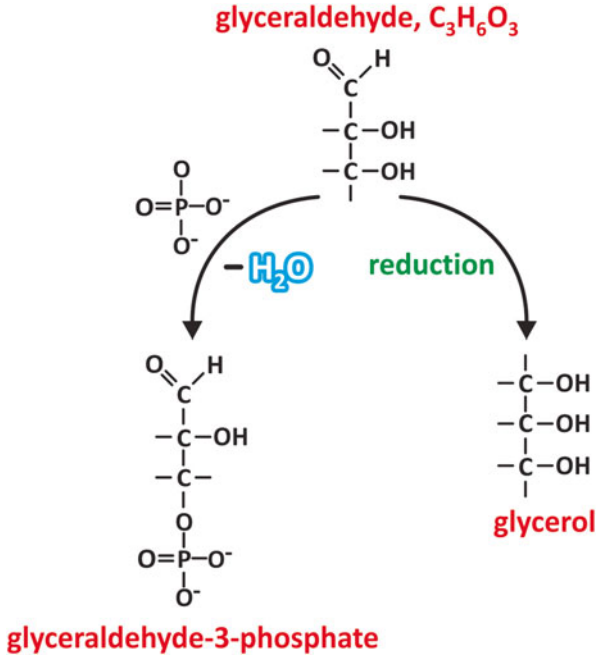


Fig. 3.16 Glyceraldehyde may be reduced to glycerol upon chemical reduction of the C=O group in carbon 1 (carbons are numbered starting with the one from the carbonyl group, similarly to fatty acids, in which carbons are numbered starting in the carboxyl group). Phosphoric acid (HPO_4^{2-}) may react with carbon 3, for instance, to form an ester, glyceraldehyde-3-phosphate. These are examples of very frequent reactions involving saccharides

The chemical structure of glyceraldehyde deserves close attention as its central carbon (C2) has four different substituents (i.e., it is bound to four different atoms or groups of atoms), being referred to as a chiral carbon or chiral center. Imagine the permutation of the H and OH substituents, for instance. A different molecule results from this switch (Fig. 3.17).

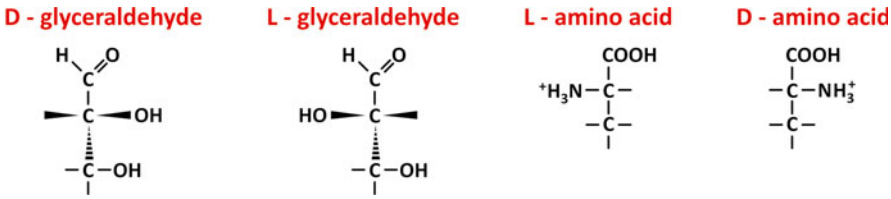


Fig. 3.17 L- and D-glyceraldehydes are mirror images. The same nomenclature was extrapolated to amino acids (alanine is shown as example)

Although having the same elemental composition and very similar structure, both molecules cannot overlap because the orientation in space of the H and OH substituents is different. The difference is clear if the molecule is represented in a 3D perspective. A closer look reveals that both molecules are mirror images of each other, i.e., they are enantiomers. To distinguish between both enantiomers of glyceraldehyde, one is named "L," and the other is named "D." These labels were arbitrarily assigned by Emil Fischer but are used to name saccharides and amino acids by extrapolation from glyceraldehyde (Fig. 3.17): D stands for dextro (right in Latin) and refers the structure having the OH group in the chiral carbon to the right when it is projected toward the observer; L stands for levo (left in Latin) and refers the structure having the OH group in the chiral carbon to the left when it is projected toward the observer.

Hermann Emil Fischer
(1852 –1919)

In chemistry research, there are two other nomenclature conventions for enantiomers independent from each other: the *R* and *S* system, which is based on a classification of the substituent group based on the atomic number of atoms bound to the central atom (chiral center), or the + and − system based on optical activity, i.e., on the direction of rotation of incident plane polarized light. Symbols + and − are sometimes replaced by *d*- (dextrorotatory) and *l*- (levorotatory) but *d*- and *l*- are easy to confuse with D- and L- and prone to misunderstanding. Both systems are more robust than Fischer's D- and L- because they are not dependent on the compari-

son with glyceraldehyde. Chemists tend to use *R/S* or +/− but biochemists are still "attached" to the D/L system for a simple reason: chiral diversity among biological saccharides and amino acids is very restricted. By far the most abundant saccharides in human biochemistry are D-. Interestingly L- is the preferred form in amino acids. Natural evolution favored one form specifically probably because it is simpler to have only one form as the building block for saccharide polymers (polysaccharides) and amino acid polymers (proteins). A small protein with 100 amino acid residues that could be D- or L- would have 2100 different possible isomeric structures. Because only L-amino acids are used, only one isomeric form is allowed. Why specifically L-amino acids and D-saccharides and not the other forms? It is not clear; probably it originated from originally random processes that later propagated and converged by evolution into the specific enantiomers found in nature nowadays.

When longer carbon chains are considered, more complex saccharides are possible depending on the:

1. Length of carbon chains
2. Position of the carbonyl group in the carbon chain
3. Number and location of chiral centers

Although many different saccharides can be found in human body, pentoses and hexoses are the most frequent in metabolic processes so we will now focus on these molecules, namely, ribose and glucose (Fig. 3.18).

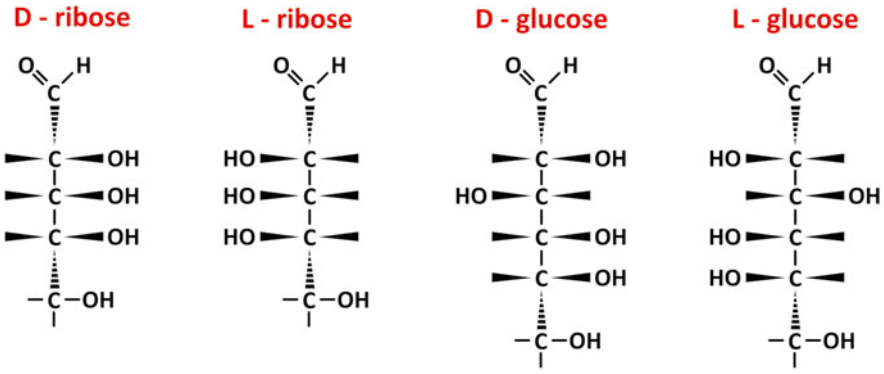


Fig. 3.18 D- and L-isomers of ribose (pentose) and glucose (hexose). D-forms are the more relevant forms in nature

Aldoses such as glyceraldehyde, ribose, and glucose react with water, forming a hydrate. This happens at the C=O group because the oxygen atom attracts the electrons leaving the C atom deficient in electrons, therefore prone to interact with the electrons of water oxygen (Fig. 3.19). The reaction is reversible, so

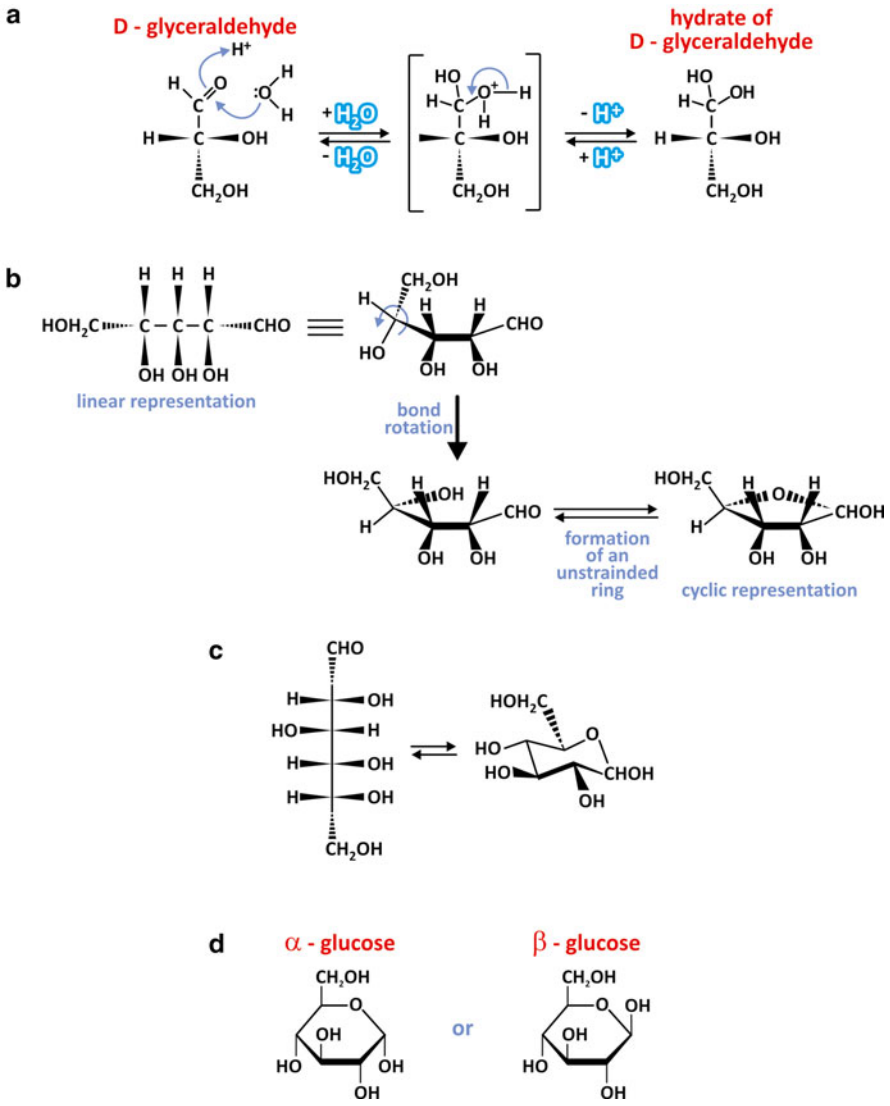


Fig. 3.19 Hydration of glyceraldehyde forms a hydrate (**a**). This reaction is reversible so glyceraldehyde coexists with its hydrate. The carbon originally present as a carbonyl group is the only carbon with two bonds to oxygen. A similar reaction may occur intramolecularly in pentoses (**b**) and hexoses (**c**). Panel (**b**) shows in detail the reaction of C=O (C1) with the OH group in C4, analogous to hydration. A cyclic pentose is thus formed. As with hydration, the reaction is reversible and both forms coexist, although the cyclic form is more abundant. The cyclic form of the hexose glucose is not planar as ribose is, as the molecular hexagon is flexible and adopts other conformations, such as the "chair" conformation (**c**). Cyclization results in the formation of two anomers (**d**) because the OH group formed at C1 may be placed on two different sides of the newly formed molecular ring: in the α-anomer the OH group in C1 is in the opposite side of the plane of the ring relative to the terminal carbon, C6 (–CH_2OH) and in the β-anomer they are both at the same side. Both the "chair" (**c**) and the more simplistic planar (**d**) representation of hexoses is used in this book

aqueous solution of aldoses contains mixtures of their aldehyde and hydrate forms. Nevertheless, pentoses and hexoses may react intramolecularly in a way that is similar to hydration. Because the carbon chain is able to bend and is dynamic (similarly to saturated aliphatic chains in lipids), the carbonyl group may contact alcohol groups (OH—hydroxyl) in the same molecule and react with it. The result is the formation of a cyclic molecule by the conversion of the carbonyl group in a hemiacetal group (Fig. 3.19). The cyclization is reversible, and in cells, the cyclic forms of ribose and glucose coexist with the linear forms, but the cyclic forms are dominant. Upon cyclization, two enantiomers are formed because the hydroxyl group in C1 may be linked to any of the two sides of the plane of the ring: in the α-anomer the OH group in C1 is in the opposite side of the plane of the ring relative to the terminal carbon, C6 ($–CH_2OH$), and in the β-anomer, they are both at the same side. Glucose adopts a "chair" conformation at variance with the strict planar ring of ribose (Fig. 3.19), but α and β anomers exist the same way. The existence of the enantiomers has drastic implications in the polymerization of hexoses.

3.2.1 From Monomers to Polymers: Polysaccharides

Saccharides such as hexoses and pentoses may react with each other forming chains that may reach considerable size. Molecules built from the association of smaller molecules of a kind are generally named polymers, and polymers made of unit saccharides such as ribose or glucose are named polysaccharides. The units forming polysaccharides are referred to as monosaccharides. A covalent association of two monosaccharides is a disaccharide. Association of "few" monomers forms "oligosaccharides"; the size boundary between oligosaccharides and polysaccharides is not well defined.

Two monosaccharides may associate by dehydration. Take the example of two glucose molecules forming a maltose molecule (a disaccharide) by dehydration (Fig. 3.20). C4 in one molecule and C1 in the other become covalently attached by an acetal linkage or *O*-glycosidic bond. Water resulting therefrom is formed with the oxygen previously attached to C1. The reverse process is the hydrolysis of maltose into two glucose monomers, which although thermodynamically favorable, is a very slow reaction. In practice, when degrading enzymes are not present, the process my be be so slow that it is irrelevant.

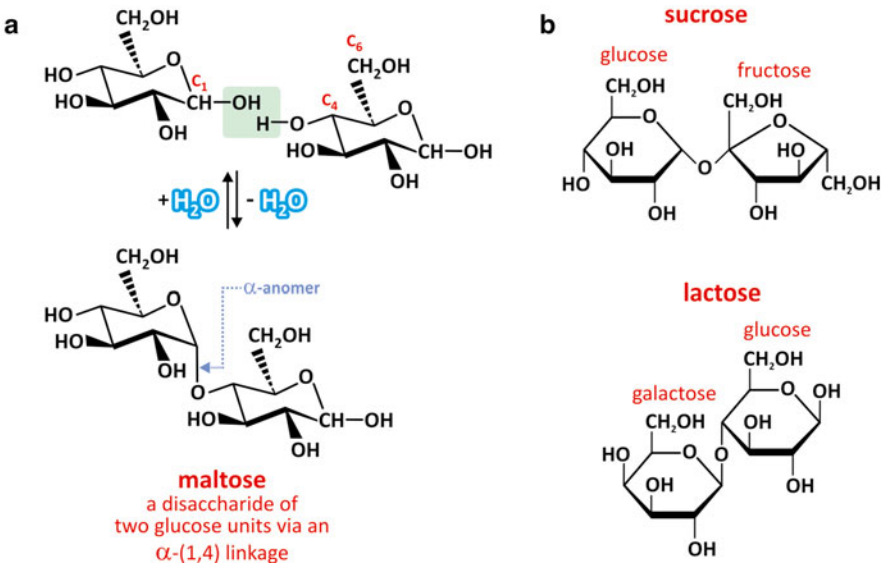


Fig. 3.20 Two glucose molecules may associate covalently by dehydration (**a**). When C1 in α-glucose (α anomer) reacts with C4 in another glucose, maltose is formed. Maltose is thus a disaccharide formed by linking two glucose molecules through an acetal or *O*-glycosidic bond. This bond is named "α-(1,4)" to stress that C1 in the α anomer binds to C4 in the other molecule. Sucrose and lactose (**b**) are other examples of disaccharides. Both are formed from the conjugation of different constituent monosaccharides: glucose and fructose in the case of sucrose and galactose and glucose in the case of lactose. Sucrose involves an α-(1,2) bond, whereas lactose involves a β-(1,4) linkage

The stereo chemistry (i.e., the spatial orientation of the chemical groups in the molecule) is very important because the covalent linkage of two molecules imposes restrictions on the way molecules can move in space. Depending on whether monomers are α or β anomers, different degrees of restriction arise. The flexibility of the conjugate is very much dependent on the enantiomers because the interaction between molecular groups in the disaccharide is very different (Fig. 3.21a). This effect is amplified in large polymers; polysaccharides may have a wide range of flexibilities, from extremely stiff and straight to coiled and deformable depending on the enantiomers used. Cellulose, for instance, is a glucose polymer formed with β(1,4) bonds that is extremely resistant mechanically, whereas amylose is an example of a flexible polymer formed by a α(1,4) backbone (Fig. 3.21b, c). Cellulose properties determined its evolutionary selection toward structural functions in plants, forming cell walls, which impacts in the macroscopic properties of wood. Amylose is a component of starch, a molecule that curls forming helices and is stored in plants for use in the energetic metabolism. Starch is the most common polysaccharide in human diet. Cellulose and amylose are striking examples of how apparently small details may actually determine profound differences in molecular properties and structures and therefore also in function.

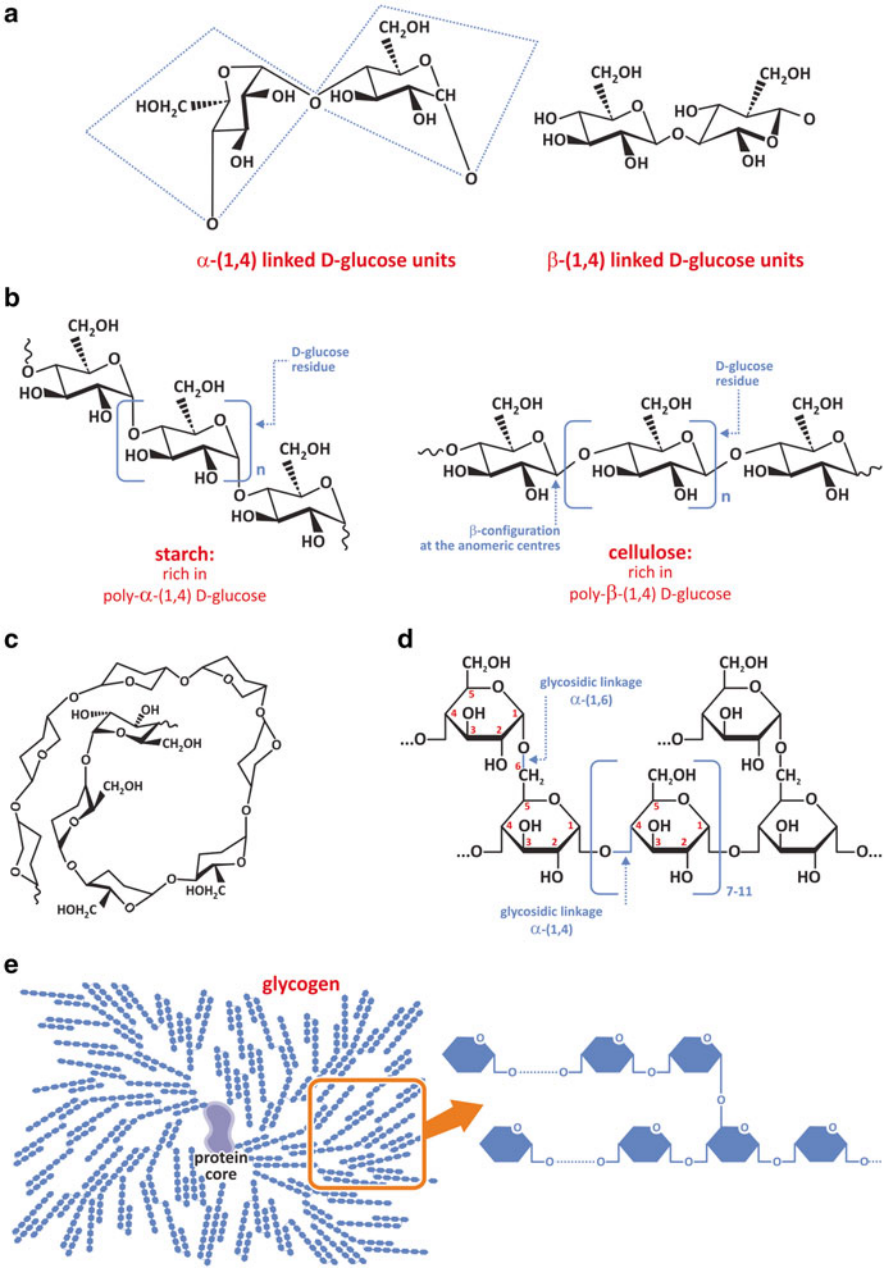


Fig. 3.21 Two monosaccharides such as D-glucose forming a disaccharide (**a**) have very different restrictions to articulate and to move depending on whether the glycosidic bond is α-(1,4) or β-(1,4). When several monomers bind to form a polymer (**b**), successive α-(1,4) or β-(1,4) bonds confer distinct properties to the polymer: α-(1,4) bonds enable bending between monomers, which results

Carbon 6 is also available for reaction so linear polymers formed of C1–C4 chains may branch when C1–C6 bonds are also formed (Fig. 3.21d). Amylose turns into amylopectin when β(1,6) links are formed. Glycogen is the human storage polysaccharide and is very similar to plant amylopectin (Fig. 3.21e). They differ only in the frequency of branching and average size of α(1,4) segments. The advantage of having glycogen as energy storage relative to a linear (unbranched) polysaccharide relates to the fact that glycogen is enzymatically degraded by saccharide hydrolysis of the terminal units. A branched molecule has several termini which can all be degraded at the same time, making glucose readily available at high rate.

Depending on their chemical composition and stereochemistry, polysaccharides found in nature have one of three functions: (1) structural/mechanical protection, (2) energy storage, and (3) water-binding (protection against dehydration). Several examples are listed in Table 3.3. Hyaluronic acid is particularly interesting for its involvement in connective tissue. Its interest extends from biochemistry to histology. Hyaluronic acid, a polysaccharide with sulfate groups (SO_4^{2-} has similar properties to PO_4^{2-}), forms an extracellular mesh with collagen, making up a flexible but resistant hydrated histological structure (Fig. 3.22).

Table 3.3 Examples of the function of polysaccharides found in nature: structural (Str), energy storage (Sto), water-binding hydration (Wat) (D-Glc, D-glucose; D-GlcNAc, *N*-acetyl-D-GLUCOSAMINE; D-GlcUA, D-glucuronic acid)

Polysaccharide	Monosaccharide 1	Monosaccharide 2	Bond	Branching	Location	Main function
Bacteria						
Peptidoglycan	D-GlcNAc	D-MurNAc	β-(1,4)	–	Bacterial wall	Str
Dextran	D-Glc	–	α-(1,6)	α-(1,3)	Capsule	Wat[a]
Animals						
Chitin	D-GlcNAc	–	β-(1,4)	–	Insects, crabs	Str
Glycogen	D-Glc	–	α-(1,4)	α-(1,6)	Liver, muscles	Sto
Hyaluronic acid	D-GlcUA	D-GlcUA	β-(1,4) β-(1,3)	–	Connective tissue	Str, Wat

[a]Capsular materials like dextrans may be overproduced when bacteria are fed with saccharides to become reserves for subsequent metabolism

Fig. 3.21 (continued) in curled polymers such as amylose (**c**) and the stiffer β-(1,4) links between monomer favor linear straight polymers, such as cellulose. Therefore, cellulose is found in structural elements of plants, while amylose is used by plants as energy storage. Humans also use a poly-α-(1,4) saccharide as energy storage. Periodic α-(1,6) branching every 10 residues, approximately, further enables a globular organization of this polysaccharide (**d**). The final result is a regularly branched polymer of D-glucose named glycogen (**e**). Glycogen synthesis is initiated by a protein and elongation requires several enzymes (see Sect. 8.2)

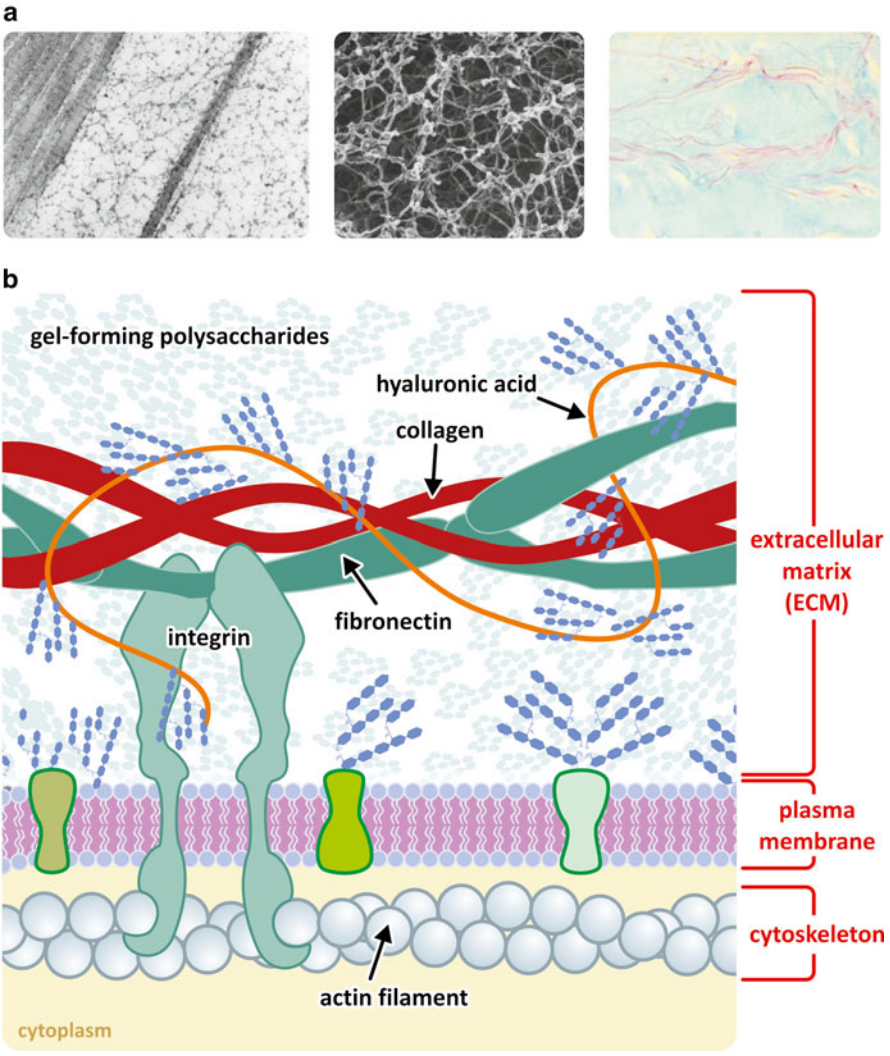


Fig. 3.22 The extracellular matrix has hyaluronic acid in its composition. (**a**) Histological preparations of cock crest highlighting the hyaluronic acid matrix (*left*, conventional electron microscopy; *center*, platinum–carbon replica; *right*, preserved blue-dyed hyaluronic and extracellular heavily glycosylated proteins, the "proteoglycan matrix"). Figure reprinted with the permission of Instituto de Histologia e Biologia do Desenvolvimento, Faculdade de Medicina, Universidade de Lisboa, Portugal. (**b**) Schematic representation of the molecular organization of extracellular matrix, which is composed of gel-forming saccharides attached to a backbone of hyaluronic acid that is intertwined among collagen fibrils. Polysaccharides form gels due to the high density of H bonds. These gels confer a moderately rigid structure and mechanical protection to cells and retain water, which prevents desiccation of the tissues

3.2.2 *Molecular Conjugates of Monosaccharides*

We have seen in previous sections that saccharide monomers offer diverse possibilities of reaction and so they are molecules that form many conjugates in nature. The most important derivatives are phosphate esters. Phosphoric acid is able to form up to three ester bonds (Fig. 3.23), although the tri-esters are not commonly found in nature. Yet, diesters are important and enable saccharide phosphates to form polymers (e.g., nucleic acids) or bridge saccharides with other organic molecules.

Phosphate groups forming esters are anionic in aqueous environment in the most common biological pH ranges. This means that neutral molecules, such as glucose, become charged when esterified with a phosphate. The consequence is an increase in solubility in water and a decrease in the ability to cross lipid bilayers, for instance. This is deemed important as glucose metabolism starts by forming glucose phosphate (Fig. 3.23a).

Phosphate ester hydrolysis is a spontaneous but very slow process, which makes it under enzymatic control in cells. In addition, many chemical processes occurring in cells, such as condensation of polymers with formation of water, are unfavorable processes (there is "excess" water in most cell environments); enzymes speed the reaction but do not shift the equilibrium toward condensation. The use of phosphate derivatives of the monomers in the process of condensation facilitates the reaction as phosphates are so-called good leaving groups: they alter the reactivity of transient chemical species in the course of the mechanism of reaction.

ATP (adenosine triphosphate, Fig. 3.23b) is among the biological molecules that are saccharide derivatives and involves a phosphate ester. A phosphate diester bond bridging two other phosphates is another interesting characteristic of this molecule. The energy involved in the phosphate–phosphate bonds makes this molecule pivotal in energetic metabolism. Divalent cations such as Mg^{2+} are usually associated to ATP and other molecules having diphosphate groups. This reduces electrostatic repulsion between the oxygen atom of water and the negative charge of phosphate groups, facilitating the hydrolysis of phosphate derivatives.

Probably not so famous as ATP, but equally important in biochemistry, is coenzyme A (Fig. 3.23c). This is a relatively small but complex molecule. Amazingly, coenzyme A has phosphate and saccharide groups but owes its reactivity to a terminal thiol (–SH) group. This thiol group may bind an acetyl residue through a thioester bond, but may also bind a fatty acid, forming acyl-CoA, which is involved in lipid metabolism.

Nicotinamide adenine dinucleotide (NAD^+) is another interesting case of saccharide derivative that also contains phosphates. NAD^+ intervenes in redox reactions as it may accept and donate electrons, changing from NAD^+ to $NADH + H^+$ or vice versa. One extra phosphate group turns NAD^+ into $NADP^+$, which has similar redox properties but can only bind to specific enzymes that usually do not bind NAD^+. This implies that there are specific metabolic roles for $NADP^+$, distinct from NAD^+. The phosphates are involved in enzyme recognition but not in the redox activity itself (Fig. 3.23d). The same happens with flavin adenine dinucleotide (FAD and $FADH_2$; Fig. 3.23e).

Nucleotides themselves deserve closer attention because they polymerize to form the so-called nucleic acids. They will be left for further discussion in the next section. To finalize, it should be stressed that many therapeutic drugs are also sac-

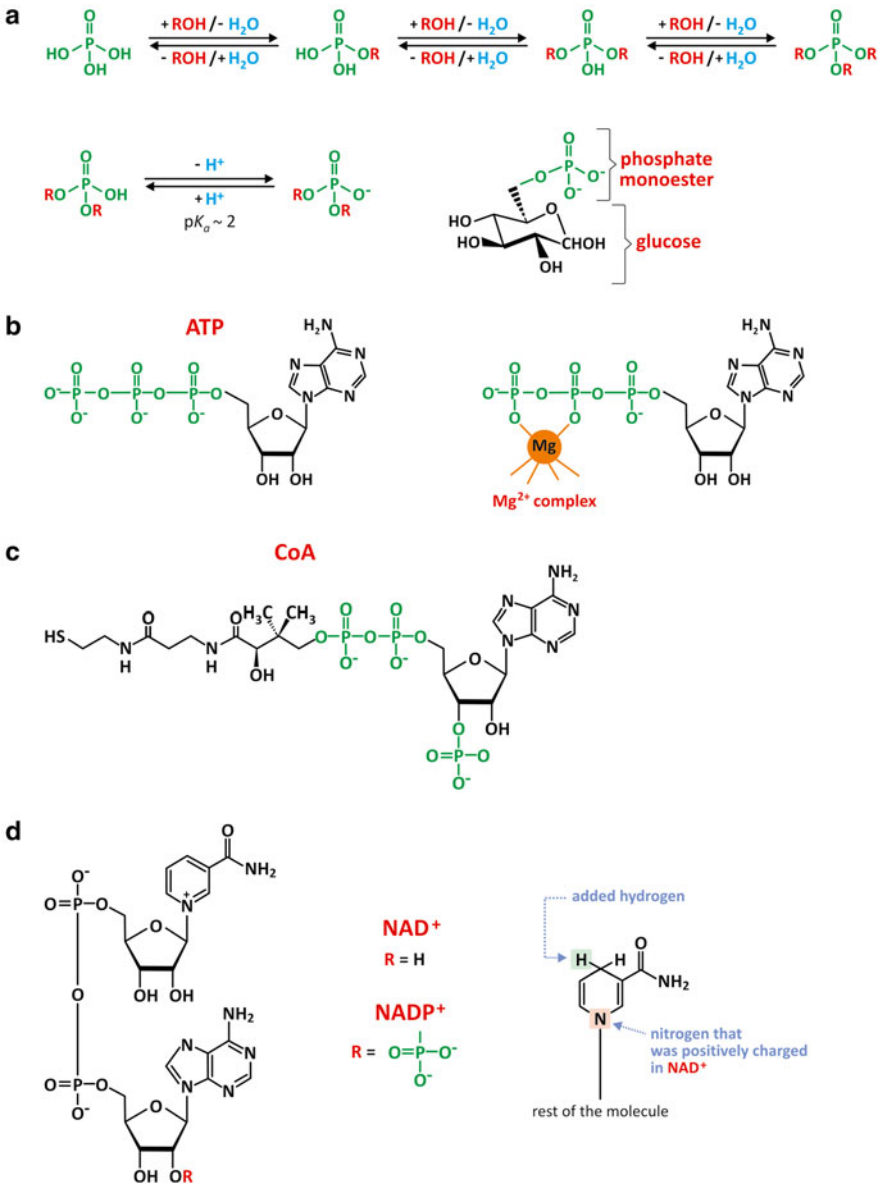


Fig. 3.23 Phosphates are able to form esters or diesters bridging two organic molecules (**a**). Phosphate confers an anionic charge to the newly formed chemical entity because the ionization of the phosphate group occurs at pH > 2, increasing its solubility in aqueous medium and decreasing its ability to translocate across lipid membranes. This is the case for glucose-6-phosphate, which is "trapped" in the cytosol of cells, where it will be processed in different metabolic pathways. Adenosine triphosphate, ATP (**b**), and coenzyme A, CoA (**c**), are important biological molecules with a saccharide residue bound to a phosphate group. ATP also contains a phosphodiester bond, very important for its reactivity in cells. CoA has a couple of phosphate groups bound to

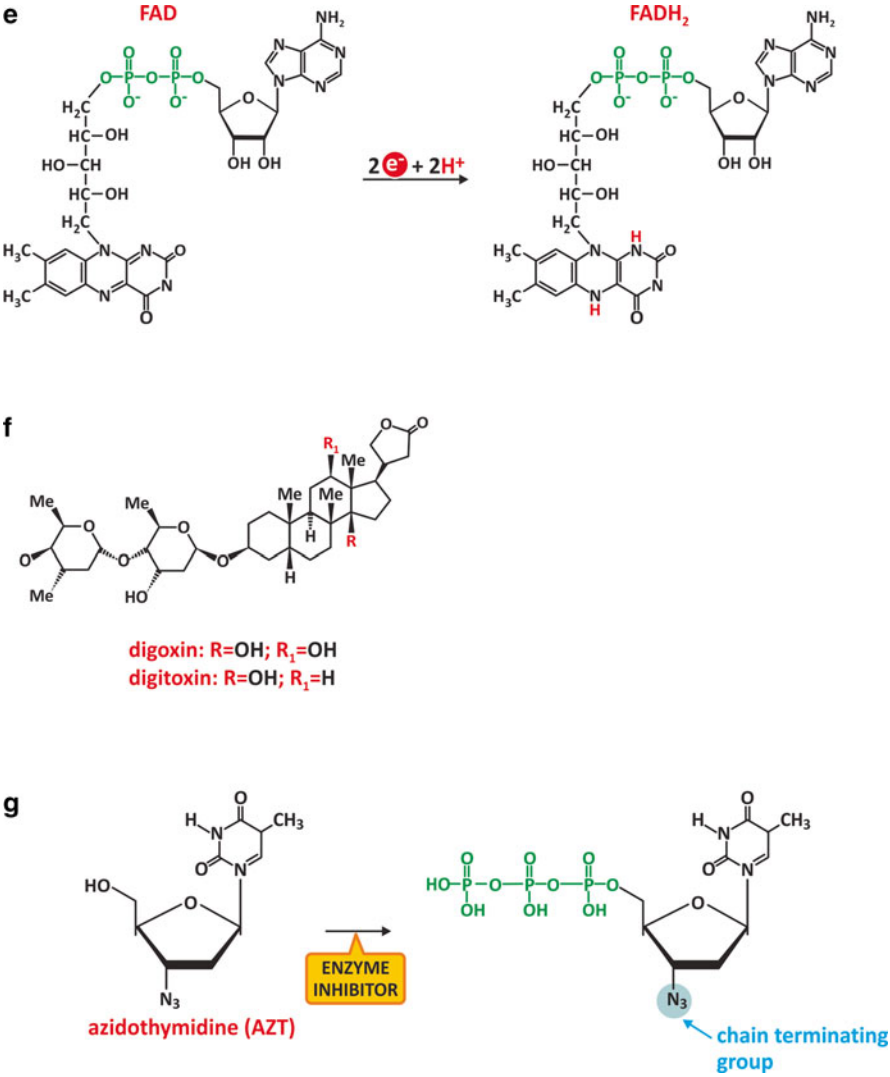


Fig. 3.23 (continued) each other, but its reactivity in cells is dictated by the sulfhydryl group (-CSH). Nicotinamide adenine dinucleotide (NAD^+) is another saccharide derivative with saccharide residues bound to phosphates. NAD^+ may be reduced to NADH (**d**). Redox reactions of NAD^+/NADH take place in a specific cyclic residue of the molecule, involving a nitrogen atom (*right*). NAD^+/NADH phosphate, $NADP^+$/NADPH (**d**), also participates in redox reactions in human metabolism. NADH and NADPH cannot be distinguished by their reducing properties because the phosphate group that distinguishes them does not interfere with the nitrogen atom that affords the redox properties. Yet enzymes use specifically NADH or NADPH and so there is no redundancy between these molecules. FAD and $FADH_2$ (**e**) are molecules similar to NAD^+ and NADH in that both constitute adenine nucleotides and their role in metabolic redox reactions. Digoxin and digitoxin (**f**) are examples of drugs with monosaccharides in their structure; more specifically, three residues are specifically combined as part of a unique structure. Another example of drug that is a saccharide derivative is azidothymidine (AZT), which is converted to a triphosphate in cells (**g**) and is able to insert in the active center of the reverse transcriptase of HIV because it is structurally similar to the natural substrate. However, the natural substrate does not have the N_3 group. The presence of this group blocks the conversion of the viral RNA into DNA

charide derivatives, such as digoxin (Fig. 3.23f), used in the treatment of heart conditions. Azidothymidine (AZT) is another example. It is an analogue of thymidine that may inhibit the action of reverse transcriptase of HIV. It was the first drug used in the treatment of AIDS. Cellular enzymes convert AZT into the effective 5-triphosphate form (Fig. 3.23g). Once bound to reverse transcriptase, the azide group, N3, is responsible for chemical inhibition. Inspired by the success of AZT (Fig. 3.24), many nucleosides are now under development to create new inhibitors of HIV reverse transcriptase to fight AIDS.

3.2.3 Molecular Conjugates of Oligosaccharides

It is worth stressing that the combination of saccharide monomers may generate a big diversity of products when compared to amino acids (Fig. 3.25). Two glucoses, for instance, can bind via 6 carbons in each monomer, thus being able to form 36

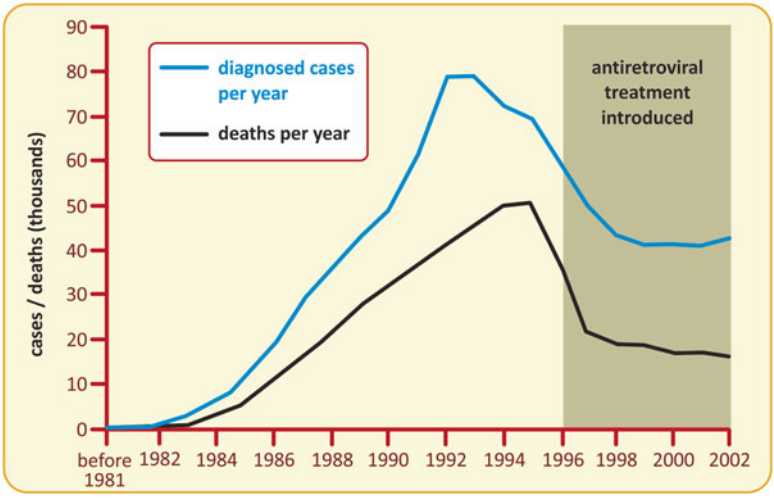


Fig. 3.24 Preventive campaigns highlighting the need to change risk behaviors had a strong impact in the spreading of AIDS in the USA, with a marked decrease in the number of diagnosed cases and deaths per year after 1993-95. The use of AZT and other drugs had a very positive additional effect in the reduction of AIDS-caused mortality

Fig. 3.25 (continued) Saccharide tags are covalently bound to lipids, usually rigid lipids such as ceramide for a better anchoring to the membrane (**b**). Glycolipids (i.e., associations of saccharides and lipids) determine blood groups, for instance (see Box 3.5). The same principle applies to oligosaccharides attached to proteins, i.e., glycoproteins (**c**). The side chain of the amino acid asparagine may react with a saccharide by dehydration forming an *N*-glycosidic bond (analogous to an *O*-glycosidic bond, but involving *N* instead of *O*). Likewise, the side chain of the amino acid serine may react with a saccharide forming an *O*-glycosidic bond. The oligomeric sequences of saccharides attached to protein IgG is shown in panel (**c**) *bottom* as an example

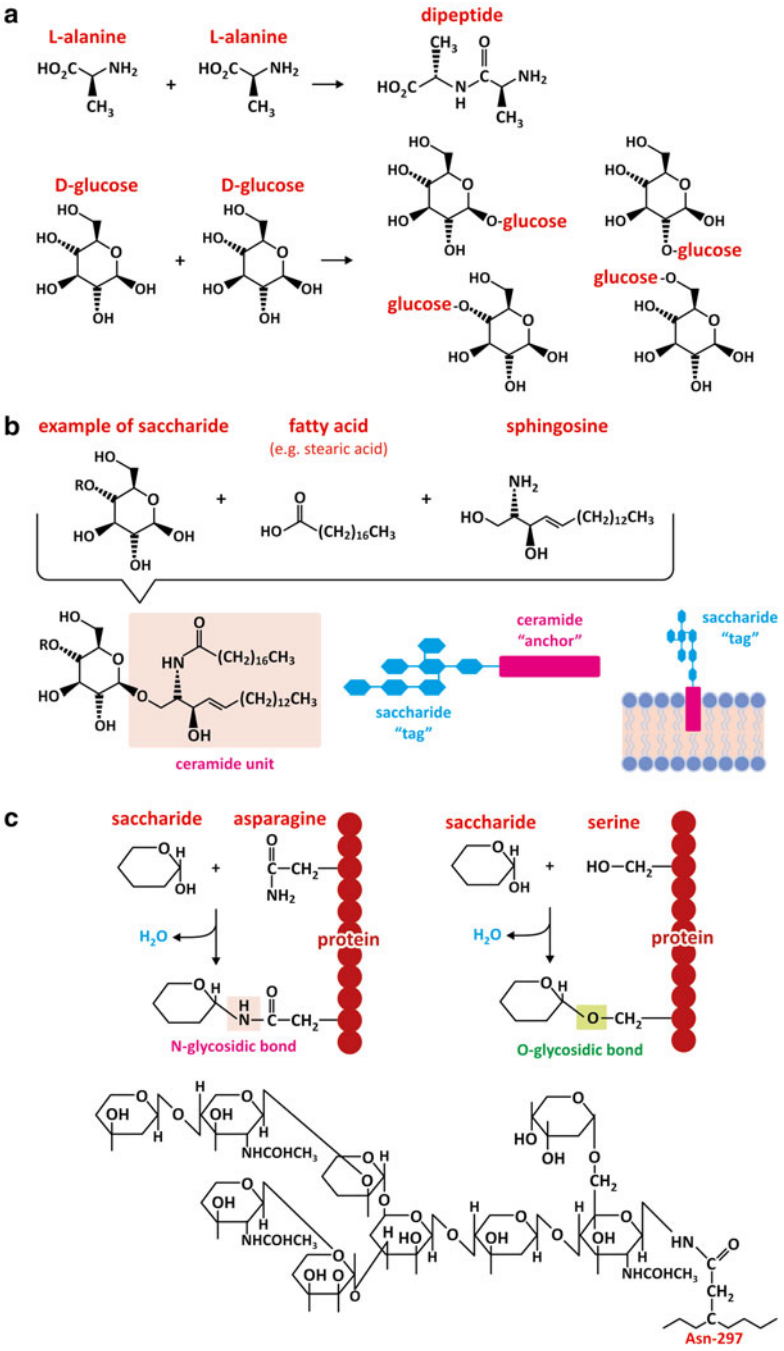
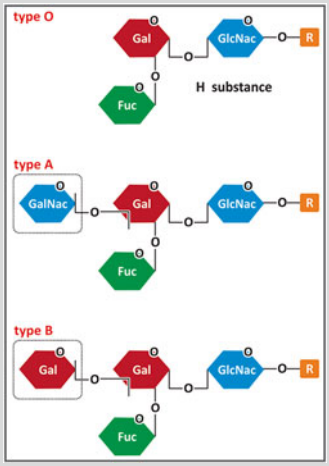


Fig. 3.25 A combination of two amino acids generates one single dimer, but there are several ways that two monosaccharides can combine to form a disaccharide (**a**). Saccharides are better suited to form highly specific structures at the surface of cell membranes (**b**) or proteins (**c**).

different molecules. Considering the anomers, the diversity increases. It is not surprising that oligosaccharides are present in the surface of cells as receptors of unique structure (see an example in Box 3.5) while amino acids form polymers (proteins) having domains with few restricted well-defined structures. Moreover, monosaccharides or oligosaccharides are frequently found in nature attached to proteins.

Box 3.5: The ABO Blood Groups

There are different blood groups according to different immunogenic molecules present in erythrocytes. The most important classification of blood groups is based on three antigens, A, B, and O that form 4 groups: A, B, O, and AB—the ABO blood groups. The ABO blood group antigens are oligosaccharide chains attached to proteins and lipids located in the outer surface of erythrocytes. One single residue of a small oligosaccharide determines whether the antigen is A, B, or O (see figure).



Fuc represents the monosaccharide fucose, Gal, galactose, GalNAc, N-acetylgalactosamine, and GlcNAc represents N-acetylglucosamine

The immune system of an individual produces antibodies against the ABO antigens not present in his own erythrocytes. Individuals in A group will have antibodies against B and vice versa. Type O, the most common, does not contain the last residue, which is the antigen, in its structure (in fact the original nomenclature was 0—zero—but became the letter O). So, individuals in blood group O will have both anti-A and anti-B antibodies. Individuals in blood group AB are rare and, naturally, have no anti-A and no anti-B antibodies. This has tremendous implications in blood transfusions as a patient cannot receive erythrocytes against which he/she has antibodies. AB individuals

(continued)

Box 3.5 (continued)

can, in principle, receive blood from any donor; O individuals can donate blood to any individual; A and B individuals can only donate and receive blood to/from individuals belonging to the same blood group.

It is believed that ABO antibody production is stimulated when the immune system contacts in foods or in microorganisms with the saccharide antigens that are absent in the erythrocytes. The functions of the ABO blood group antigens are not known. Individuals who lack the A and B antigens are healthy, suggesting that any function the antigens have is not important, at least not in modern times.

Hemolytic disease of the newborn (HDN) is a serious medical problem that occurs almost exclusively in infants of blood group A or B who are born to group O mothers. This is because the anti-A and anti-B formed in group O individuals tend to be of the IgG type, which can cross the placenta. HDN tends to be relatively mild mainly because fetal erythrocytes do not express adult levels of A and B antigens. However, the precise severity of HDN cannot be predicted.

3.2.4 Polymers of Saccharide Conjugates: Nucleic Acids

Nucleotides that compose deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are formed by 2-deoxyribose or ribose, respectively, linked to a heterocyclic base, a purine (adenine, guanine) or a pyrimidine (cytosine and uracil or thymine), and a phosphate group attached to carbon 5 of the ribose residue. To avoid ambiguity with numbering of carbons of the heterocyclic base, the carbon numbers of the ribose and deoxyribose are identified with a prime: phosphate ester linkage occurs at C5′ (Fig. 3.26). The physical and chemical characteristics of the heterocyclic bases are extremely important as they are determinant for the way nucleotide polymers (nucleic acids) organize. The bases are planar, cyclic, aromatic molecules with N and O atoms able to participate in hydrogen bonding in the plane of the ring. The bases are low polarity groups poorly solvated, so both faces of the plane of the base rings will be fairly hydrophobic and thus subject to significant entropic effects.

There are four different possible nucleotides in RNA and DNA. RNA is formed by adenosine-5′-monophosphate (AMP), guanosine-5′-monophosphate (GMP), cytidine-5′-monophosphate (CMP), and uridine-5′-monophosphate (UMP). DNA is formed by deoxyribose, which is denoted by the prefix d in dAMP, in dGMP, in dCMP, and in dTMP. dTMP stands for thymine-5′-monophosphate using deoxyribose; DNA does not contain dUMP.

Because nucleotides are phosphate monoesters, they can form additional ester links to other alcohols, such as the OH groups in other nucleotides. In other words, they can polymerize by dehydration reactions. Nucleic acids are formed by phosphodiester bonds between C5′of one nucleotide and C3′ of another nucleotide (Fig. 3.26c). The result is a linear polymer having the heterocyclic bases and the

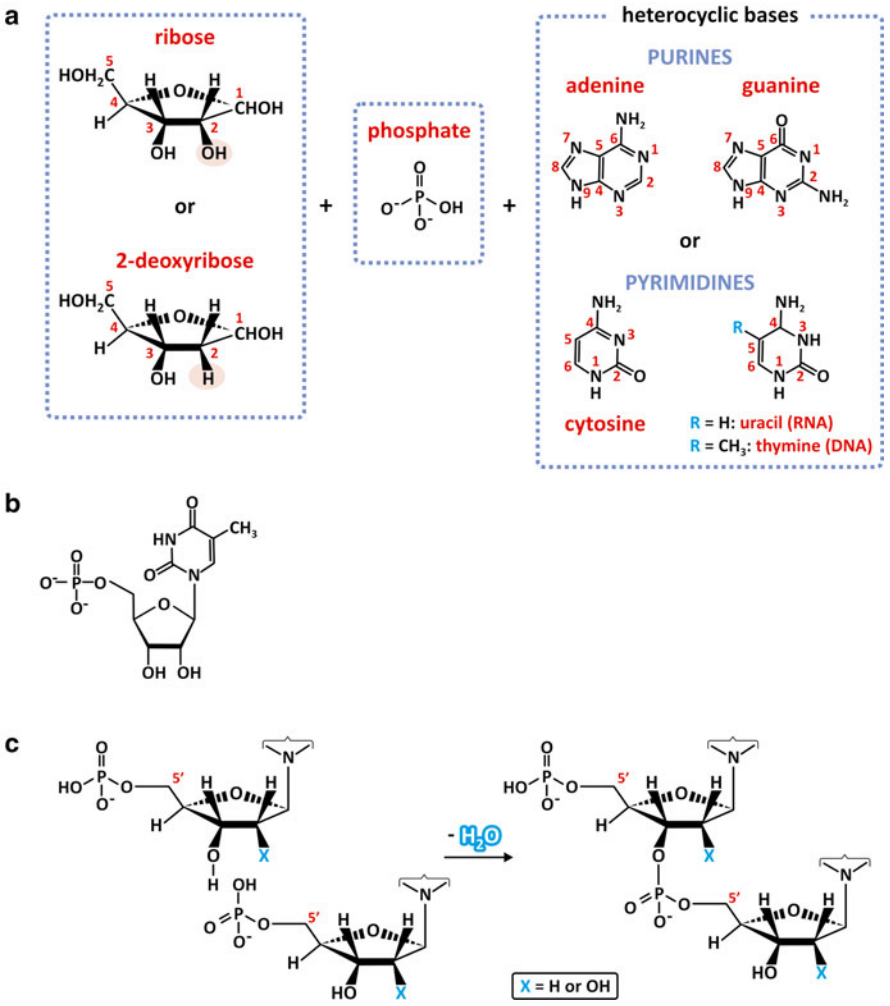


Fig. 3.26 Nucleotides are formed with ribose or 2-deoxyribose, phosphate and a heterocyclic base, purine (adenine or guanine) or pyrimidine (cytosine, uracil, or thymine) (**a**). The phosphate group forms a phosphodiester bond in C5 and the heterocyclic base binds to C1. The nucleotide deoxythymidine phosphate is shown as an example (**b**). A dimer of nucleotides may be formed by dehydration, which creates a phosphodiester linkage between the monomers via C5′ and C3′ (**c**). In RNA X=OH, in DNA X=H

phosphate groups in opposing sides, the phosphate groups being anionic (Fig. 3.27). Some simplified representations of nucleic acids pinpoint this characteristic (e.g., Fig. 3.27b), which remains elusive when the nucleic acid is simply represented by a sequence of letters identifying the nucleotides (T, thymine; C, cytosine; G, guanine; A, adenine; U, uracil). By convention, nucleic acid sequence is written from the C5′ to the C3′ endings, 5′ → 3′ (Fig. 3.27).

Unlike polysaccharides, nucleic acids are amphiphilic molecules (Fig. 3.27) so the entropic effect will be a significant driving force for folding in aqueous environ-

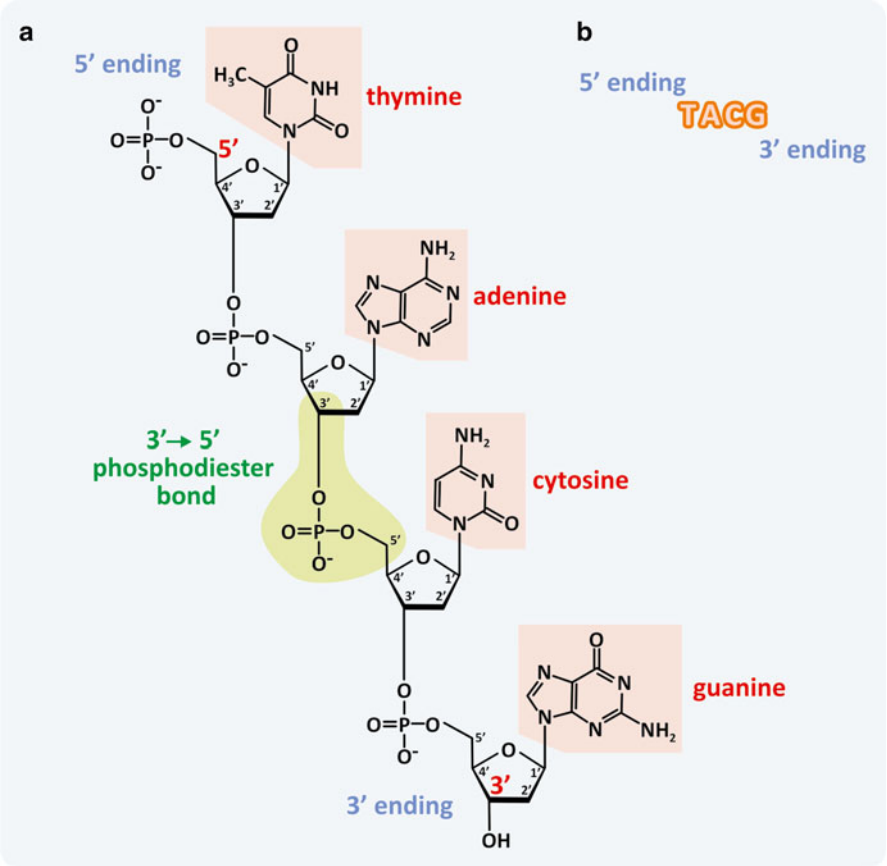


Fig. 3.27 Natural polymers of nucleotides are named nucleic acids. Desoxyribonucleic acid (DNA) has 2-deoxyribose residues and uses guanine, cytosine, adenine and thymine but not uracil (**a**). Ribonucleic acid (RNA) has ribose residues and uses guanine, cytosine, adenine and uracil but not thymine. Both polymers are formed by C3′–C5′ phosphodiester bonds. For the sake of simplicity, the chemical structure of the monomers is usually omitted, and other forms of presenting the nucleotide residues sequence are preferred. The simplest and most common form represents the nucleotides by a one-letter code (the first letter of the base name: G, A, C, T or U). Which end is the free, C5′ or C3′, is not explicitly mentioned but it is established by convention that the sequences are presented in the directions 5′ to 3′ (**b**)

ment. The heterocyclic bases will tend to nucleate to minimize their contact with water molecules. The crystal structure of transfer RNA (tRNA) shows that the bases stack parallel to each other, which is favored by their strictly planar structure. In addition, the nucleic acid tends to twist along its major axis forming a helix that exposes the phosphates to the aqueous medium and has the bases stacking in its core. In addition, most of the helical regions in tRNA consist of two sequences of the same RNA chain running in opposite directions with bases in opposite sequences contacting each other close enough and with the adequate stereochemical arrangement to establish hydrogen bonding between them. This adequate arrangement only occurs if purines pair with pyrimidines, as in pairs A–U and G–C (Fig. 3.28). This

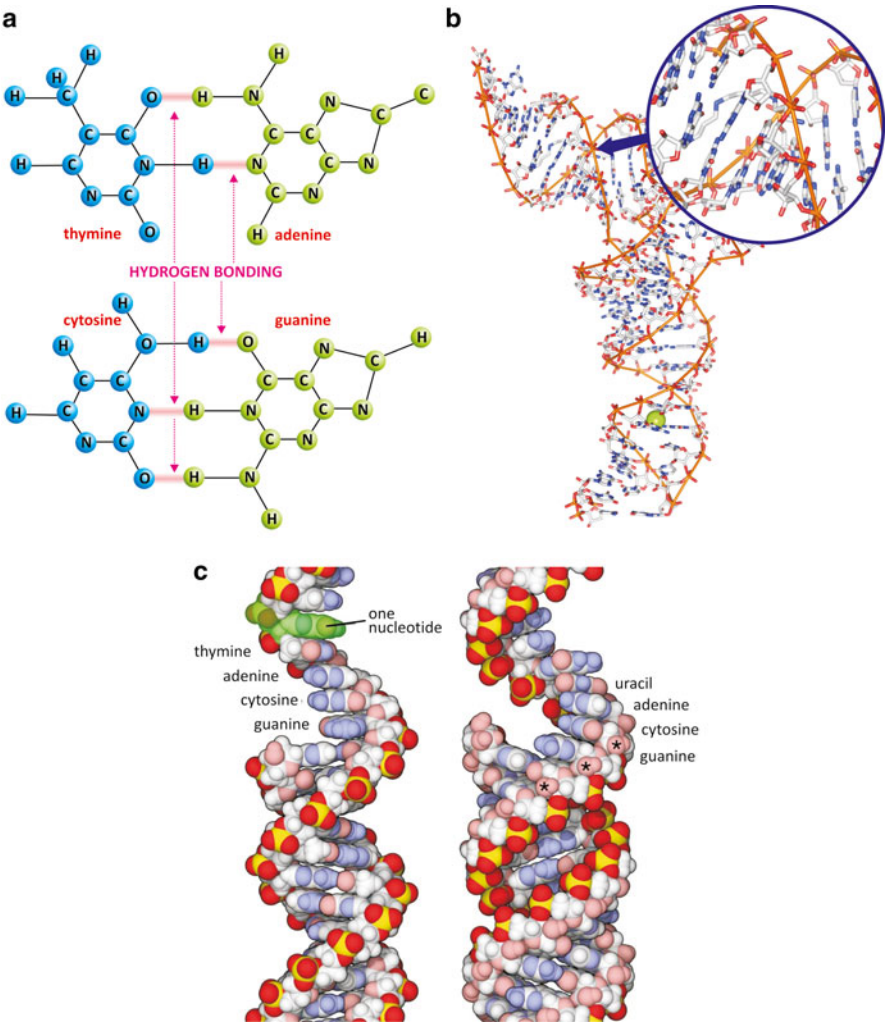


Fig. 3.28 Heterocyclic bases are flat. Hydrogen bonding occurs in the plane of the rings. Purines and pyrimidines are able to interact because of the match in the number and orientation of H doning/acceptor groups, forming pairs T-A and C-G (**a**), which is known as Watson–Crick base pairing. Because bases are so flat, relatively hydrophobic on both sides, and undergo base pairing, nucleic acids may bind complementary sequences of nucleotides in the same polymer or from a different polymer. The entropic effect will cause this arrangement to twist around its long axis forming a double helix in which the polar parts of the molecule, phosphate and pentose residues, are exposed to the aqueous medium shielding the relatively hydrophobic bases. In the center of this helix, the bases stack parallel to each other and are slightly rotated relatively to each other. This kind of organization can be found even in some domains of the transfer RNAs (**b**; PDB 2TRA). The OH groups present in C2′ groups of RNA (*asterisk* in panel (**c**) *right*) but not DNA (panel (**c**) *left*) have structural implications in the conformation of nucleic acids as these groups contribute to shield the core of the double strands from the aqueous environment. Panel (**c**) was reproduced from Goodsell, The Machinery of Life, 2009

is known as Watson–Crick base pairing. Hydrogen bonding occurs at the edges of heterocyclic bases, in the plane of the rings.

In regions in which the two opposing antiparallel sequences of tRNA have a considerable array of complementary base pairs, both RNA sequences fold into a helical structure to keep the parallel stacked base pairs in the core surrounded by the pentose and phosphate esters backbone. This is the double helix structure frequently associated to DNA but equally present in some regions of the RNA molecule. Ribosomal RNA structure is similar to that of tRNA (Fig. 3.28). It is also worth stressing that DNA polymers may have complementary RNA polymers, associate with them and even fold into helices. However, the alcohol group, OH, at C2′ makes the structure of RNA less compact due to its volume and polarity.

The structure of RNA is not only less compact, but it is also less chemically stable. The OH group at C2′ is close to the phosphate diester bond with which it can react to hydrolyze RNA (Fig. 3.29). Because RNA molecules have transient functionalities and are not stored for very long periods in the cells, this limitation of RNA is not a problem. DNA is less prone to hydrolysis because it lacks the OH group in C2′, being the molecule that natural evolution selected to store genetic information for longer periods.

Some drugs target DNA taking advantage from the parallel stacking of heterocyclic bases. Notably, most of these molecules are composed of hydrophobic planar heterocyclic groups able to intercalate the base pairs of DNA (Fig. 3.30). Some of these molecules are used in the treatment of cancer because they prevent cell multiplication.

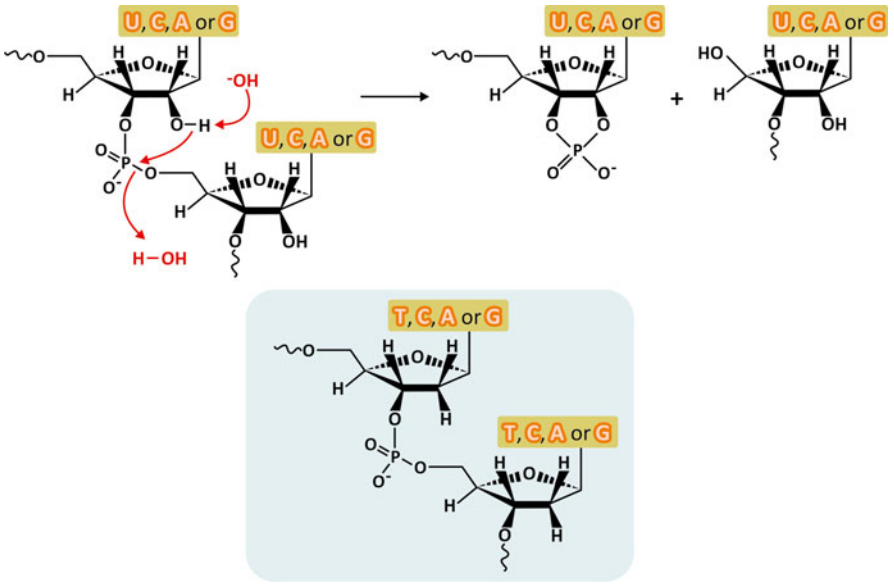


Fig. 3.29 Intramolecular hydrolysis of the phosphodiester bonds of RNA caused by a base (*top*). The absence of a hydroxyl group at C2′ increases the hydrolytic stability of DNA relative to RNA (*bottom shaded structures*)

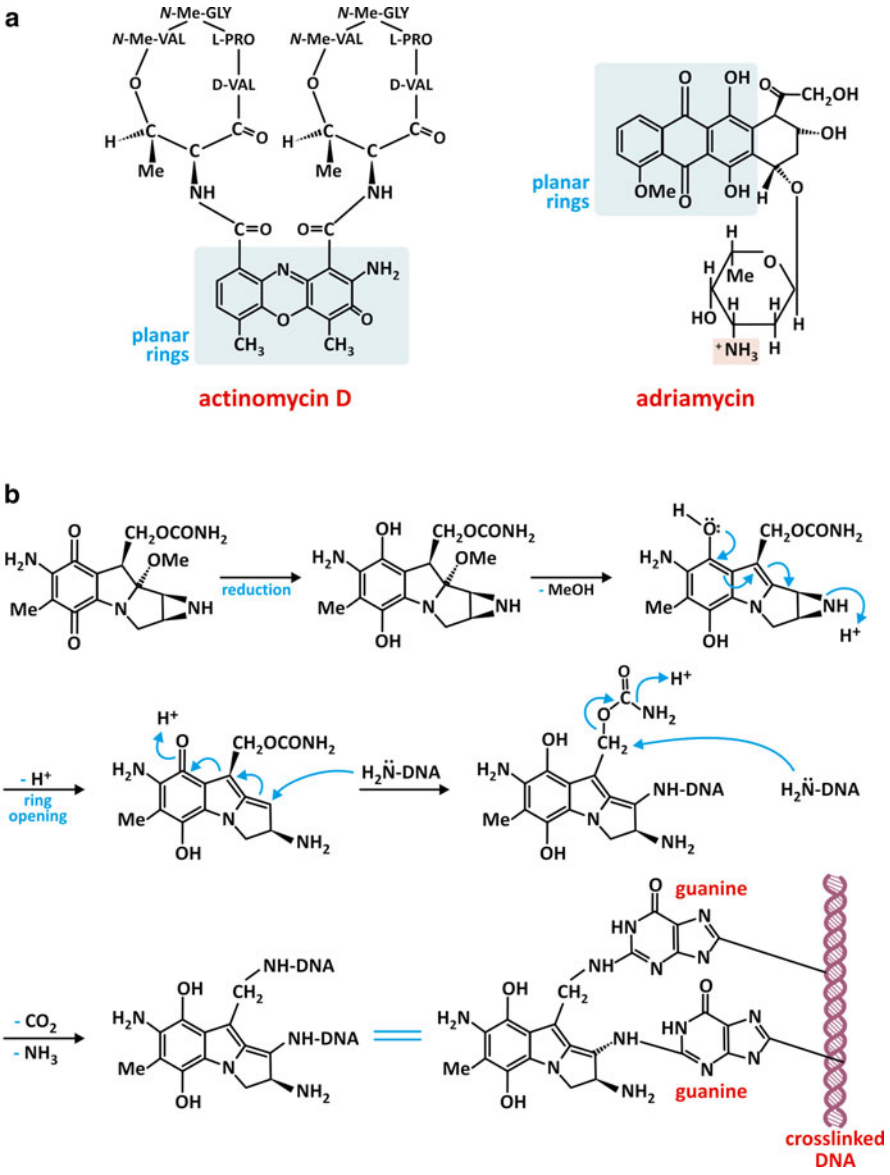


Fig. 3.30 Actinomycin D is an antibiotic with anticancer activity (**a**). It binds DNA because it has a flat polycyclic and relatively hydrophobic group able to intercalate the stacked bases of DNA, preventing RNA synthesis. Adriamycin (**a**) is also an anticancer drug that operates with the same mechanism of action: intercalation of a polycyclic flat hydrophobic group between the base pairs of DNA, preventing cell proliferation. Mitomycin C (**b**) has a different mechanism of action: it is a DNA cross-linker by covalently linking two guanines. The direct contact with these nucleotide residues is possible because this drug is a polycyclic flat and relatively hydrophobic molecule

3.3 Amino Acids and Their Polymers: Peptides and Proteins

Chemically speaking amino acids are molecules that simultaneously have carboxyl (–COOH) and amine ($–NH_2$) groups. Biochemists focus on α-amino acids, in which these groups are bound to the same terminal carbon (the so-called α-carbon in older organic chemistry nomenclatures), because naturally occurring proteins are polymers of α-amino acids. These amino acids have the structure depicted in Fig. 3.31. Besides the amino and carboxylic acid groups, the α-carbon (also named central carbon) attaches a hydrogen atom and another group, the so-called side chain that is represented by R. In nature R is one of 20 more common groups with few exceptions, which are usually derivatives of these 20 groups.

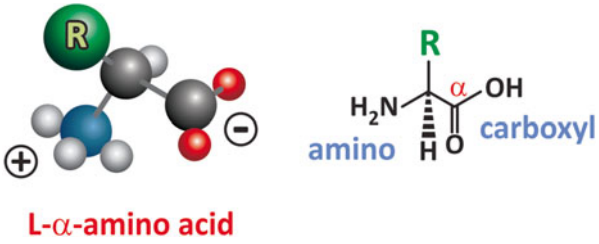


Fig. 3.31 The structure of α-amino acids. Depending on pH, in aqueous solution the amino group may be protonated and the carboxylic acid deprotonated, which makes amino acids potential zwitterions, i.e., bearing two opposite charges. Being weak bases and acids, amino acids have the ability to constitute buffers themselves (see Sect. 2.1.1)

Another interesting peculiarity of naturally occurring amino acids besides being α-amino acids is that they are almost exclusively L-enantiomers as the α-carbons are chiral centers. The other enantiomer is named D. The L and D nomenclature for the stereochemistry of the amino acids was established by Emil Fisher in analogy with glyceraldehyde, which also has a single chiral center with two possible enantiomers. Another nomenclature, more complex and following modern rules, exists to describe the stereochemistry of amino acids, but the predominance of L-amino acids and the simplicity of the L vs. D system resulted in the long-term longevity and universality of this system.

One simple empirical rule to distinguish L- from D-enantiomers is to adopt the perspective of the chemical structure of the amino acid along the H-αC axis (Fig. 3.32). The groups COOH, NH_2, and H appear projected as the vertices of a triangle. You can now recognize "corn" written clockwise in L-enantiomers or counterclockwise in D-enantiomers. This is the CORN rule of thumb.

The chemical nature of the lateral chain, R, is determinant for biochemical processes in which amino acids participate and for the structure that proteins adopt when such amino acids are present. Broadly speaking, amino acids can be grouped in four different categories based on polarity and acidic/basic nature of R (Fig. 3.33): acidic, basic, neutral polar, and neutral nonpolar. Other classification systems are based on the chemical nature of R: hydrocarbons, carboxylic acids, amides ($–CONH_2$), acyclic nitrogen containing, hydroxyl, sulfur containing, and nitrogen heterocycles. Figure 3.33

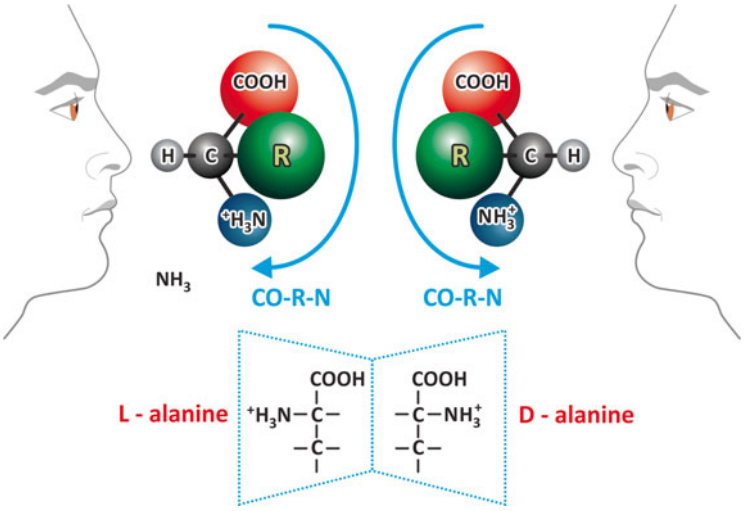


Fig. 3.32 The L vs. D nomenclature revealed by the CORN rule of thumb. When R=H (this happens in glycine, the simplest amino acid), chirality does not exist as two equal substituents (H, in this case) are attached to the central carbon. The example of L- and D-alanine is presented highlighting that they are mirror images

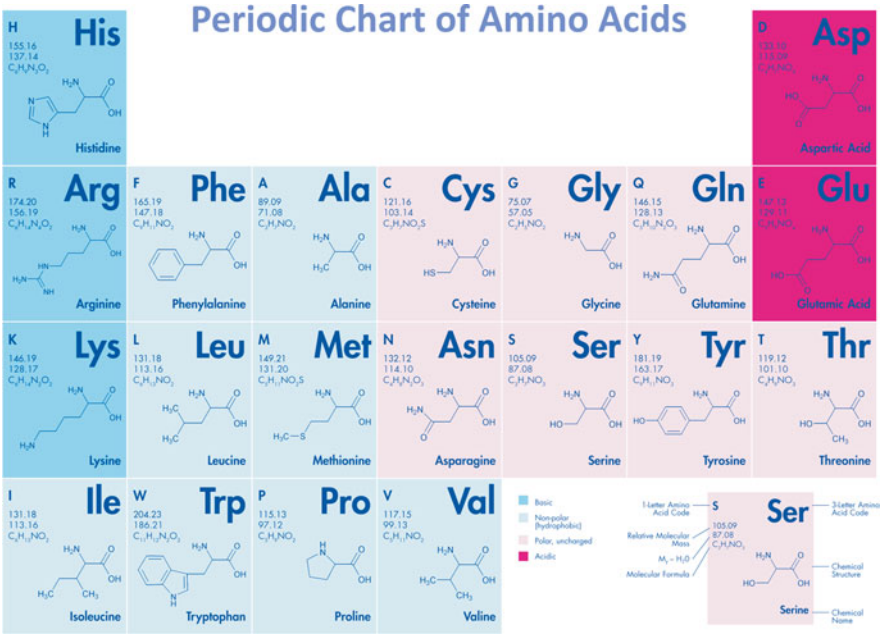


Fig. 3.33 Periodic chart-like arrangement of the natural amino acids. Figure reprinted with the permission of Bachem, Bubendorf, Switzerland

includes grouping of the amino acids according to polarity and charge and shows the chemical nature of the side chains. Table 3.4 clarifies the relationship between the names of amino acids in full and the three-letter and one-letter code abbreviated nomenclature. It also summarizes the most relevant properties of amino acids.

Table 3.4 Natural amino acid nomenclature (three-letter and one-letter code) and main properties. An essential amino acid cannot be synthesized *de novo* and therefore must be obtained from diet

Nomenclature rationale for one-letter code	Amino acid	Three-letter code	One-letter code	Main properties
First letter is unique	**Cys**teine	Cys	**C**	Thiol side chain susceptible to oxidization to form disulfides
	Histidine	His	**H**	Essential amino acid with imidazole side chain. The imidazole side chain has a pK_a of approximately 6.0, which implies that relatively small shifts in most frequent physiologically relevant pH values will change its average charge, which in turn may impact significantly on protein structure
	Isol**eu**cine	Ile	**I**	Essential amino acid isomer of leucine. Chiral side chain
	Methionine	Met	**M**	Side chain possesses a S-methyl thioether, which may be a source of sulphur for cartilage healing. It has been suggested that methionine is able to strengthen the structure of hair and nails because its side chains may cross react
	Serine	Ser	**S**	Residues of serine are found in some phospholipids (besides proteins)
	Valine	Val	**V**	Essential amino acid. Like Leu and Ile, has a branched side chain
First letter not unique. Most frequent amino acids have priority	**Ala**nine	Ala	**A**	D-Alanine occurs in bacterial cell walls and in some peptide antibiotics. Side chain is very small (methyl group)
	Glycine	Gly	**G**	Side chain consists in H, making Gly the only achiral and the smallest possible amino acid
	Leucine	Leu	**L**	Essential branched side-chain amino acid
	Proline	Pro	**P**	The amine nitrogen is bound to two alkyl groups forming a cyclic side chain, which gives proline an exceptional conformational rigidity compared to other amino acids. When proline is involved in a peptide bonding, its nitrogen is not bound to any hydrogen, meaning it cannot act as a hydrogen bond donor, causing a disruption of α-helices and β-sheets
	Threonine	Thr	**T**	Essential amino acid. Chiral side chain. The hydroxyl group in the side chain is prone to glycosylation and phosphorylation

(continued)

Table 3.4 (continued)

Nomenclature rationale for one-letter code	Amino acid	Three-letter code	One-letter code	Main properties
First letter not unique and less frequent: letter with phonetic similarity or side chain chemical nature	**Arg**inine	Arg	**R**	The guanidinium group in the side chain is positively charged at physiological pH ranges therefore prone to binding negatively charged groups. This group has also the ability to form multiple H bonds
	Asparagi**n**e (side chain contains N)	Asn	**N**	Its side chain is curiously an amide (like in peptide bonds). Ows its name to asparagus because it was first detected in aspargus juice
	Aspartate	Asp	**D**	Together with glutamic acid, aspartate is an acidic amino acid because of the carboxylic group in the side chain
	Glutamate	Glu	**E**	In addition to its role in proteins and amino acid metabolism, in neurosciences glutamate is a very important neurotransmitter
	Glutami**n**e	Gln	**Q**	Its side chain is curiously an amide (like in peptide bonds) formed by replacing the side-chain hydroxyl of Glu with an amine functional group.
	Phenylalanine	Phe	**F**	Essential amino acid with a benzyl side chain, which makes it fluorescent and neutral
	Tyrosine	Tyr	**Y**	Tyrosine has a phenol group in the side chain, which makes it fluorescent. More importantly, the phenol group functions as a receiver of phosphate mediated by protein kinases (so-called tyrosine kinases) resulting in alterations in the activity of the target protein.
	Tryptophan (side chain with double ring)	Trp	**W**	Essential amino acid having a fluorescent indole functional group in the side chain. The indole group is bulky and hydrophobic, so Trp is commonly found in lipid-contacting domains of proteins, such as transmembrane regions of membrane proteins or fusion domains of viral proteins
Nearest first letter	**Lys**ine	Lys	K	Essential amino acid. Like in Arg, Lys side chain participates in hydrogen bonding and is cationic at physiological pH range, therefore prone to binding negatively charged groups
Unknown	(Unknown amino acid)	–	X	Unidentified amino acids in peptide or protein structure are generically represented by X

It should be kept in mind that the ionization states of amino acids vary with pH, so depending on pH, amino acids may have different net charges. The example of histidine is presented in Fig. 3.34. Histidine is peculiar as the side chain changes ionization ($pK_a \sim 6$) not far from the range of plasma and cytoplasmic pHs. The intermediate value of the neutrality range (from pH 6 to 9), the so-called isoelectric point, pI, is 7.6, within the range of plasma and cytoplasmic pH range, which happens only for histidine.

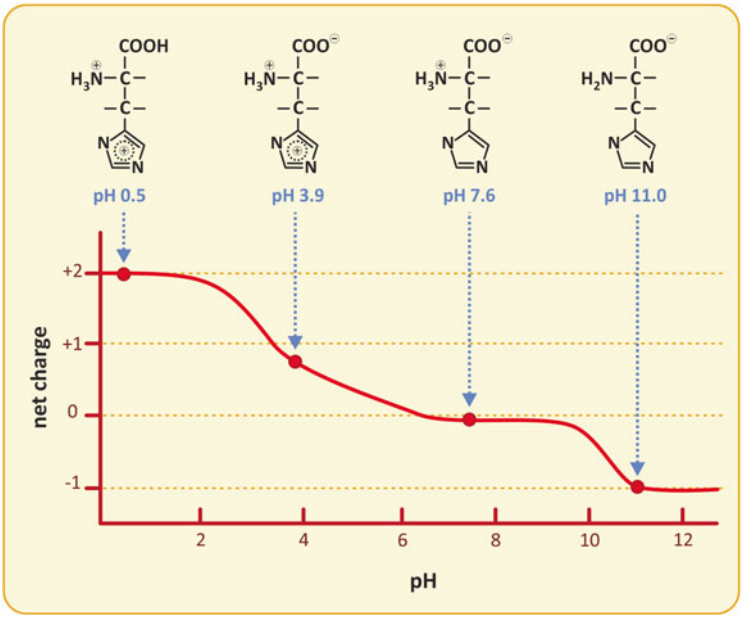


Fig. 3.34 Variation of the net charge of His with pH. The molecule has three ionizable groups, one acidic and two basic. Therefore, allowed global charges range from −1 to +2. However, in most common physiological pHs, the global charge is nearly nil

3.3.1 From Monomers to Polymers: Peptides and Proteins

Amine and carboxylic groups may react, forming amide bonds (Fig. 3.35). Amide bonds connecting several amino acids form a peptide. Many amino acids connected through amide bonds form a protein. There is no precise limit to separate the number of amino acid monomers in peptides and proteins although 30 is usually taken as a reference value.

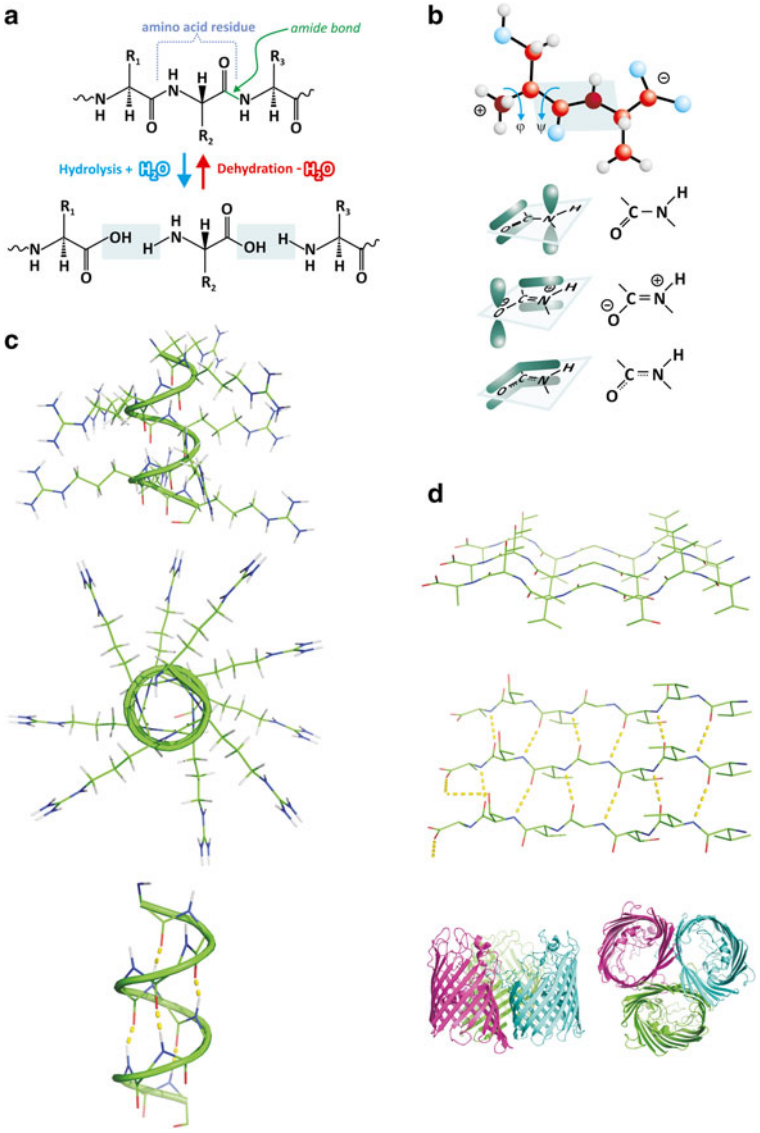


Fig. 3.35 Dehydration reactions among amino acids lead to polymerization through amide (also known as "peptide") bonds (**a**). The CO, CN, and NH bonds are coplanar because of the electronic distribution among the connected OCN set of atoms (**b**). This implies that when the polymeric chain folds, flexibility is limited and specific arrangements tend to be adopted, which include α-helices (**c**) and β-sheets (**d**). Whether a certain sequence of amino acid residues adopts the conformation of α-helix, β-sheet, or any other depends largely on the amino acids involved, their order, and environmental factors such as solvent polarity, pH, and temperature. Both α-helices and β-sheets are conformations that enable the occurrence of frequent intramolecular hydrogen bonding and externalize the location of side chains [(**c**) and (**d**)]. Proteins may be formed almost exclusively of α-helices, such as hemoglobin (see figure 3.42), or β-sheets, such as porin (**d**, bottom) or be a mixture of both. Images of porins are a courtesy of Dr. Claudio Soares, ITQB-UNL, Portugal

Among biochemists, amide bonds forming peptides or proteins are generally referred to as peptide bonds. Because peptide bonds are very planar (C=O, C–N, and N–H bonds are coplanar) due to electron distribution limitations imposed by specific molecular orbitals and have the R groups in close vicinity, the chain of peptide bonds forms a polymer that is not freely flexible. It articulates with spatial constraints, which means that the polymeric chain tends to adopt fixed angles between its amide groups; these angles are the ones that allow accommodating the side chains of the amino acids and adapting the orientation of the amide bonds to each other (Fig. 3.35). As shown in Fig. 3.33, there is a wide diversity of side chains in charge, polarity, and size. All these parameters influence the way a protein folds to cope with the electrostatics, hydrogen bonding, entropic effects (hydrophobicity), and occupation of 3D space. In the end, all these factors determine that amino acid polymers have two different favored kinds of regular conformations: α-helices and β-sheets. Many other folds exist but are not as common because these two are the ones that better accomodate the stabilization of amino acid sequences.

In a α-helix, the peptide bond sequence (i.e., the peptide or protein "backbone") adopts a helical structure projecting the side chains, R, to the exterior of the helix. It is a very stable structure because there are almost no constraints to spatially accommodate R and because the vast array of C=O and N–H groups in the backbone interact strongly through frequent hydrogen bonds. Many proteins, such as hemoglobin, are composed of several helical segments in their amino acid residues sequence. To facilitate protein representation and reading, helical segments are usually represented as a helical ribbon or a cylinder. This highlights the conformation of the segments, although it overlooks what specific amino acids are involved.

β-sheets are extended conformations that turn in specific points resulting in several linear amino acid residue sequences antiparallel to each other. Like in helices, this enables frequent hydrogen bonding in the protein backbone and projection on the side chain groups to the exterior of this compact arrangement. A certain degree of bending is allowed, and big extensions of β-sheets are usually associated to very stable proteic structures, such as membrane pores. β-sheet representation is usually done with straight ribbons and/or arrows (Fig. 3.35d).

The complete protein structure is described in three or four levels. The primary level is simply the sequence of amino acids that compose the protein, conventionally counted from the free amine terminal to the free carboxyl terminal. This elucidates the chemical nature of the protein but tells us little about what are the domains engaging α-helices, β-sheets, or none, which form the secondary-level structures. α-helices and β-sheets from different parts of the same protein tend to interact with each other toward mutual stabilization by means of electrostatic forces, hydrogen bonding, and entropic effects contributions (Fig. 3.36). The tertiary level arises there from α-helices, β-sheets, and other local arrangements that organize in space to form the protein structure itself. Occasionally, there are different parts of the

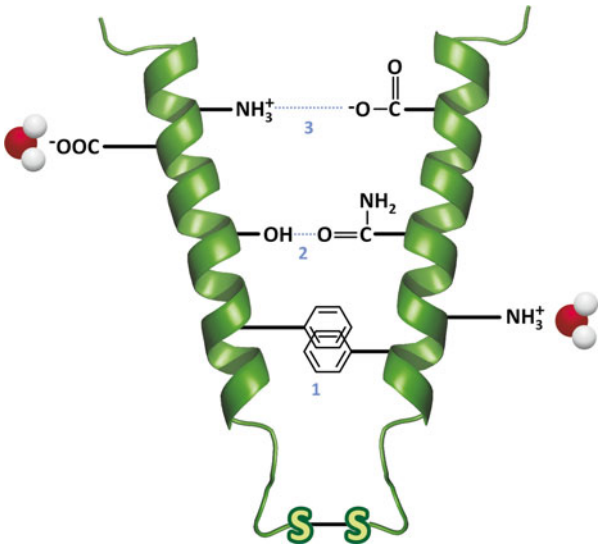


Fig. 3.36 Secondary-level structures such as helices interact intramolecularly or intermolecularly through electrostatic forces (3), hydrogen bonding (2) or entropic effect factors that result in exposure of polar groups such as $–COO^-$ and $–NH_3^+$ to aqueous solvent, and association of hydrophobic groups with minimal exposure to the aqueous environment (1). Two Cys residues in contact may react through the thiol groups (–SH) in the side chains forming disulfide bonds (S–S) that strongly contribute to the stabilization of the structure of proteins (see as an example the structure of insulin in Fig. 3.37)

global geometry of the protein that form fairly independent and separable parts, frequently having specific dynamics and specific functions. These are known as domains. An upper level exists for proteins that associate with other proteins, equal or not, to form organized protein assemblies: the quaternary-level structure. The different levels for protein structure are illustrated in Fig. 3.37, using as example insulin, whose structure was discovered by Dorothy Hodgkin, whom also discovered the structure of cholesterol (see Fig. 2.10).

Hydrogen bonding is frequently the strongest non-covalent factor in keeping the tertiary and quaternary levels of the structure of proteins. Enolase is a good example. Although there are no covalent bonds between both proteins in the dimer,

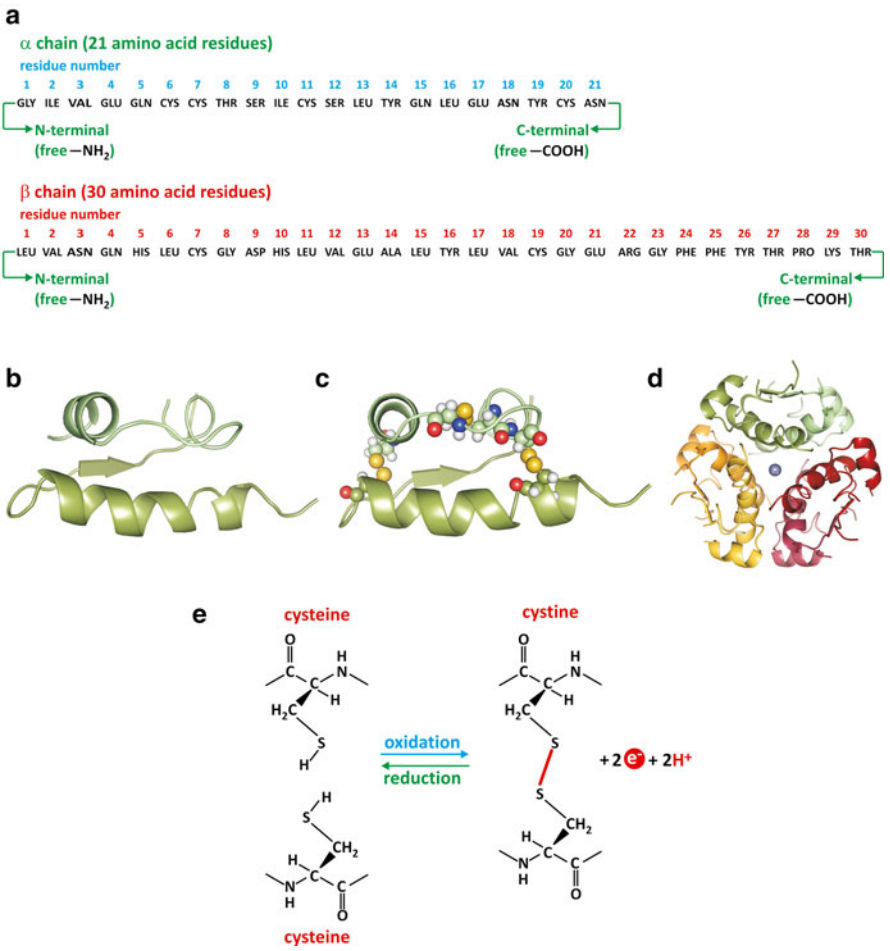


Fig. 3.37 The hormone insulin, from primary to quaternary-level structure. (**a**) Amino acid (three-letter code) sequence, the primary-level structure. (**b**) Segments engaging α-helical secondary-level structure are represented as helical ribbons. (**c**) The protein folds into a tertiary-level structure that is stabilized by disulfide bonds (*yellow* in the protein structure). Disulfide bonds are the result of oxidation of two thiol (–SH) groups to form an S–S bond (**e**). It is common that Cys residues react this way in extracellular proteins. (**d**) Six insulin monomers associate forming a homohexamer, the quaternary-level structure. The quaternary-level structure is stabilized by the presence of two zinc ions (central sphere) and due to contacts between hydrophobic surfaces of monomers (entropic effect). Insulin is stored in the pancreatic beta cells and secreted into the bloodstream in the form of aggregates of these compact hexamers. Upon dilution in the blood, insulin dissociates, and the active form is believed to be the monomer

hydrogen bonds are frequent (Fig. 3.38). Altogether, the sum of all hydrogen bonds creates a strong network of adhesion forces in the contact surface of the proteins. Hydrogen bonds are directional; they occur in a well-defined direction between chemical groups at a definite distance; this further contributes to maintain the structure of proteins. The extreme contribution of hydrogen bonding to polymer structure

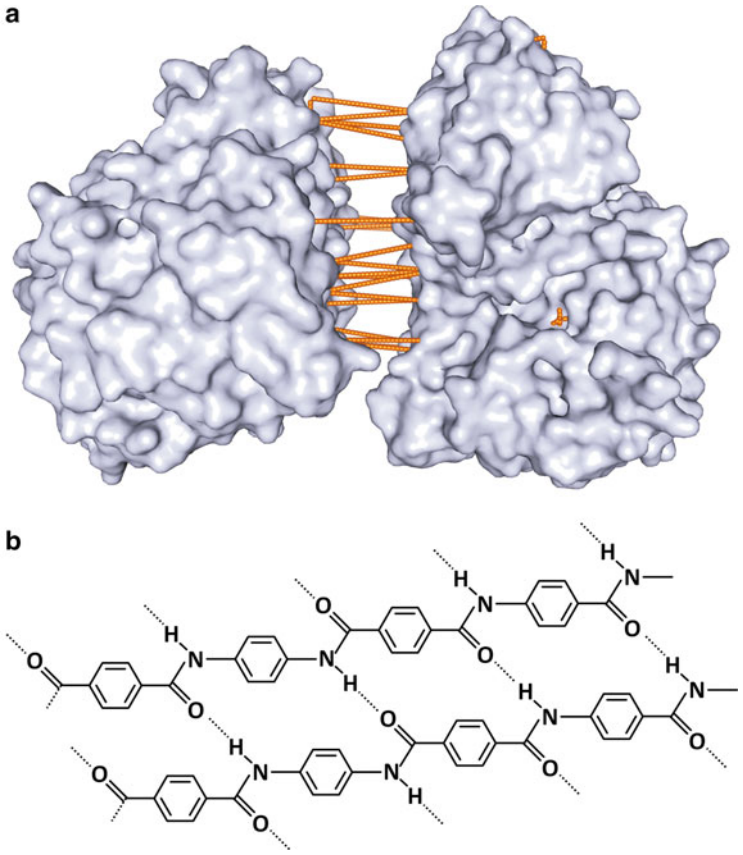


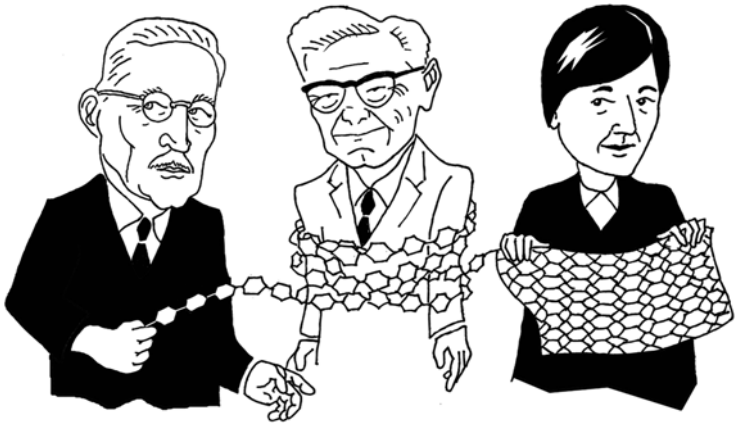
Fig. 3.38 Examples of hydrogen bonding contribution to polymer structure. (**a**) Enzyme enolase (PDB 1IYX) is a dimer in which both subunits are attached by a dense array of hydrogen bonds in the contact surface between them. (**b**) The structure of an amide polymer (such as proteins) commercially known as Kevlar. It involves a dense network of hydrogen bonds, which confers high resistance, and Kevlar is used in protective materials such as helmets and bullet proof vests. Compare the molecular-level details of Kevlar and β-sheets in proteins; there is a parallelism between the resistance of Kevlar and the extreme stability of aggregates formed by the juxtaposition of β-sheets in amyloid plaques (see Box 3.6)

may not be intuitive, but one should bear in mind that Kevlar, an extremely resistant material used in protective items such as bulletproof vests, owes its properties in part to hydrogen bonding (Fig. 3.38).

Kevlar was named after its inventor, the chemist Stephanie Louise Kwolek, whom had planned to attend a medical school but started a temporary job in chemistry and finally quit a medical career. The historic parallelism between artificial polymeric materials and biological molecules dates back to 1920, when Hermann Staudinger proposed that rubber and other polymeric molecules such as starch, cellulose, and proteins are long chains of short repeating molecular units linked by covalent bonds, a disruptive concept at that time. Staudinger used the term macromolecule

("makromoleküle") for the first time, a term now very popular among biochemists. Paul Flory, a chemist pioneer of the studies of three-dimensional organization of polymers and its relation to dynamics, also worked for the rubber industry during certain periods of his career. His work opened the field to the establishment of structure–function relationships in macromolecular biochemistry. Flory was awarded the Nobel Prize in Chemistry in 1974 "for his fundamental achievements, both theoretical and experimental, in the physical chemistry of macromolecules."

Thinking of natural protein fabrics, such as silk and spider webs, and artificial fabrics made of nylon and other polymers helps us realize that in molecular world the boundaries between nature and human artifacts are extremely tenuous.



As mentioned before, there are also covalent contributions to the tertiary level of structure of proteins, namely, disulfide bonds (or "bridges") and attachment of metal or other non-proteic groups to more than one amino acid residue. Disulfide bonds are formed by oxidation of two contacting Cys thiol (-SH) groups originating an S–S bond between the Cys residues (cystine). Cell cytosol is a relatively strong reducing environment, and the contribution of disulfide bonds to the stabilization of cytosolic proteins is limited. However, in other circumstances, disulfide bridges form and are strong stabilizers of protein structure at the tertiary level. Insulin, a peptide hormone, is an example (see Fig. 3.37).

Metals can bind multiple ligands and covalently link different amino acid residues in a protein, therefore also contributing to stabilize a tertiary-level structure. Frequently, metals bind to the thiol group of Cys. In the electron transfer chain proteins, several metallic complexes are present, which in addition to chemical functions also contribute to the stability of the proteins (Fig. 3.39; see also Sect. 6.2.2).

When one refers to quaternary-level structure, one usually refers to proteins that associate with high specificity and well-defined function, such as hemoglobin. This does not include pathological cases in which aggregation of proteins leads to loss of function and increase in toxicity. Extensive tertiary-level alterations are observed when amyloid fibers form upon aggregation of proteins or when prions trigger

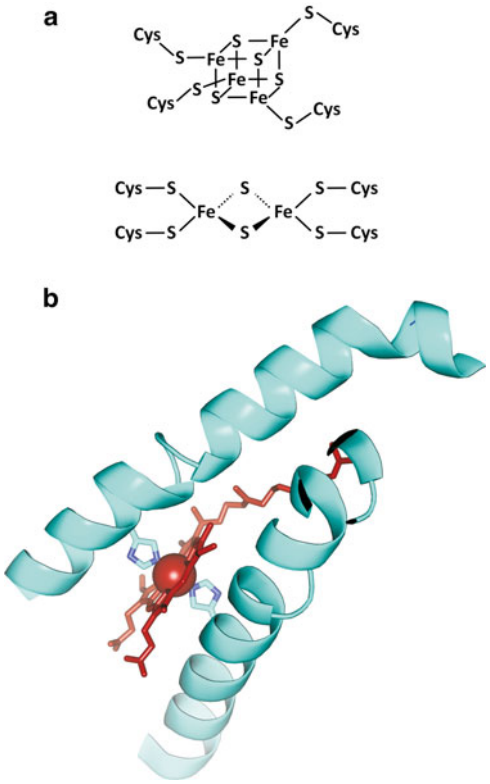


Fig. 3.39 Metal complexes, such as iron–sulfur centers (**a**), are common among the proteins of the electron transfer chain. Iron complexes with sulfur atoms but also with the thiol group of the side chain of Cys residues, resulting in stabilization of the structure of the proteins where they insert. The nitrogen atoms in the side chains of His are also prone to complex formation with metals. Panel (**b**) shows a detail of complex formation with metals in complex IV (PDB 1OCC) of the electron transfer system (see also Sect. 6.2.3). An iron ion (*red sphere*) complexes simultaneously the N atoms of two His side chains (*blue*), stabilizing the tertiary-level structure of the protein. It also binds to a nonprotein molecule, the heme a (*red organic structure*), which is also associated to complex IV

conformational changes of native proteins (see Box 3.6), for instance. These are referred to as protein-folding diseases. Folding is the expression used to comprise secondary- and tertiary-level structure altogether.

3.3.2 Structure and Function in Proteins

Proteins can adopt many different structures at the tertiary level, from extended rods to compact globules. Extended proteins may associate in fibers and globular proteins may have flexible domains able to bind other molecules. This gives the impression that proteins with extended conformation, like keratin, collagen, or silk fibroin, are adequate to maintain the structure of tissues or biomaterials, whereas globular proteins intervene in dynamical processes, which explains why enzymes

Box 3.6: Amyloids and Prions: When Misfolding Turns into Disease
The relationship between the structure and function of proteins has been one of the main issues of modern biochemistry for decades. Mutated proteins may have important changes in their structure and may thus display a defective function. However, the knowledge that proteins without mutations can fold in diverse forms, some of them pathogenic, is recent. Protein folding is the key to important diseases such as Alzheimer's, in which massive stacks of β-sheet-folded proteins accumulate in the brain. These stacks form plaques of insoluble protein in the extracellular tissue, which cannot be broken down by enzymes. When these plaques were found for the first time, they were described as related to saccharides and named amyloids. Although the chemical nature of the plaques is now known not to be related to saccharides, the name "amyloid" is still used and the group of diseases is known as amyloidoses.

Amyloid plaques grow with an ordered structure forming long filaments (fibrils). There are about 20 different proteins that can act as the building blocks of these fibrils, each of which is associated with a different disease. In so-called systemic amyloidoses, the precursors of these plaques are transported through the bloodstream from their point of origin to their point of deposition. Localized amyloidoses are of great clinical significance, as they mainly affect the central nervous system, the extracellular tissue of which is particularly susceptible to damage.

Transmissible spongiform encephalopathies (TSEs), which include mad cow disease (bovine spongiform encephalopathy; BSE) and Creutzfeldt–Jakob disease (CJD) in humans, are forms of amyloidoses in which the diseased brain degenerates to a porous sponge-like structure. These diseases appear when human proteins called prions misfold. The human prion is a component of the membrane of healthy nerve cells (called PrPc) that may misfold in a particular way. Amazingly, the misfolded prion may induce misfolding in a neighboring prion if contact among both molecules occurs. This has the appearance of an infection-like process in which the misfolded molecule "infects" the "healthy" molecule. "Infectious" prions can be transmitted in the diet, triggering a domino effect on healthy prions.

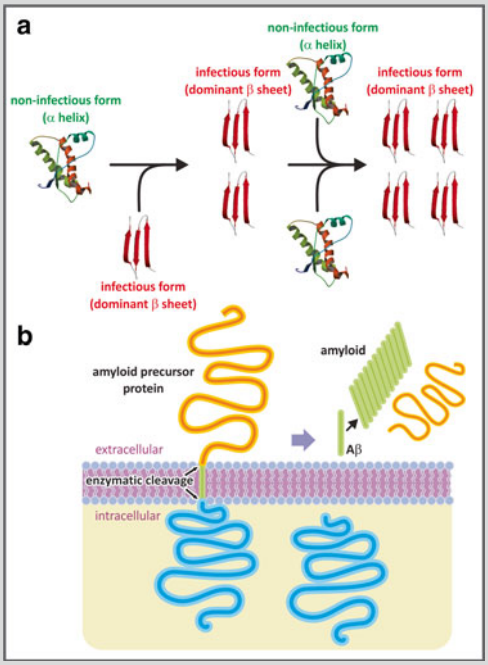
In Alzheimer's disease β-amyloid plaques are formed by cleavage of the amyloid-precursor protein (APP) by two different enzymatic activities, which release peptide fragments that are 40 or 42 amino acids long. When these peptides fold into β-sheets and aggregate, fibrils are formed, surrounding neurons and causing damage. This does not happen when the same peptides fold differently. It is only in β-sheets that hydrophobic amino acids are exposed, and they rapidly bind to hydrophobic groups of other peptides due to the entropic effect. The β-sheet structure, being highly ordered, is prone to regular stacking, ultimately leading to fibril formation.

In panel (a) of the following figure, a prion protein domain PrP(121–231) is shown in the noninfectious form (mouse, PDB 1AG2). The infectious form, having a structure not known in detail but dominated by beta strands

(continued)

Box 3.6 (continued)

(tentatively depicted; red), may induce conformational changes in the noninfectious protein to produce a replica of itself. The process may amplify and aggregates of the infectious form thus accumulate in pathological conditions. In Alzheimer's disease, β-amyloid plaques are not formed by an infection-like propagation of the proteins dominated by beta-strand domains. Panel B shows cleavage of the APP with concomitant release of peptides (green) that fold into β-sheets and aggregate.



are globular proteins. While this is generally true, it should also be acknowledged that the frontier between both is not always clear. For instance, actin is a globular protein that binds other actin molecules to form quaternary-level fibers in the cytoskeleton and muscle contractile system. Actin is an example of a globular protein having structural functionality (Fig. 3.40; see also Sect. 10.1.1). Another interesting

Fig. 3.40 (continued) exposed in collagen and form cross-links that strengthen the collagen fibers. The chemical process of this cross-linking reaction between the endings of the Lys side chains depends on vitamin C (ascorbic acid). Lys residues are also involved in cross-links of elastin (**d**), a connective tissue protein of unusual elasticity. The cross-links involve four Lys residues forming a desmosine arrangement. Actin (**e**) is a globular protein that self-associates forming fibers important for muscle contraction. Myosin (**f**), on the other hand, is a combination of an extended domain with a globular domain ("head") having catalytic activity. The extended domain associates to other proteins forming fiber-like oligomers responsible for muscle contraction (see Sect. 10.1.2) Hyp - Hydroxyproline

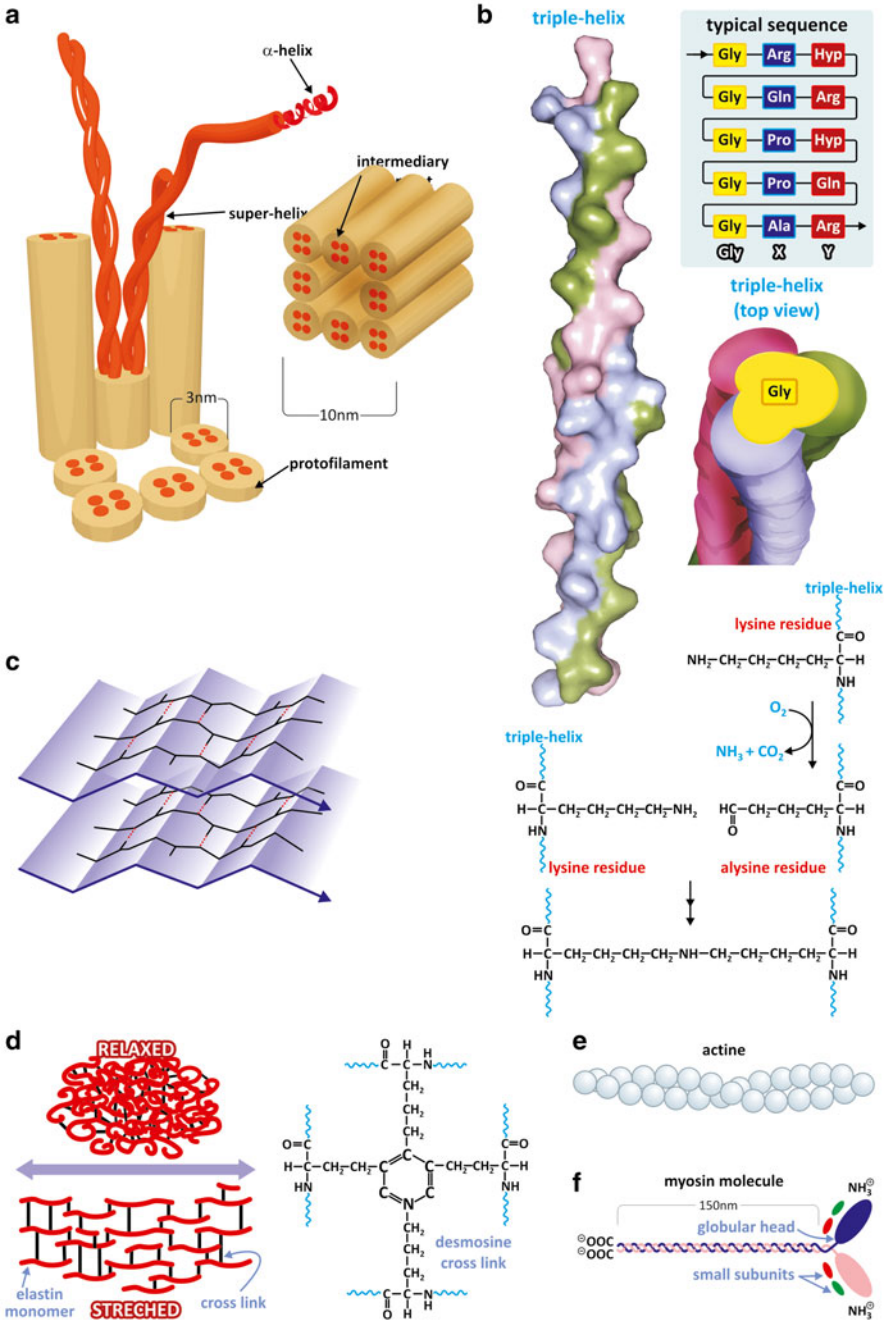


Fig. 3.40 Proteins such as keratin (**a**) and collagen (**b**, PDB 1BKV) adopt string-like conformations (very extended helices) that associate to form fibers that stabilize structures such as hair or nails (keratin), or the connective tissue (collagen). Silk fibroin (**c**) forms very extensible Gly- and Ala-rich β-sheets that associate in an antiparallel fashion forming a very resistant biomaterial such as spider silk. It is important to realize that collagen has a Gly residue at every third position of its amino acid sequence. Gly is a very small amino acid because the side chain is H. This creates a line along the helix surface where two other similar helices may dock in close contact. Lys side chains are

molecule that challenges the classical dichotomy of structural/extended vs. functional/globular proteins is lung elastin. When relaxed, elastin is a globular protein but it stretches allowing lung expansion. Elastin molecules are interconnected covalently by the side chain of four Lys residues: cross-linked desmosine bonds (Fig. 3.40). In this way, the continuous alteration between the globular and extended conformation of elastin confers to the lung the ability to expand and contract without histological lesions. Finally, one should stress the fact that there are also proteins that have both extended and globular domains. This is the case of myosin, another central protein in muscle contraction, in which an extended domain forms oligomeric fibers (Fig. 3.40; see also Sect. 10.1.1).

Besides tight parallel packing in collagen fibers, proteins bind covalently to other proteins in different fibrils of collagen through side chains of Lys residues (Fig. 3.40). When these strong covalent meshing of collagen is disrupted, the properties of the connective tissue are very much affected causing diseases (Box 3.7). Likewise, mutations of the Gly residues impact dramatically on collagen structure and connective tissue function. This has a particular effect on bones from early age: They loose resistance and break easily. This disease is known as osteogenesis imperfecta, Lobstein syndrome, or, more commonly, "brittle bone disease."

Box 3.7: Scurvy: An Example of a Pathology Directly Associated to Protein Structure

Scurvy is a pathology characterized by fatigue, anemia, gingivitis (gum disease), and skin hemorrhages caused by diets with a prolonged deficiency of ascorbic acid (vitamin C). It was a frequent disease of sailors in long voyages during the pioneering intercontinental discoveries of the fifteenth century. Many men died until it was discovered that scurvy could be cured and prevented by consuming citrus, such as oranges, lemons, and limes. The Portuguese sailor Vasco da Gama was the first European to lead a fleet that reached India by sea, linking Europe and Asia and the Atlantic and the Indic. The drama of scurvy in his first trip to India (1497–99) is eloquently described in an epic poem by Luís de Camões, in *The Lusiads* (1572):

"And 'twas that sickness of a sore disgust,
the worst I ever witness'd, came and stole
the lives of many; and far alien dust
buried for aye their bones in saddest dole.
Who but eye-witness e'er my words could trust?
of such disform and dreadful manner swole
the mouth and gums, that grew proud flesh in foyson
till gangrene seemed all the blood to poyson:
"Gangrene that carried foul and fulsome taint,
spreading infection through the neighbouring air:
No cunning Leach aboard our navy went,
much less a subtle Chirurgeon was there;
but some whose knowledge of the craft was faint
strove as they could the poisoned part to pare,

(continued)

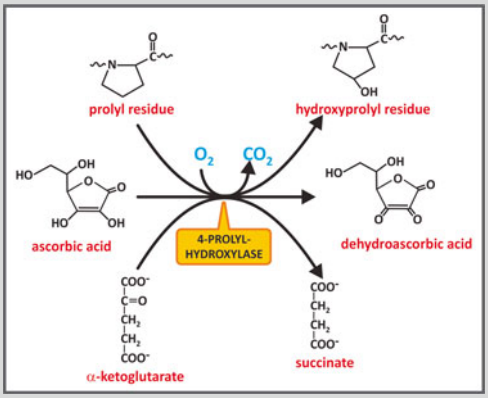
Box 3.7 (continued)

as though 'twere dead; and here they did aright; —
all were Death's victims who had caught the blight.
"The Lusiads" (Canto V, 81 and 82) version reproduced here was translated to English in 1880 by Richard Burton

Nearly two thirds of the sailors of the entire fleet of Vasco da Gama died during the trip, although documents from that time clearly show that the by Portuguese sailors knew that a diet based on fruits and other unprocessed foods was a treatment for scurvy. A manuscript from a pilot in the fleet of Pedro Álvares Cabral, discoverer of Brazil in 1500, says that a diet of fresh foods, including sheep, chicken, ducks, lemons, and oranges, was used to heal scurvy.

It was only in the eighteenth century that the Scottish doctor James Lind related scurvy to diets poor in citrus on a reasonably scientific way. Ascorbic acid was discovered by the biochemist Albert Szent-Gyorgyi (Born Hungarian, later US citizen), who was awarded the Nobel Prize in Medicine or Physiology in 1937 (Szent-Gyorgyi also performed important studies on muscle contraction; see Box 10.1). Ascorbic acid takes part of in biochemical pathways, the synthesis of collagen being one of them. Specifically it is mandatory in protein hydroxylation, which is a posttranslational modification in which a hydroxyl group (–OH) is added to a protein residue. Collagen is naturally hydroxylated in healthy individuals. Ascorbic acid is also mandatory in the biosynthesis of carnitine. Impaired synthesis of carnitine and collagen accounts for the common symptoms of scurvy.

The primary defects behind rotten or loose teeth, rigid tendons, or cartilage fragility observed in scurvy reside in the connective tissue. Without ascorbic acid, collagen is not hydroxylated, and a nonfibrous, defective incomplete collagen is formed instead of fibrous collagen due to the impairment of cross-linking. The enzyme prolyl 4-hydroxylase, for instance, hydroxylates a Pro residue using an iron atom that is oxidized in the process. Ascorbic acid is needed to reduce the iron and make the enzyme active again (see figure).



(continued)

Box 3.7 (continued)

Lysyl hydroxylases are also operative. Hydroxylation of Pro and Lys residues, both exposed in the triple helix of collagen, favors intermolecular adhesion interactions by hydrogen bonding and further chemical modifications, which are strengthened with age (this is one of the reasons why meat from young animals is tenderer than from older animals).

Carnitine is involved in the transport of fatty acids into the mitochondria, where they are oxidized (see Sect. 7.4.3). Ascorbic acid is used by two different enzymes in the carnitine biosynthesis. Without ascorbic acid, production of carnitine declines and fatty acids cannot be used as energy source. This leads to fatigue, which is one of the symptoms of scurvy. Curiously, fatigue appears prior to other symptoms. This may be explained by the fact that the enzymes in carnitine biosynthesis require higher concentrations of ascorbic acid to function (they have "lower affinity" for ascorbic acid) when compared to hydroxylases.

Other pathologies, such as the Ehlers–Danlos syndrome (EDS) and osteogenesis imperfecta (OI), are genetic pathologies associated to deficiencies on protein–protein interactions in collagen. A fraction of the Lys residues in collagen react with one another forming covalent cross-links in collagen fibers. In some forms of EDS, this cross-linking is impaired, rendering the skin less firm and less resistant, hyperelastic. Collagen contributes to the mechanical strength of the skin, joints, muscles, ligaments, blood vessels, and visceral organs. In OI, replacement of Gly residues in the collagen amino acid residue sequence destroys the capacity of the protein to assemble in perfect triple helices because all other residues are bulkier than Gly. This leads to an extremely severe condition that is characterized by alterations in the physical properties of collagen and perturbations in the biochemical processes involving collagen homeostasis. The relationship between the collagen fibrils and hydroxyapatite crystals when bones are formed is altered, causing brittleness. For this reason, OI is also known as "brittle bone disease."

One remarkable property of many globular proteins is the ability to both bind other molecules and change conformation upon binding. Taking adenylate cyclase as example, binding of molecules such as ATP or ADP causes very mobile domains of the protein to change position (Fig. 3.41). This often leads to dynamic distortions of the ligand molecules because of the contact with amino acid residues. The results may be such that covalent bonds are formed or broken, thus chemically transforming the ligand molecule into a product. In practice, the protein action is to increase the rate of transformation of the ligand in the product. If only the ligand, not the protein, is chemically transformed in this process, this can be seen as an enzymatic catalysis, i.e., increase of the velocity of chemical reactions caused by proteins, the enzymes. In these cases, the ligand (reactant) is named substrate.

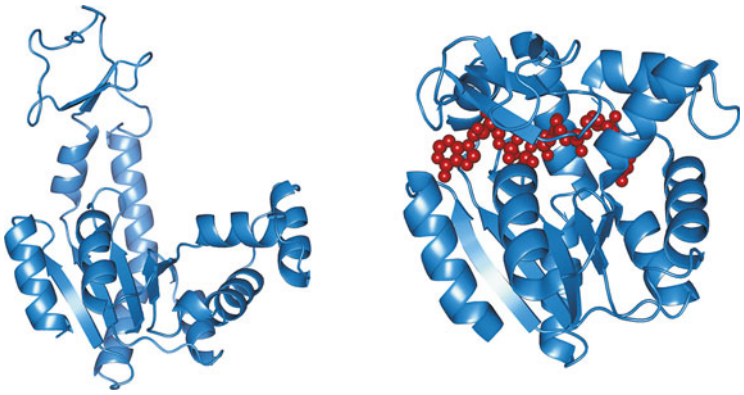


Fig. 3.41 Adenylate cyclase (or adenylyl cyclase), unbound (PDB 4AKE; *left*) and bound (PDB 1AKE; *right*) to a dinucleotide analogue (*red*). Upon binding to the dinucleotide analogue, the mobile domains adapt by changing position

3.3.3 Cooperative Interplay Between Tertiary-Level and Quaternary-Level Structure

As discussed in the previous section, the tertiary-level structure of a protein is determined and maintained by arrays of sites in which attractive or repulsive forces between groups of atoms exist. This is a relatively delicate balance. When a significant number of such "force spots" are altered, the conformation of the protein adapts by adopting a different tertiary-level structure, which corresponds to the new balance of forces. Likewise, in cases in which a quaternary-level structure exists, the changes in the conformation of one protein monomer at the surface of contact with other monomer may impose alterations in the "force spots" (hydrogen bounds, electrostatic repulsion and attraction, entropic factors, etc.) so that the second monomer changes conformation to adapt. In practice, this means that conformational changes in one protein may be transmitted to a neighboring protein that is in contact with it. In other words, tertiary-level structural changes may be transmitted and amplified to other proteins throughout the quaternary-level structure. Hemoglobin is a good example. It is formed by two subunits, α and β, forming a dimer that associates to other dimer—a tetramer that is in fact a dimer made of dimers. Dimers are numbered 1 and 2; therefore, hemoglobin is a tetramer of four subunits β1, α1, β2, and α2 (Fig. 3.42). α – β attractive forces are stronger than 1–2 attractive forces, but both are sufficient to transmit to neighboring monomers conformational changes. This affects the affinity of the hemoglobin monomers to oxygen. Each hemoglobin monomer is covalently associated to a non-proteic

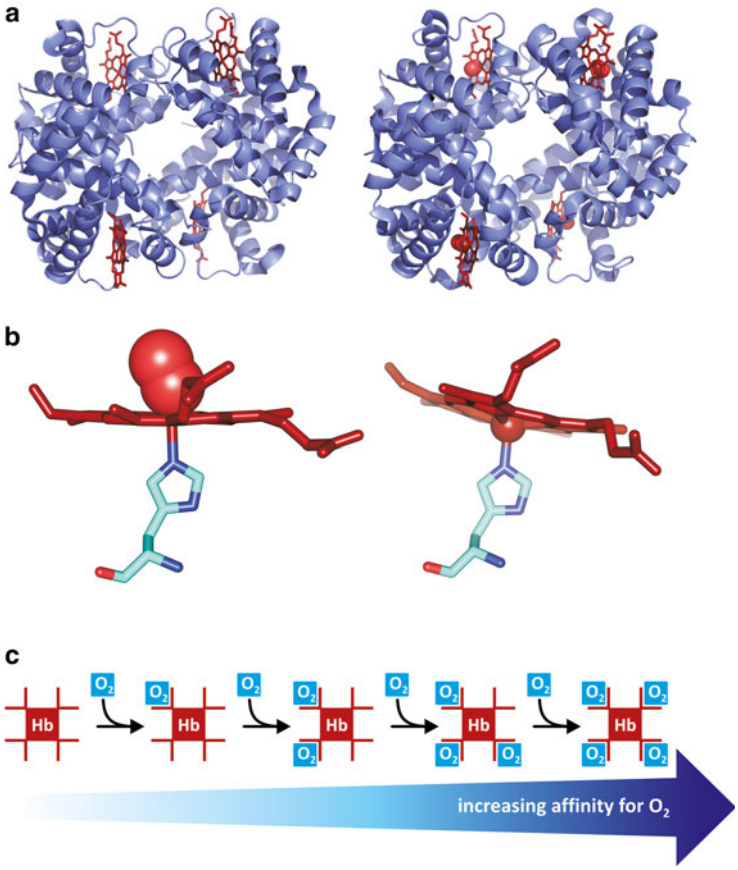
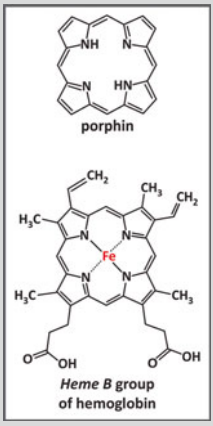


Fig. 3.42 Hemoglobin is a tetramer (**a**), each subunit containing a heme group (*red*). Human deoxyhemoglobin (*left*; PDB 2HHB) and oxyhemoglobin (*right*; PDB 1GZX) show subtle but important changes in conformation. Binding of molecular oxygen to the heme group [(**b**), *left*; PDB 1HHO] causes a shift in the orientation of the heme relative to the His residue when compared to the heme in deoxyhemoglobin [(**b**), *right*; PDB 4HHB). This slight distortion in the position of the heme leads to a variation in the conformation of the protein, which propagates to neighboring monomers in the tetramer. The neighboring monomers acquire higher affinity for O_2 (**c**)

group, i.e., a prosthetic group, of the porphyrin family (Box 3.8). In this case the porphyrin binds in its center an iron ion, forming a heme. The iron ion complexes with the heme through four bonds in the plane of the porphyrin and to a His residue side chain orthogonally to the heme plane. Another orthogonal bond, opposite to His is established with small molecules having electron donor atoms, such as O_2 or CO. When an O_2 molecule binds to the iron in the heme, the position of the iron slightly shifts, which in turn affects the position of the His residue. When the His

Box 3.8: The Importance of Heme Groups in Proteins.



Many natural proteins are associated to prosthetic groups of similar chemical nature called porphyrins. Porphyrins are macrocyclic compounds related to porphin (see figure). Hemoglobins, for instance, bind porphyrin groups, such as heme *b* (see figure). Hemes are porphyrins that bind iron ions in the center. The remarkable capacity to bind metallic ions of charge +2 or +3 in the center of the ring may explain the success of these molecules during natural selection and subsequent ubiquity of porphyrins in nature. The porphyrin macrocycle has 26 delocalized (π) electrons in total, being classified as aromatic from a chemical point of view. This system of delocalized electrons extends to the nitrogen atoms and is available to bind the cationic metals. It is also responsible for the intense absorption bands in the visible region of electromagnetic radiation. This is the reason why compounds with porphyrins, such as hemoglobin and chlorophyll, are intensely colored. Heme-containing proteins, hemoproteins, and other metal-containing proteins, metalloproteins, due to their unique electronic properties, are adequate for the transient binding of diatomic gases that occurs during their transportation in blood and electron transfer (i.e., electron donation and reception to and from other compounds). It is possible that hemoproteins evolved from ancient proteic forms, whose function was electron transfer in sulfur-based photosynthesis in the ancestors of cyanobacteria before molecular oxygen existed in atmosphere.

In addition to the peculiar intrinsic properties of hemes and other porphyrin–metal associations, the interaction between the porphyrin and the amino acid residues at the site where it inserts in proteins is of extreme importance. Variations in the shape, volume, and chemical composition of the binding site, in the mode of heme binding, and in the number and nature of heme–protein

(continued)

Box 3.8 (continued)
interactions result in significantly different heme environments in proteins having different biological roles. The outcome is a fine-tuning of the heme properties. Take the 3D structure of the hemoglobin chain as an example. The position of a His residue is such that its protonation interferes with oxygen release from the iron ion. Acidification of the medium causes the protonation of His, which in turns facilitates the release of oxygen. This is known as Bohr effect (see Fig. 3.44 in main text) and is not a simple curiosity: In the pulmonary vasculature, the pH is higher than in the peripheral tissues because the pH is affected by the local abundance of CO_2; therefore, the Bohr effects help in increasing the efficacy of binding oxygen in the lungs and releasing in peripheral tissues.

The porphyrin groups are equally important to stabilize protein structure and resistance to proteolysis, although these properties are frequently overlooked. Even in cases in which the porphyrins are not covalently bound to the protein, the interplay between amino acid residues and the non-proteic groups is very specific.

residue is pulled, the whole structure of the protein changes slightly. This change in conformation induces a change in conformation of the neighboring monomers. As a consequence, the neighboring monomers acquire increased affinity to bind an oxygen molecule. So, binding of O_2 to a monomer increases the chances that a second O_2 molecule binds to another monomer in the hemoglobin tetramer relative to a monomer in a tetramer without bound O_2 molecules. This is called positive cooperativity, i.e. several entities influencing each other favouring a certain event.

Hemoglobin monomers interact cooperatively to bind up to four oxygen molecules. Myoglobin, a protein abundant in muscles, also binds O_2, but this protein occurs as a monomer, in contrast to hemoglobin. Comparing hemoglobin to myoglobin makes the effect of cooperativity clear. The fraction of myoglobin binding O_2 relative to total myoglobin increases nearly linearly with partial pressure of oxygen up to saturation. To be more precise, the variation is hyperbolic. In contrast, hemoglobin binds O_2 critically at a triggering narrow concentration range, in which it reaches saturation. Hemoglobin changes from highly unsaturated to almost saturated in a narrow O_2 partial pressure interval. Interestingly, the narrow interval of transition to near saturation corresponds to the partial pressure of oxygen found in peripheral tissues (Fig. 3.43), away from the lung alveoli. Therefore, cooperativity among the hemoglobin monomers enables hemoglobin to saturates with oxygen in the lungs and delivers its cargo to peripheral tissues. Myoglobin would not be adequate for this function as it is almost saturated in both situations. Myoglobin is fit for oxygen storage in muscle cells. Release of oxygen occurs when the consumption in mitochondria is such that the cell is almost depleted in oxygen (Fig. 3.43).

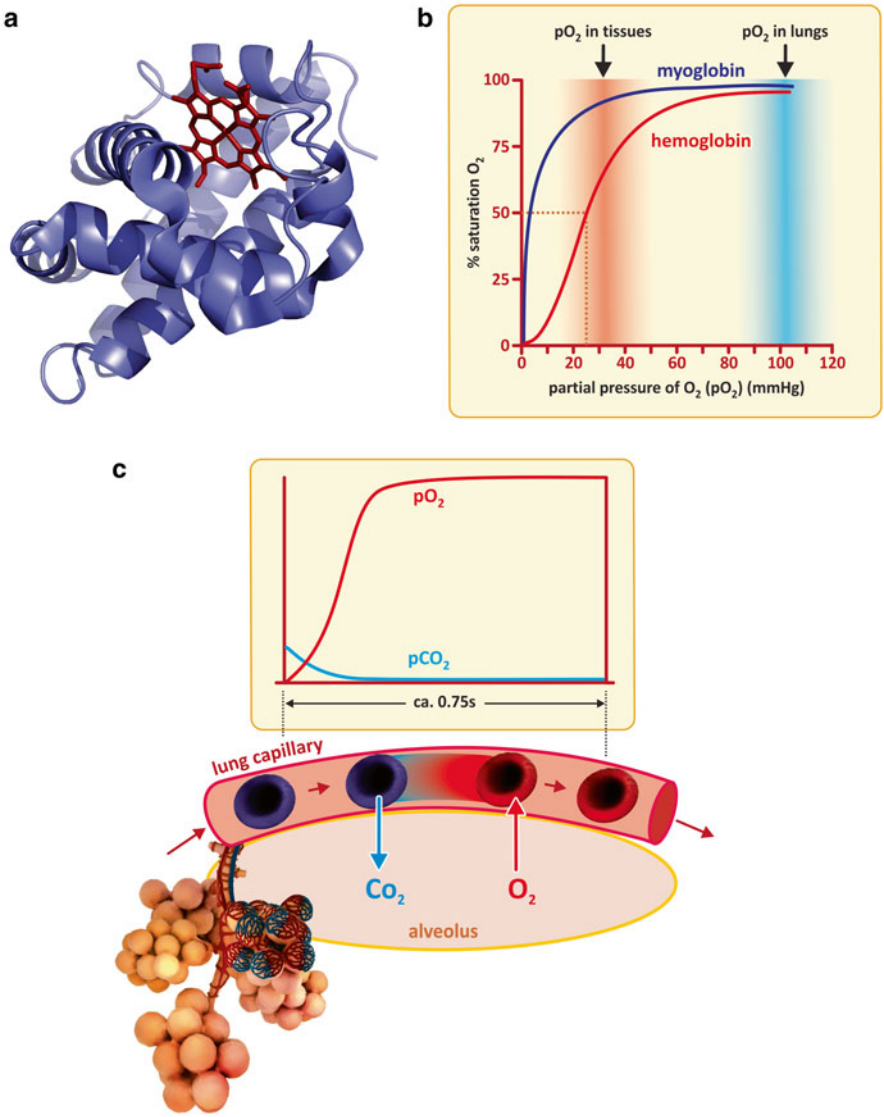


Fig. 3.43 Myoglobin (**a**) is a monomeric oxygen-binding protein. Like a hemoglobin monomer, it also binds oxygen through a heme group (*red organic structure*). The absence of cooperativity in myoglobin when compared to hemoglobin implies distinct binding capacities at different oxygen partial pressures, pO_2 (**b**). Hemoglobin has an abrupt transition from low to high binding when pO_2 changes from values typical from peripheral tissues to values typical of the lungs. As the erythrocytes pass adjacent to alveoli, O_2 and CO_2 diffuse freely across arterial and lung cells driven by partial pressure gradients (**c**)

It is worth stressing that hemoglobin transports most, but not all, oxygen used in tissues. Oxygen, like carbon dioxide, is a small, hydrophilic molecule, which easily dissolves in aqueous media. Although being hydrophilic, it is a very small molecule and diffuses freely in tissues. This is the reason why cells do not need oxygen transporters or channels. The same happens with CO_2 but it has lower affinity for hemoglobin. In addition, CO_2 is converted to HCO_3^- that equilibrates with H_2CO_3 (see Sect. 2.1.1). Therefore, plasma CO_2 transport is not dependent on a specific protein.

Although the direct binding of CO_2 to hemoglobin is not significant, hemoglobin is very important in the chemistry and physiology of CO_2 in the human body as the protein itself is a weak base or weak acid depending on pH. In the peripheral tissues, in which CO_2 is present at higher partial pressure, CO_2 diffuses to plasma and therefore in erythrocytes. Carbonic anhydrase then converts CO_2 to H_2CO_3 that acidifies the medium. Acid pH leads to the protonation of hemoglobin, which has lower affinity for O_2. Near the lung alveoli, plasma CO_2 diffuses to the alveoli due to the gradient in the partial pressure of CO_2 (Fig. 3.43). This drop in plasma partial pressure of CO_2 in plasma causes carbonic anhydrase to convert H_2CO_3 in CO_2, therefore shifting the equilibrium HCO_3^-/H_2CO_3 toward the consumption of HCO_3^- and H^+. The slight drop in pH causes the deprotonation of hemoglobin, which has higher affinity for oxygen. Thus, there is a coupling between pH and the efficiency of oxygen capture, transport, and release (Fig. 3.44). The coupling of hemoglobin structure with pH is known as the "Bohr effect."

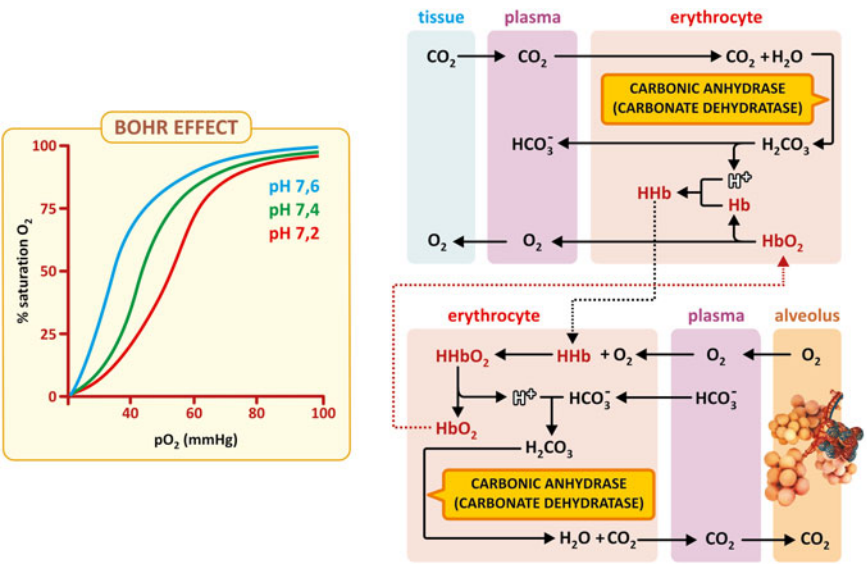


Fig. 3.44 The Bohr effect associated to hemoglobin. CO_2 levels in blood influence the transport of O_2 by hemoglobin through plasma pH because protonation/deprotonation of hemoglobin affects cooperativity in O_2 binding. Lower pH (found in tissues due to the release of CO_2 in plasma) favors hemoglobin protonation and unbinding of oxygen. Hb - Hemoglobin, HHb - protonated hemoglobin

In the same way oxygen binding, transport, and release by hemoglobin are affected by pH, binding of 2,3-bisphosphoglycerate (2,3-BPG) also influences oxygen fixation in a way such that the binding curve of oxygen by hemoglobin is shifted toward higher partial pressures of oxygen (Fig. 3.45). This is far from being a simple curiosity: 2,3-BPG forms from 1,3-BPG, a metabolite of the glycolysis pathway (see Sect. 6.1.3). When glycolysis is highly active, 2,3-BPG is formed and release of O_2 from hemoglobin becomes more effective. This is convenient for the cell as higher glycolytic activity implies, in principle, a higher demand for oxygen by the human cells. The fixation of oxygen in lungs remains unaffected.

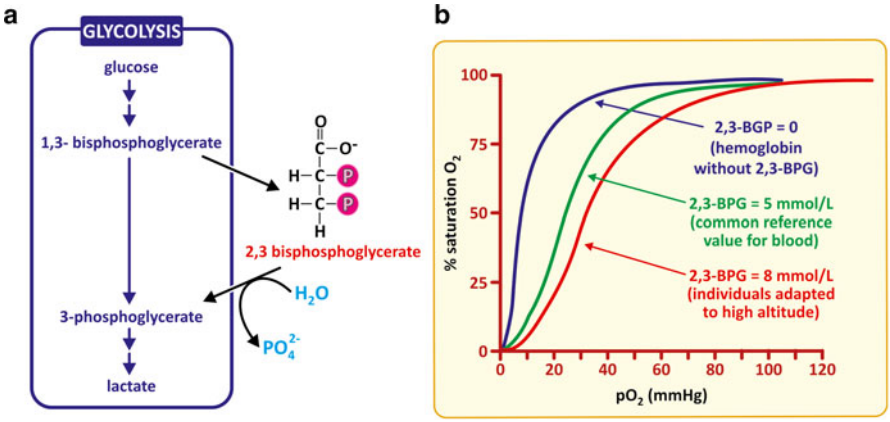


Fig. 3.45 2,3-BPG forms from 1,3-bisphophoglycerate, an intermediate metabolite of glycolysis; it is therefore, a chemical signal of glycolytic activity (**a**). 2,3-BPG binding to hemoglobin affects cooperativity in such a way that the O_2 binding curves are shifted in a way that O_2 release in peripheral tissues is facilitated, but the O_2 captured in the lungs is not affected (**b**)

3.3.4 Enzymes

The previous section showed how dynamic the binding of proteins to other molecules may be. In the case of hemoglobin, oxygen binds to the iron of the heme group. However, frequently more complex molecules bind directly to side chains of certain amino acid residues. Specific sets of amino acid residues in a protein may precisely locate and orient in space and have the correct physical properties (charge, polarity, hydrogen donor/acceptor groups, etc.) to specifically bind molecules that establish attractive forces with them (Fig. 3.46). The electronic clouds of these molecules are distorted by the contact with the amino acid residues, which in turn adapt their tertiary-level structure to the presence of the molecules. This mutual adaptation between protein and bound molecules frequently weakens some chemical bounds of the molecules, which may thus be destroyed. Likewise, formation of other bonds is possible. The result is that the molecule that bound to the protein is

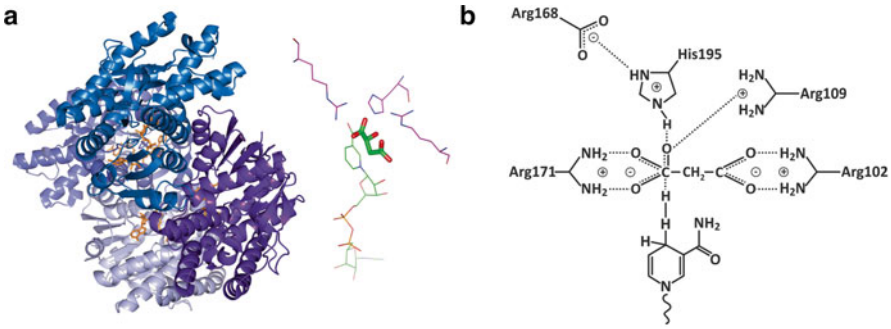


Fig. 3.46 Malate dehydrogenase (PDB 2DFD) is a homodimeric enzyme ([A]) that catalyzes the oxidation of malate to oxaloacetate concomitantly with the reduction of NAD^+ to NADH [highlighted in *orange* in (**a**)]. A specific set of amino acids of the enzyme has the right properties (charge, H-binding ability, etc.), the right location, and the right orientation to simultaneously fit the malate molecule (**b**). This set of residues forms the so-called active site (or active center) of the enzyme. NAD^+ binds to other site, specific for it, in close vicinity to the active site and participates in the oxidation of malate facilitated by the action of the amino acids [(**b**), *bottom structure*]

converted in a different molecule. If the resulting product dissociates from the protein and the protein returns to the same state as before binding the original molecule, then the protein is an enzyme, i.e., a proteic natural catalyst. The initial molecule that binds the enzyme and undergoes a chemical reaction is said to be a substrate, as previously mentioned in Sect. 3.3.2.

Because the enzymes interact with the substrate and facilitate its conversion to products, they increase the velocity of chemical reactions enormously, typically, above 10^7-fold. In some cases, the increase may be 10^{17}-fold, which is a figure difficult to conceive intuitively. Considering that a 2×10^8-fold increase in the velocity of a relaxed walk (~1.5 ms^{-1}) would leave us traveling at the speed of light (~3×10^8 ms^{-1}), this intuitive perception becomes clearer. 10^{17}-fold is more than the difference between a relaxed walk and 100 million faster than the speed of light in vacuum. One will see in Chap. 4 that enzymes accelerate reactions but cannot turn impossible reactions in possible ones. However, enzymes turn very slow reactions (so slow that in practice they seem unable to occur) into fast reactions. So, in practice it is almost like if an impossible one was transformed into a possible reaction by the intervention of an enzyme.

3.3.4.1 The Importance of Studying Enzymes

Enzymes are interesting molecules because the dynamics of their tertiary-level structure implies catalytic activity, which has shaped life as it exists today. Even viruses need enzymes to be effective. Because enzymes are so proficient in speeding reactions, controlling the activity of enzymes is, in practice, controlling the course of chemical reactions in a cell. This is an essential piece to impose order in the chemistry of the cells. Also controlling the activity of enzymes ensures that certain

reactions only occur to a significant extent *when* and *where* the enzyme is inside the cell. This prevents conflicting reactions in a well regulated cell and permits certain reactions to be coupled. Imagine substrate A and substrate B; now imagine that enzyme E_A converts A in B and a second enzyme, E_B, converts B in C. The simultaneous presence of both enzymes in the same cell compartment has the practical consequence that A is converted in C. This coupling of reactions may reach considerable complexity, with many substrates, reactions, and enzymes being involved, sometimes with branched and cyclic reaction sequences (recall Figs. 1.3 and 1.4). Such sets of reactions are referred to generally as "metabolisms." Regulation of metabolism is largely dependent on enzymes. The mechanisms of metabolic regulation are extremely important and will be addressed in Chap. 5. A regulated metabolism is a sine qua non condition for a particular state of organisms we call "health."

Yet, the structure–activity relationship in enzymes and their significance in metabolic regulation is only part of the importance of studying enzymes. Enzymes can operate outside cells and be used in industrial processes in pharma, food, or detergent processing and manufacturing, for instance. More importantly, in biomedical sciences and clinical practice, they can be used as valuable tools for diagnosis. When enzymes that were supposed to be confined in cellular compartments in specific tissues are found with increased levels in plasma, this is a sign of tissue lesions with rupture of cell membranes (and consequent leakage of enzymes to the plasma). The death of cells in tissues implies a constant flow of intracellular contents to the plasma, but in non-injured tissues this occurs to a very limited extent. A severe lesion in the liver, heart, or other organ leads to unusually high increased levels in the plasma of enzymes that are specific for that organ. Prostatic specific acid phosphatase (PSAP, Fig. 3.47), for instance, is an enzyme produced by the prostate that can be found in

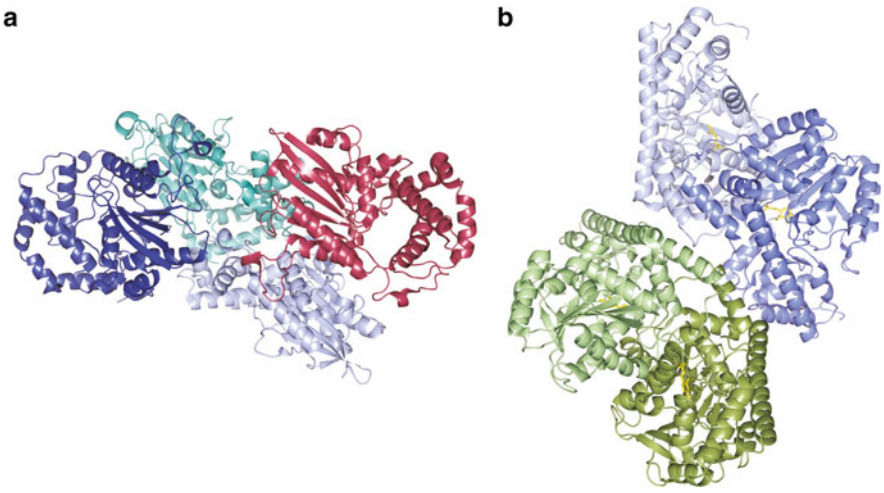


Fig. 3.47 Human prostatic specific acid phosphatase, PSAP [(**a**); PDB 1cvi), and aspartate aminotransferase, ASP [(**b**); PDB 3II0], are markers of prostate cancer and hepatitis, respectively. Pyridoxal 5-phosphate is a cofactor that is shown bound to ASP (*yellow*)

high amounts in the blood of men who have prostate cancer. A short number of other diseases cause moderate increased levels of PSAP, but only direct lesion of the prostate such as that provoked by tumors in this organ causes high levels of the protein in plasma. PSAP is then classified as a marker for prostate cancer.

Being highly irrigated (see Box 8.1) and particularly exposed to the action of drugs and other exogenous chemicals and viruses, the liver is an organ that suffers frequent insults that lead to the presence of hepatic enzymes in plasma. Two transaminases, alanine transaminase (ALT) and aspartate transaminase (AST), are markers of lesions frequently assayed in blood samples when hepatitis, poisoning, or alcoholic liver disease is suspected. However, one should bear in mind that these enzymes are also present in other organs, albeit in smaller concentrations. A full diagnosis is composed of data that takes into consideration not only biochemical analysis but also the symptoms, the history, lifestyle, and other diagnostic results.

The enzymes lactate dehydrogenase (LDH) and creatine kinase (CK) are of particular interest because they are markers of heart muscle lesion. Although these enzymes also exist in muscles other than the heart, there are differences in the amino acid composition that enable detection of the heart variants. Enzymes with similar activity and extensive structure homology are known as isoenzymes. For instance, three CK isoenzymes have been discovered: CK-MM or CK3, found mostly in skeletal muscle; CK-BM or CK2, found mostly in myocardium; and CK-BB or CK1, which is concentrated in the lungs and brain. Because of this distribution of CK isoenzymes, a pulmonary embolism is associated with elevated levels of CK-BB. On the other hand, an acute myocardial infarction is associated with elevated levels of CK-MB, and injuries of skeletal muscle cause elevated levels of CK-MM. LDH isoenzymes are tetramers, both in the heart and other muscles. These tetramers may disassemble and reassemble in the form of mixed heterogeneous tetramers because the structure of the monomers is very similar. In any case, the presence of dominant heart isoenzymes can be detected in plasma in case of myocardial infarction. The plasma enzyme changes in acute myocardial infarction are shown in Fig. 3.48. CK-MB isoenzyme peaks first, AST next, and LDH last.

There are also nonenzymatic markers that are used in the diagnosis of acute myocardial infarction: myoglobin and two cardiac troponins, troponin I (cTnI) and troponin T (cTnT). CK-MB and the heart LDH isoenzyme are the most important ones due to their heart specificity. Cardiac troponins are also important as their serum levels are frequently elevated during the first hours of acute myocardial infarction, even at a time when CK and CK-MB activities are still within the reference range, but are not fully established as the enzymatic markers.

While the scientific and clinical discipline of studying enzymes for direct clinical diagnosis, clinical enzymology, is expanding and gaining importance, it is curious to mention that dead brain tissue does not release into the blood any significant amounts of enzymes. Despite the frequency of cerebral infarcts (strokes), no test for

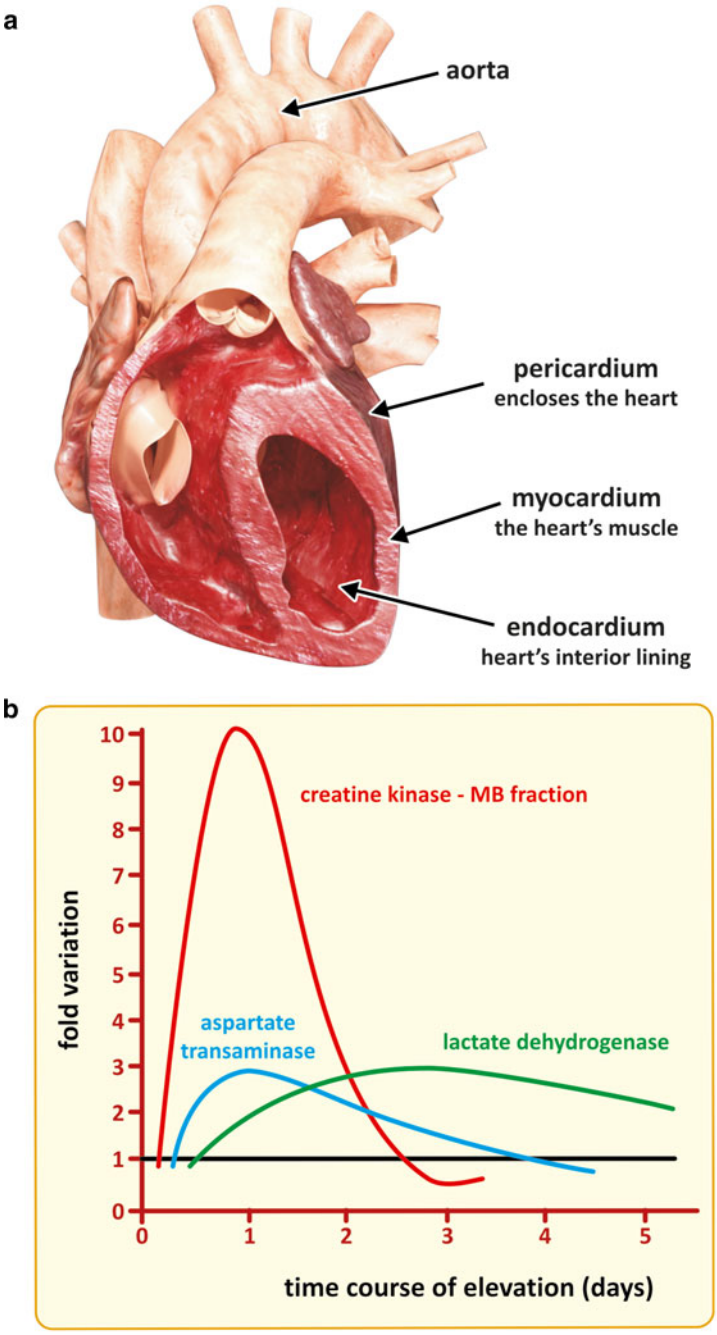


Fig. 3.48 Myocardium is the heart's muscle (**a**). Myocardium infarction involves partial tissue death (necrosis) caused by a local deficit of oxygen supply, as a consequence of an obstruction of the tissue's blood flow. Cardiac muscle enzymes, such as CK-MB, AST, and LDH, appear in the blood after the infarct (**b**). The combined information of CK-MB and LDH allows to estimate the time of the infarct, which in turn helps devising a therapeutical strategy. CK - creatine kinase, AST - Aspartate transaminase, LDH - Lactate dehydrogenase

brain enzymes is currently available due to the blood–brain barrier (BBB, Fig. 3.49), the network of capillaries that irrigates the central nervous system. The cells of these capillaries are connected by tight junctions and adhesion molecules that severely restrict the diffusion of hydrophilic macromolecules into the cerebrospinal fluid (CSF). Small gas molecules, such as O_2 and CO_2, diffuse passively through the barrier, and some nutrients and hormones are actively transported with specific proteins (this will be revisited, for instance, in Box 9.3, which discusses the glucose transport through the BBB).

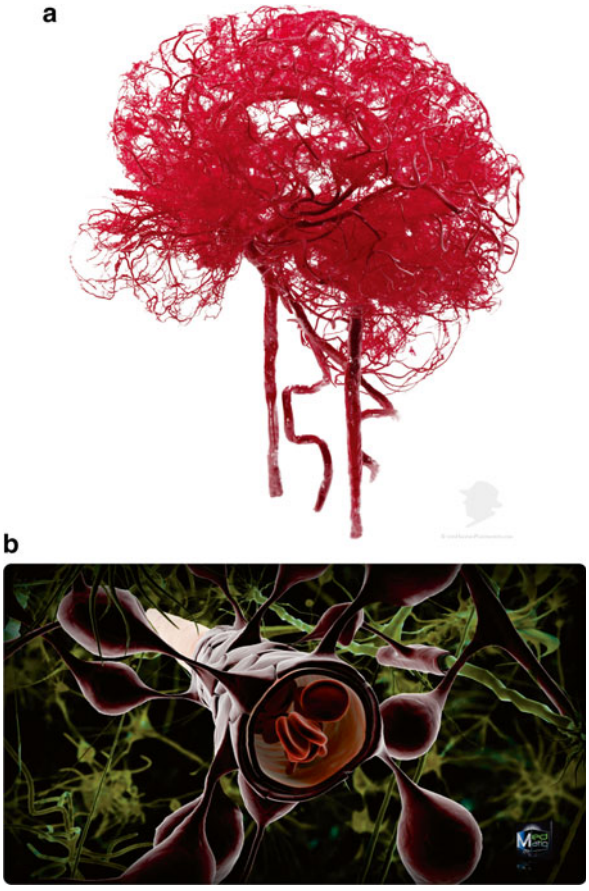


Fig. 3.49 The blood–brain barrier (BBB) is the network of brain capillaries (*top*). This photograph shows the result of a technique used to conserve anatomical structures named plastination applied to the human BBB (reprinted with permission of von Hagens Plastination, Germany; © www.vonHagens-Plastination.com). The network of very thin arteries that penetrate the brain forms a very reticulated mesh. The capillaries are associated to a thick basement membrane and astrocytic end feet (brown cells covering the endothelium; *bottom panel*). Passage of molecules across the endothelial cells of the BBB is highly selective. Enzymes released from nerve cells upon a stroke cannot reach the blood, a fact that makes diagnosis very difficult. Bottom panel is a Ben Brahim Mohammed work reproduced under the Creative Commons license

3.3.4.2 The Nomenclature of Enzymes

Because an enzyme is very specific for the substrate or for a family or related molecules of very similar structure, it is very unique. In the early days of metabolic studies, enzymes were named individually, one by one, with no concern for general rules of nomenclature. With time, the diversity of names and multitude of criteria to indentify newly discovered enzymes was such that the lack of a nomenclature that could be used worldwide was indering to the progress of enzymology (the scientific discipline devoted to study enzymes). The International Union of Pure and Applied Biochemistry (IUPAB), an international organization emerging from the joint efforts of many national societies of biochemists around the globe, appointed a working group to propose general rules that could be used to classify and identify enzymes. The result was a nomenclature based on the kind of reaction catalyzed by the enzyme consisting of:

- A name based on the contraction of "substrate + suffix ase" (e.g., urea + ase = urease, an enzyme that catalyzes a reaction with urea). This name has some flexibility.
- A rigid four-number code preceded by EC (for "Enzyme Commission"), which is unique for each enzyme (or sets of isoenzymes). The numbers refer to a family of enzymes and three successive subfamilies (Box 3.9). EC 5.2.1.3, for instance, identifies an enzyme from family 5 ("isomerases—catalyzes an isomerization reaction), first subfamily 2 ("cis-trans isomerization"), and the total code identifies specifically retinal isomerase, an enzyme involved in vision (see Sect. 2.2). Part of the whole tree of enzyme nomenclatures is presented in Box 3.9.

Box 3.9: Enzyme Classification and Nomenclature

IUBMB, the International Union of Biochemistry and Molecular Biology, is the organization responsible for recommendations on the nomenclature and classification of enzymes. Enzyme classification and strict nomenclature rules allow the unambiguous identification of enzymes. A working group, named Enzyme Commission, was established in 1956 to propose a universal classification and nomenclature system. Nearly 659 enzymes were known by then and the chaos in enzyme naming was clear. Nowadays more than 5500 enzymes are known, and it would be virtually impossible to communicate in enzymology if an official universal classification and nomenclature systems had not been established.

In 1961, the Enzyme Commission presented its report, in which enzymes are divided in six classes according to the reaction they catalyze. Classes and three levels of subclasses are numbered. The Enzyme Commission thus identified enzymes through a four-number code preceded by the letters EC to

(continued)

Box 3.9 (continued)

clearly identify that the numeric code corresponds to the classification set by the Enzyme Commission. The Enzyme Commission itself has been renamed but the classification system is still the same. The initials EC have remained although the commission they refer to has not.

Besides the numeric EC code, a name is also used because names are more intuitive and immediate than numeric codes. The most commonly used name for the enzyme is preferred, provided that it is unambiguous, but there are alternative systematic names that attempt to describe unambiguously the catalysis. Systematic names consist of two parts. The first contains the name of the substrate or, in the case of a bimolecular reaction, of the two substrates separated by a colon. The second part, ending in -ase, indicates the nature of the reaction, e.g., oxidoreductase, oxygenase, transferase (with a prefix indicating the nature of the group transferred), hydrolase, lyase, racemase, epimerase, isomerase, mutase, and ligase.

In practice, the enzyme classification and nomenclature stem from the classification of enzyme-catalyzed reactions, not from protein structures. A single protein may have two or more EC numbers if it catalyzes two or more reactions. This is the case, for example, for two proteins in *Escherichia coli*, each of which catalyzes the reactions both of aspartate kinase and of homoserine dehydrogenase. It may also happen that two or more proteins with no detectable evidence of homology catalyze the same reaction. For example, various different proteins catalyze the superoxide dismutase reaction and share a single EC number, EC1.15.1.1. This latter case is relatively rare, but it is almost universal that proteins catalyzing the same reaction in different organisms, or sets of isoenzymes in one organism, are homologous, with easily recognizable similarities in sequence.

Take Class EC 1 of enzymes, oxidoreductases, as example. This class contains the enzymes catalyzing oxidation reactions. Since the oxidation of one group must be accompanied by the reduction of another, they are grouped together as oxidoreductases. The systematic enzyme name is in the form "*donor*/*acceptor* oxidoreductase." The substrate that is being oxidized is regarded as being the hydrogen donor. The name is commonly "*donor* dehydrogenase." Although the term reductase is sometimes used as an alternative, it is important to remember that the recommended name does not define the equilibrium position of the reaction or the net direction of flux through the enzyme in vivo. The term "*donor* oxidase" is used only when O_2 is the acceptor.

Enzyme Classes

There are six classes of enzymes:

EC 1: Oxidoreductases catalyze reactions in which a substrate donates one or more electrons to an electron acceptor, becoming oxidized in the process.

(continued)

Box 3.9 (continued)

EC 2: Transferases catalyze reactions in which a chemical group is transferred from a donor substrate to an acceptor substrate.

EC 3: Hydrolases catalyze reactions in which a bond in a substrate is hydrolyzed to produce two fragments.

EC 4: Lyases catalyze non-hydrolytic reactions in which a chemical group is removed from a substrate leaving a double bond.

EC 5: Isomerases catalyze one-substrate one-product reactions that can be regarded as isomerization reactions.

EC 6: Ligases catalyze the joining together of two or more molecules coupled to hydrolysis of ATP or an analogous molecule. These enzymes are also sometimes called synthetases, a name that was already in use before creation of the original Enzyme Commission.

In reality all of the enzymes in classes 1–3 satisfy the definition of transferases. However, as these three classes are all large compared to the other three groups, it is convenient to break them into three classes and to reserve the name transferase for enzymes that are not oxidoreductases or hydrolases.

Enzyme Subclasses

Each of the six classes is divided into subclasses on the basis of the salient differences between the enzymes in the class. In **EC 1**, for example, the subclasses define the type of substrate acted on:

EC 1.1: Acting on the CH–OH group of donors
EC 1.2: Acting on the aldehyde or oxo group of donors
EC 1.19: Acting on reduced flavodoxin as donor
EC 1.97: Other oxidoreductases

This last subclass is numbered **EC 1.97** because it is provisional. In due course the enzymes it contains may be reclassified more appropriately. The original report had two subclasses **EC 1.99** and **EC 1.98** that were removed when sufficient information was available to place the enzymes they contained elsewhere.

Classes **EC 3–5** are divided into subclasses on the basis of types of substrate, in much the same way as in **EC 1**. In **EC 2**, however, it was more useful to emphasize the nature of the transferred group. So, for example, we have:

EC 2.1: Transferring one-carbon groups
EC 2.2: Transferring aldehyde or ketone residues
EC 2.3: Acyltransferases
EC 2.8: CoA-transferases

(continued)

Box 3.9 (continued)

In **EC 6** the division into subclasses is made on the basis of the type of product:

EC 6.1: Forming carbon–oxygen bonds
EC 6.2: Forming carbon–sulfur bonds
EC 6.3: Forming carbon–nitrogen bonds
EC 6.4: Forming carbon–carbon bonds
EC 6.5: Forming phosphoric ester bonds

Enzyme Sub-subclasses

The subclasses are divided into sub-subclasses in much the same way as the way the subclasses themselves are defined. For example, **EC 1.16** (oxidoreductases oxidizing metal ions) contains two sub-subclasses:

EC 1.16.1: With NAD^+ or $NADP^+$ as acceptor
EC 1.16.2: With oxygen as acceptor

As with the numbering of subclasses, 99 (or a smaller number if necessary) is used for sub-subclasses containing a miscellaneous group of enzymes. For example, subsection **EC 1.6** contains oxidoreductases acting on NADH or NADPH, and within this there is **EC 1.6.99** for miscellaneous acceptors.

There are also sub-subsubclasses so that each enzyme is identified by four different numbers. The division of sub-subclasses into sub-subsubclasses follows the same rationale as before. An exhaustive visit of the 4th level of classes is not justified here.

Final note: *Text based on "Enzyme Classification and Nomenclature" by S Boyce and K Tipton (Encyclopedia of Life Sciences, 2001) and "Current IUBMB recommendations on enzyme nomenclature and kinetics" by A, Cornish-Bowden (Perspectives in Science, 2014, 1, 74–87)*

Selected Bibliography

Boyce S, Tipton KF (2001) Enzyme classification and nomenclature, Encyclopedia of Life Sciences, 1–11. http://www.kois.sk/bioorg/bioorganicka_chemia/BIOORG2/SKLADIK/LITERATURA/A0000710-Enzyme%20classification%20and%20nomenctature.pdf

Cornish-Bowden A (2014) Current IUBMB recommendations on enzyme nomenclature and kinetics. Perspect Sci 1:74–87

Dobson CM, Gerard JA, Pratt, AJ (2001) Foundations of Chemical Biology, Oxford Chemistry Primers. Oxford University Press

IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (1985) Nomenclature and symbolism for amino acids and peptides. J Biol Chem 260:14–42

Stevenson J, Brown AJ (2009) How essential is cholesterol? Biochem J 420:e1–e4

Westheimer FH (1987) Why nature chose phosphates. Science 235:1173–1178