Overview

A. Classification ●

The **lipids** are a large and heterogeneous group of substances of biological origin that are easily dissolved in organic solvents such as methanol, acetone, chloroform, and benzene. By contrast, they are either insoluble or only poorly soluble in water. Their low water solubility is due to a lack of polarizing atoms such as O, N, S, and P (see p. 6).

Lipids can be classified into substances that are either *hydrolyzable*— i. e., able to undergo hydrolytic cleavage—or *nonhydrolyzable*. Only a few examples of the many lipids known can be mentioned here. The individual classes of lipids are discussed in more detail in the following pages.

Hydrolyzable lipids (components shown in brackets). The simple **esters** include the *fats* (triacylglycerol; one glycerol + three acyl residues); the *waxes* (one fatty alcohol + one acyl residue); and the *sterol esters* (one sterol + one acyl residue). The **phospholipids** are esters with more complex structures. Their characteristic component is a phosphate residue. The phospholipids include the *phosphatidic acids* (one glycerol + two acyl residues + one phosphate) and the *phosphatides* (one glycerol + two acyl residues + one phosphate + one amino alcohol). In the **sphingolipids**, glycerol and one acyl residue are replaced by sphingosine. Particularly important in this group are the sugar-containing **glycolipids** (one sphingosine + one fatty acid + sugar). The *cerebrosides* (one sphingosine + one fatty acid + one sugar) and *gangliosides* (one sphingosine + one fatty acid + several different sugars, including neuraminic acid) are representatives of this group.

The components of the hydrolyzable lipids are linked to one another by **ester bonds**. They are easily broken down either enzymatically or chemically.

Non-hydrolyzable lipids. The **hydrocarbons** include the *alkanes* and *carotenoids*. The **lipid alcohols** are also not hydrolyzable. They include long-chained *alkanols* and cyclic *sterols* such as cholesterol, and *steroids* such as estradiol and testosterone. The most important **acids** among the lipids are *fatty acids*. The *eicosanoids* also belong to this group; these are derivatives of the polyunsaturated fatty acid arachidonic acid (see p. 390).

B. Biological roles ●

1. Fuel. Lipids are an important source of energy in the diet. In quantitative terms, they represent the principal energy reserve in animals. Neutral fats in particular are stored in specialized cells, known as *adipocytes*. Fatty acids are released from these again as needed, and these are then oxidized in the mitochondria to form water and carbon dioxide, with oxygen being consumed. This process also gives rise to reduced coenzymes, which are used for ATP production in the respiratory chain (see p. 140).

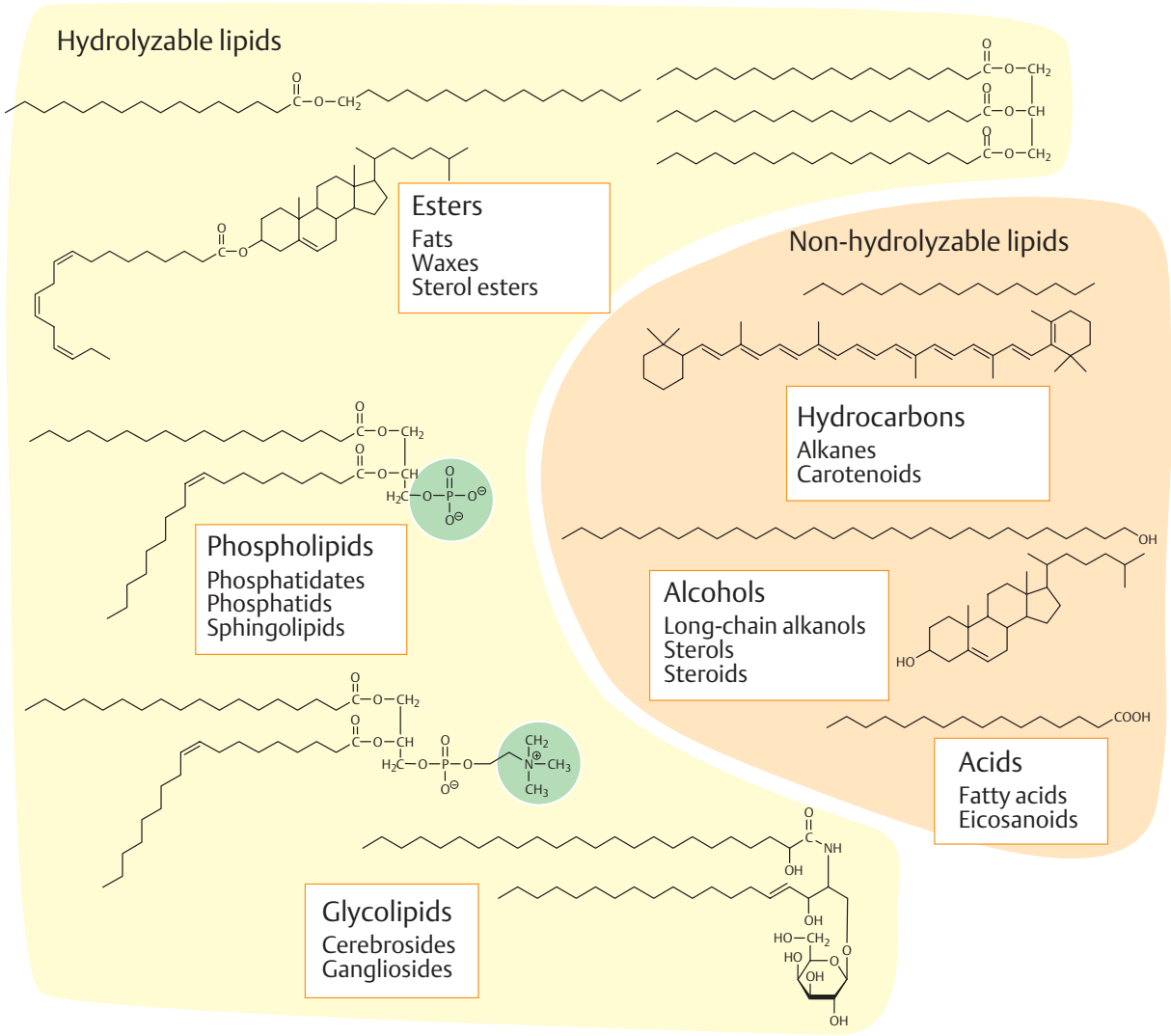
2. Nutrients. Amphipathic lipids are used by cells to build membranes (see p. 214). Typical membrane lipids include phospholipids, glycolipids, and cholesterol. Fats are only weakly amphiphilic and are therefore not suitable as membrane components.

3. Insulation. Lipids are excellent insulators. In the higher animals, neutral fats are found in the subcutaneous tissue and around various organs, where they serve as mechanical and thermal insulators. As the principal constituent of cell membranes, lipids also insulate cells from their environment mechanically and electrically. The impermeability of lipid membranes to ions allows the formation of the membrane potential (see p. 126).

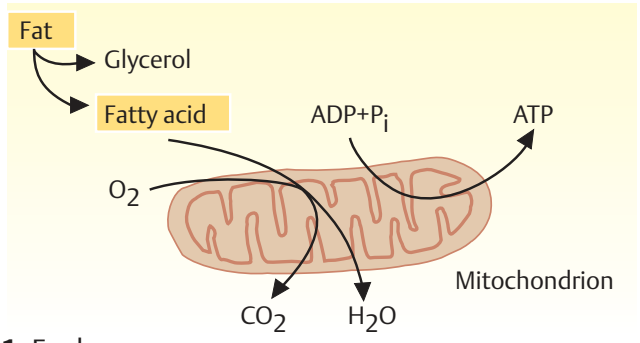
4. Special tasks. Some lipids have adopted special roles in the body. Steroids, eicosanoids, and some metabolites of phospholipids have *signaling functions*. They serve as hormones, mediators, and second messengers (see p. 370). Other lipids form *anchors* to attach proteins to membranes (see p. 214). The lipids also produce *cofactors for enzymatic reactions*—e. g., vitamin K (see p. 52) and ubiquinone (see p. 104). The carotenoid retinal, a light-sensitive lipid, is of central importance in the *process of vision* (see p. 358).

Several lipids are not formed independently in the human body. These substances, as **essential fatty acids** and **fat-soluble vitamins**, are indispensable components of nutrition (see pp. 364 ff.)

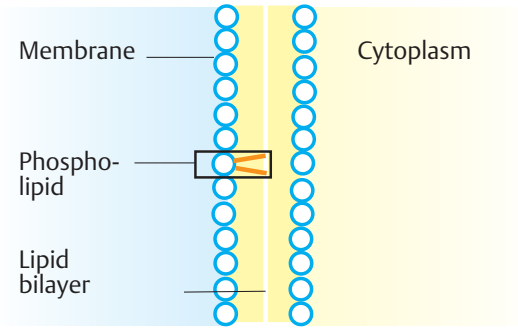
A. Classification



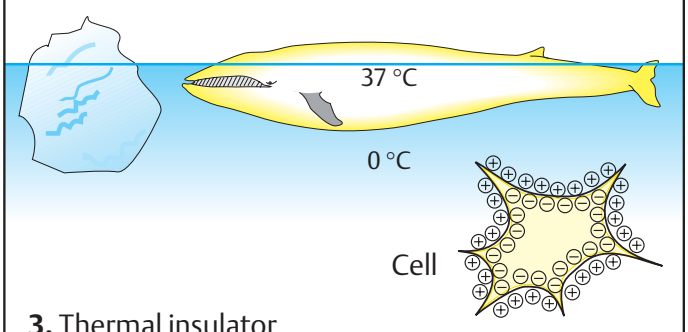
B. Biological roles



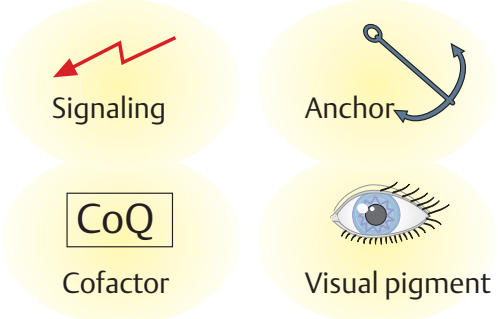
1. Fuel



2. Building block



3. Thermal insulator



4. Special tasks

Fatty acids and fats

A. Carboxylic acids ◑

The naturally occurring **fatty acids** are carboxylic acids with unbranched hydrocarbon chains of 4–24 carbon atoms. They are present in all organisms as components of fats and membrane lipids. In these compounds, they are esterified with alcohols (glycerol, sphingosine, or cholesterol). However, fatty acids are also found in small amounts in unesterified form. In this case, they are known as *free fatty acids* (FFAs). As free fatty acids have strongly amphipathic properties (see p. 28), they are usually present in protein-bound forms.

The table lists the full series of aliphatic carboxylic acids that are found in plants and animals. In higher plants and animals, unbranched, longchain fatty acids with either 16 or 18 carbon atoms are the most common—e. g., palmitic and stearic acid. The number of carbon atoms in the longer, natural fatty acids is always even. This is because they are biosynthesized from C_2 building blocks (see p. 168).

Some fatty acids contain one or more isolated *double bonds*, and are therefore *"unsaturated."* Common **unsaturated fatty acids** include oleic acid and linoleic acid. Of the two possible *cis–trans* isomers (see p. 8), usually only the *cis* forms are found in natural lipids. Branched fatty acids only occur in bacteria. A shorthand notation with several numbers is used for precise characterization of the structure of fatty acids—e g., 18:2;9,12 for linoleic acid. The first figure stands for the number of C atoms, while the second gives the number of double bonds. The positions of the double bonds follow after the semicolon. As usual, numbering starts at the carbon with the highest oxidation state (i. e., the carboxyl group corresponds to C-1). Greek letters are also commonly used (α = C-2; β = C-3; ω = the last carbon, ω-3 = the third last carbon).

Essential fatty acids are fatty acids that have to be supplied in the diet. Without exception, these are all polyunsaturated fatty acids: the C_{20} fatty acid *arachidonic acid* (20:4;5,8,11,14) and the two C_{18} acids *linoleic acid* (18:2;9,12) and *linolenic acid* (18:3;9,12,15). The animal organism requires arachidonic acid to synthesize eicosanoids (see p. 390). As the organism is capable of elongating fatty acids by adding C_2 units, but is not able to introduce double bonds into the end sections of fatty acids (after C-9), arachidonic acid has to be supplied with the diet. Linoleic and linolenic acid can be converted into arachidonic acid by elongation, and they can therefore replace arachidonic acid in the diet.

B. Structure of fats ◑

Fats are esters of the trivalent alcohol *glycerol* with three fatty acids. When a single fatty acid is esterified with glycerol, the product is referred to as a *monoacylglycerol* (fatty acid residue = acyl residue).

Formally, esterification with additional fatty acids leads to *diacylglycerol* and ultimately to *triacylglycerol,* the actual fat (formerly termed "triglyceride"). As triacylglycerols are uncharged, they are also referred to as *neutral fats*. The carbon atoms of glycerol are not usually equivalent in fats. They are distinguished by their *"sn"* number, where sn stands for "stereospecific numbering."

The three acyl residues of a fat molecule may differ in terms of their chain length and the number of double bonds they contain. This results in a large number of possible combinations of individual fat molecules. When extracted from biological materials, fats always represent mixtures of very similar compounds, which differ in their fatty acid residues. A chiral center can arise at the middle C atom (*sn* -C-2) of a triacylglycerol if the two external fatty acids are different. The monoacylglycerols and diacylglycerols shown here are also chiral compounds. Nutritional fats contain palmitic, stearic, oleic acid, and linoleic acid particularly often. Unsaturated fatty acids are usually found at the central C atom of glycerol.

The length of the fatty acid residues and the number of their double bonds affect the melting point of the fats. The shorter the fatty acid residues and the more double bonds they contain, the lower their melting points.

A. Carboxylic acids

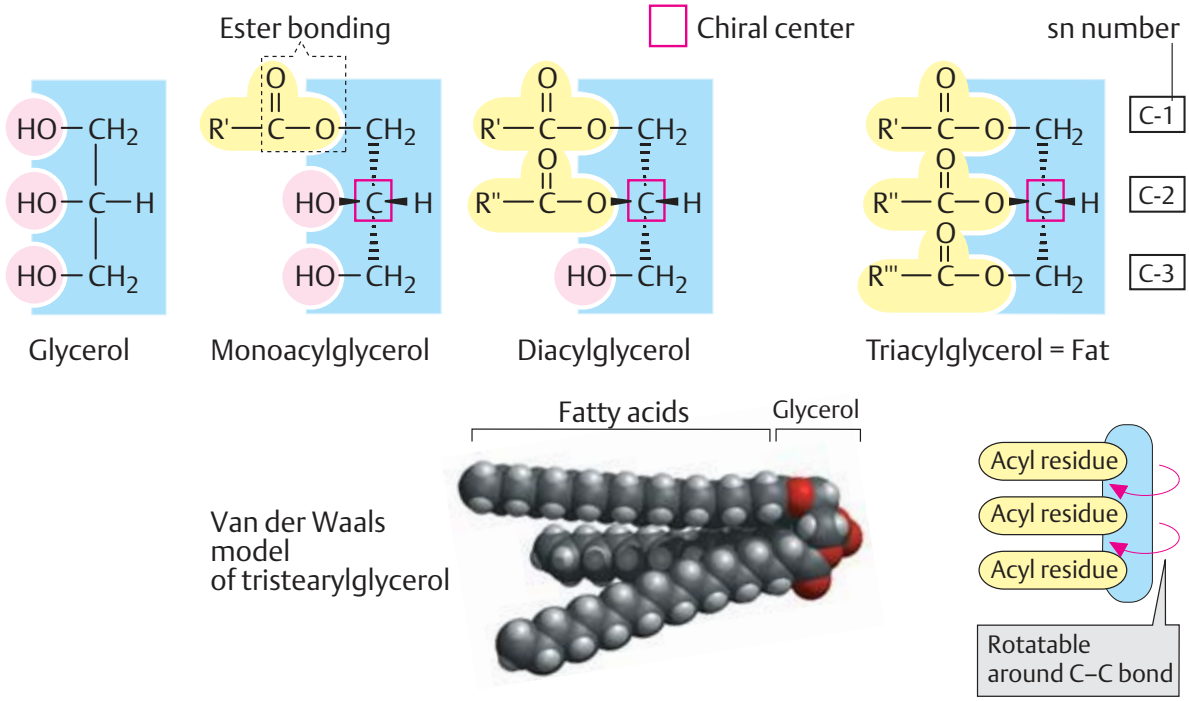
Name	Number of carbons : Number of double bonds; Position of double bonds	
Formic acid	1 : 0	Not contained in lipids
Acetic acid	2 : 0	
Propionic acid	3 : 0	
Butyric acid	4 : 0	
Valerianic acid	5 : 0	
Caproic acid	6 : 0	
Caprylic acid	8 : 0	
Capric acid	10 : 0	
Lauric acid	12 : 0	
Myristic acid	14 : 0	
Palmitic acid	16 : 0	
Stearic acid	18 : 0	
Oleic acid	18 : 1; 9	
Linoleic acid	18 : 2; 9,12	
Linolenic acid	18 : 3; 9,12,15	
Arachidic acid	20 : 0	
Arachidonic acid	20 : 4; 5,8,11,14	
Behenic acid	22 : 0	
Erucic acid	22 : 1; 13	
Lignoceric acid	24 : 0	
Nervonic acid	24 : 1; 15	

$HOOC-CH_2-CH_2-CH_2-CH_2-CH_3$

Caproic acid

Essential in human nutrition

B. Structure of fats



Phospholipids and glycolipids

A. Structure of phospholipids and glycolipids ◑

Fats (triacylglycerol, **1**) are esters of glycerol with three fatty acids (see p. 48). Within the cell, they mainly occur as fat droplets. In the blood, they are transported in the hydrophobic interior of lipoproteins (see p. 278).

Phospholipids (**2**) are the main constituents of biological membranes (see pp. 214–217). Their common feature is a phosphate residue that is esterified with the hydroxyl group at C-3 of glycerol. Due to this residue, phospholipids have at least one negative charge at a neutral pH.

Phosphatidates (anions of the phosphatidic acids), the simplest phospholipids, are phosphate esters of diacylglycerol. They are important intermediates in the biosynthesis of fats and phospholipids (see p. 170). Phosphatidates can also be released from phospholipids by phospholipases.

The other phospholipids can be derived from phosphatidates (residue = phosphatidyl). Their phosphate residues are esterified with the hydroxyl group of an amino alcohol (*choline*, *ethanolamine*, or *serine*) or with the cyclohexane derivative *myo-inositol*. *Phosphatidylcholine* is shown here as an example of this type of compound. When two phosphatidyl residues are linked with one glycerol, the result is *cardiolipin* (not shown), a phospholipid that is characteristic of the inner mitochondrial membrane. *Lysophospholipids* arise from phospholipids by enzymatic cleavage of an acyl residue. The hemolytic effect of bee and snake venoms is due in part to this reaction.

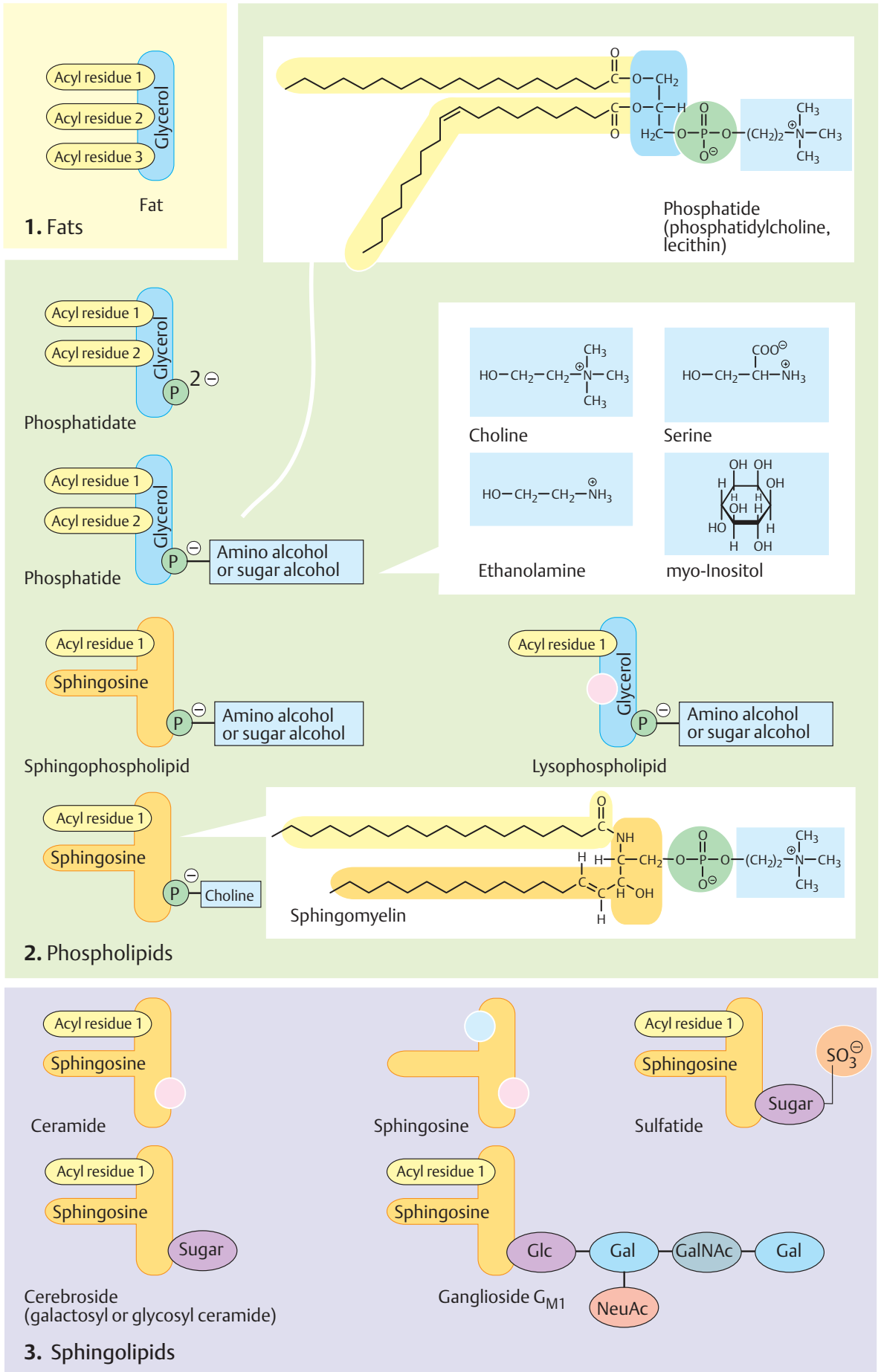
Phosphatidylcholine (lecithin) is the most abundant phospholipid in membranes. *Phosphatidylethanolamine* (cephalin) has an ethanolamine residue instead of choline, and *phosphatidylserine* has a serine residue. In *phosphatidylinositol,* phosphatidate is esterified with the sugarlike cyclic polyalcohol *myo*-inositol. A doubly phosphorylated derivative of this phospholipid, phosphatidylinositol 4,5-bisphosphate, is a special component of membranes, which, by enzymatic cleavage, can give rise to two *second messengers*, diacylglycerol (DAG) and inositol 1,4,5trisphosphate ($InsP_3$; see p. 386).

Some phospholipids carry additional charges, in addition to the negative charge at the phosphate residue. In phosphatidylcholine and phosphatidylethanolamine, the *N*-atom of the amino alcohol is positively charged. As a whole, these two phosphatides therefore appear to be neutral. In contrast, phosphatidylserine—with one additional positive charge and one additional negative charge in the serine residue—and phosphatidylinositol (with no additional charge) have a negative net charge, due to the phosphate residue.

Sphingolipids (**3**), which are found in large quantities in the membranes of nerve cells in the brain and in neural tissues, have a slightly different structure from the other membrane lipids discussed so far. In sphingolipids, *sphingosine*, an amino alcohol with an unsaturated alkyl side chain, replaces glycerol and one of the acyl residues. When sphingosine forms an amide bond to a fatty acid, the compound is called *ceramide* (**3**). This is the precursor of the sphingolipids. *Sphingomyelin* (2)—the most important sphingolipid—has an additional phosphate residue with a choline group attached to it on the sphingosine, in addition to the fatty acid.

Glycolipids (**3**) are present in all tissues on the outer surface of the plasma membrane. They consist of sphingosine, a fatty acid, and an oligosaccharide residue, which can sometimes be quite large. The phosphate residue typical of phospholipids is absent. *Galactosylceramide* and *glucosylceramide* (known as cerebroside) are simple representatives of this group. Cerebrosides in which the sugar is esterified with sulfuric acid are known as *sulfatides*. *Gangliosides* are the most complex glycolipids. They constitute a large family of membrane lipids with receptor functions that are as yet largely unknown. A characteristic component of many gangliosides is *N*-acetylneuraminic acid (sialic acid; see p. 38).

A. Structure of fats, phospholipids, and glycolipids



Isoprenoids

A. Activated acetic acid as a component of lipids ◑

Although the lipids found in plant and animal organisms occur in many different forms, they are all closely related biogenetically; they are all derived from **acetyl-CoA**, the "activated acetic acid" (see pp. 12, 110).

1. One major pathway leads from acetyl-CoA to the activated fatty acids (**acyl-CoA**; for details, see p. 168). *Fats, phospholipids,* and *glycolipids* are synthesized from these, and fatty acid derivatives in particular are formed. Quantitatively, this is the most important pathway in animals and most plants.

2. The second pathway leads from acetyl-CoA to isopentenyl diphosphate (*"active isoprene"*), the basic component for the **isoprenoids**. Its biosynthesis is discussed in connection with biosynthesis of the isoprenoid, cholesterol (see p. 172).

B. Isoprenoids ◑

Formally, isoprenoids are derived from a single common building block, isoprene (2-methyl-1,3-butadiene), a methyl-branched compound with five C atoms. Activated isoprene, *isopentenyl diphosphate,* is used by plants and animals to biosynthesize linear and cyclic oligomers and polymers. For the isoprenoids listed here—which only represent a small selection—the number of isoprene units (I) is shown.

From activated isoprene, the metabolic pathway leads via dimerization to activated *geraniol* (I = 2) and then to activated *farnesol* (I = 3). At this point, the pathway divides into two. Further extension of farnesol leads to chains with increasing numbers of isoprene units—e. g., *phytol* (I = 4), *dolichol* (I = 14–24), and *rubber* (I = 700–5000). The other pathway involves a "head-to-head" linkage between two farnesol residues, giving rise to *squalene* (I = 6), which, in turn, is converted to *cholesterol* (I = 6) and the other *steroids*.

The ability to synthesize particular isoprenoids is limited to a few species of plants and animals. For example, rubber is only formed by a few plant species, including the rubber tree (*Hevea brasiliensis*). Several isoprenoids that are required by animals for metabolism, but cannot be produced by them independently, are vitamins; this group includes *vitamins A, D, E,* and *K*. Due to its structure and function, vitamin D is now usually classified as a steroid hormone (see pp. 56, 330).

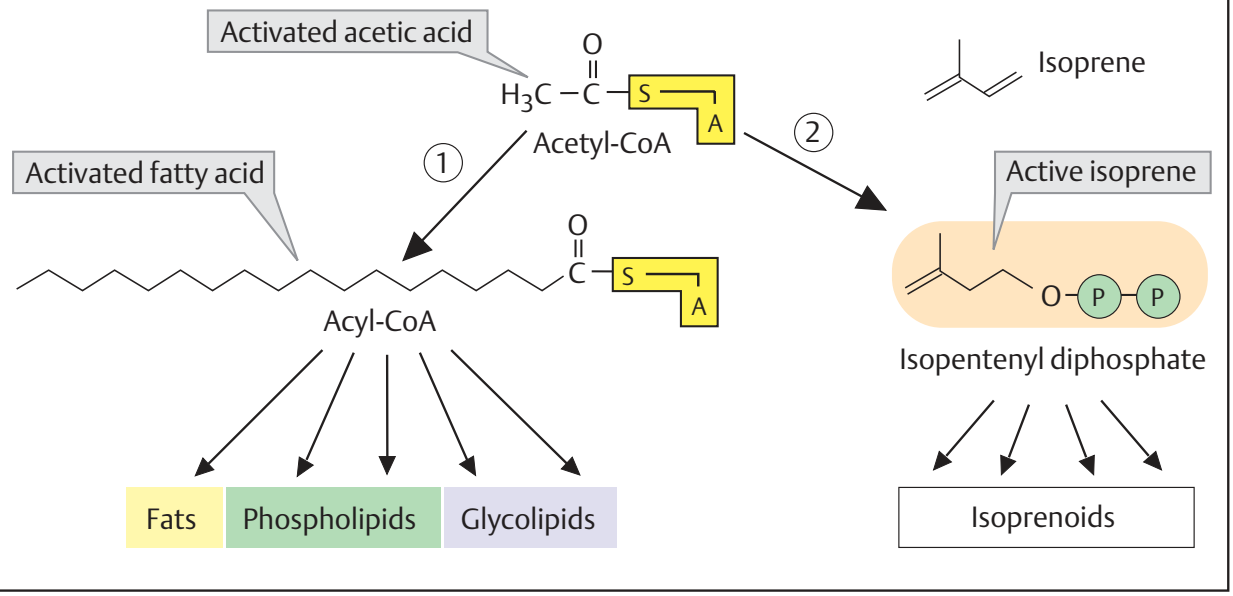
Isoprene metabolism in plants is very complex. Plants can synthesize many types of aromatic substances and volatile oils from isoprenoids. Examples include *menthol* (I= 2), *camphor* (I = 2), and *citronellal* (I = 2). These C_{10} compounds are also called *monoterpenes*. Similarly, compounds consisting of three isoprene units (I = 3) are termed *sesquiterpenes*, and the steroids (I = 6) are called *triterpenes*.

Isoprenoids that have hormonal and signaling functions form an important group. These include *steroid hormones* (I = 6) and *retinoate* (the anion of retinoic acid; I = 3) in vertebrates, and *juvenile hormone* (I = 3) in arthropods. Some plant hormones also belong to the isoprenoids—e. g., the cytokinins, abscisic acid, and brassinosteroids.

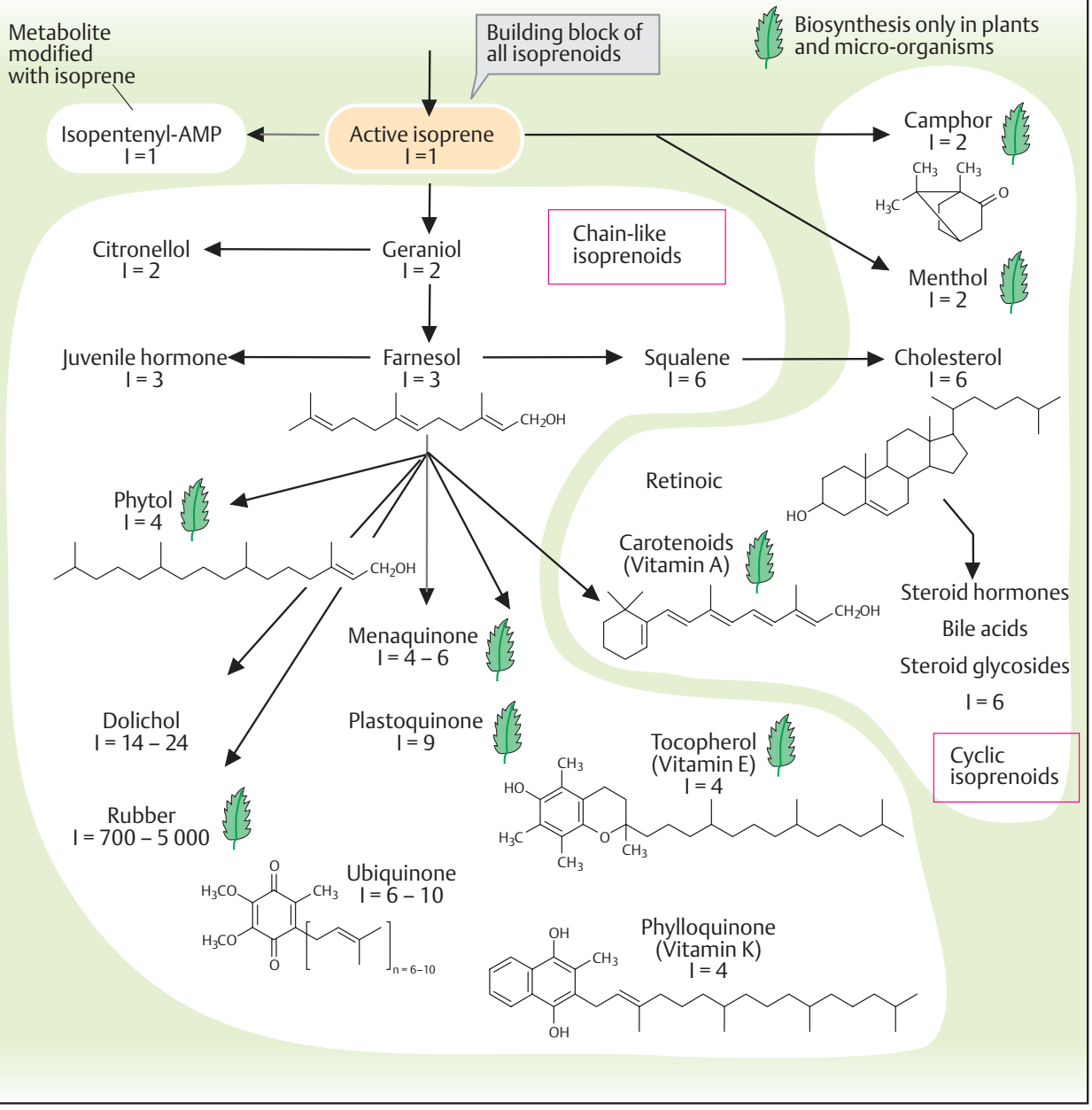
Isoprene chains are sometimes used as lipid anchors to fix molecules to membranes (see p. 214). Chlorophyll has a *phytyl* residue (I = 4) as a lipid anchor. Coenzymes with isoprenoid anchors of various lengths include *ubiquinone* (coenzyme Q; I = 6–10), *plastoquinone* (I = 9), and *menaquinone* (vitamin K; I = 4–6). Proteins can also be anchored to membranes by *isoprenylation*.

In some cases, an isoprene residue is used as an element to modify molecules chemically. One example of this is *N'*-isopentenyl-AMP, which occurs as a modified component in tRNA.

A. Activated acetic acid as a component of lipids



B. Isoprenoids



Steroid structure

A. Steroid building blocks ◑

Common to all of the steroids is a molecular core structure consisting of four saturated rings, known as *gonane*. At the end of the steroid core, many steroids also carry a side chain, as seen in *cholestane*, the basic component of the *sterols* (steroid alcohols).

B. Spatial structure ○

The four rings of the steroids are distinguished using the letters A, B, C, and D. Due to the tetrahedral arrangement of the single carbon bonds, the rings are not flat, but puckered. Various *ring conformations* are known by the terms "chair," "boat," and "twisted" (not shown). The *chair* and *boat* conformations are common. Fivemembered rings frequently adopt a conformation referred to as an "envelope". Some rings can be converted from one conformation to another at room temperature, but with steroids this is dif cult.

Substituents of the steroid core lie either approximately in the same plane as the ring (e = *equatorial*) or nearly perpendicular to it (a = *axial*). In threedimensional representations, substituents pointing toward the observer are indicated by an unbroken line (β position), while bonds pointing into the plane of the page are indicated by a dashed line (α position). The so-called *angular* methyl groups at C-10 and C-13 of the steroids always adopt the β position.

Neighboring rings can lie in the same plane (*trans*; **2**) or at an angle to one another (*cis*; **1**). This depends on the positions of the substituents of the shared ring carbons, which can be arranged either *cis* or *trans* to the angular methyl group at C-10. The substituents of steroid that lie at the points of intersection of the individual rings are usually in *trans* position. As a whole, the core of most steroids is more or less planar, and looks like a flat disk. The only exceptions to this are the ecdysteroids, bile acids (in which A:B is *cis*), cardiac glycosides, and toad toxins.

A more realistic impression of the threedimensional structure of steroids is provided by the space-filling model of *cholesterol* (**3**). The four rings form a fairly rigid scaffolding, onto which the much more mobile side chain is attached.

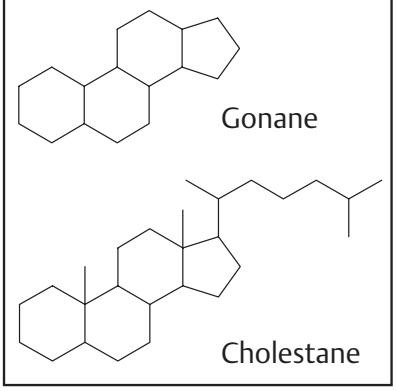
Steroids are relatively apolar (hydrophobic). Some polar groups—e. g., hydroxyl and oxo groups—give them amphipathic properties. This characteristic is especially pronounced with the bile acids (see p. 314).

C. Thin-layer chromatography ○

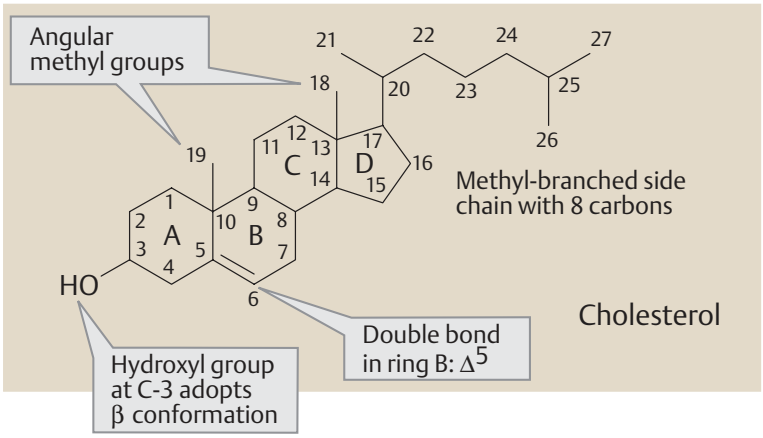
Thin-layer chromatography (TLC) is a powerful, mainly analytic, technique for rapidly separating lipids and other small molecules such as amino acids, nucleotides, vitamins, and drugs. The *sample* being analyzed is applied to a *plate* made of glass, aluminum, or plastic, which is covered with a thin layer of silica gel or other material (**1**). The plate is then placed in a chromatography chamber that contains some *solvent*. Drawn by capillary forces, the solvent moves up the plate (**2**). The substances in the sample move with the solvent. The speed at which they move is determined by their distribution between the *stationary phase* (the hydrophilic silica), and the *mobile phase* (the hydrophobic solvent). When the solvent reaches the top edge of the plate, the chromatography is stopped. After evaporation of the solvent, the separated substances can be made visible using appropriate staining methods or with physical processes (e. g., ultraviolet light) (**3**). The movement of a substance in a given TLC system is expressed as its R_f value. In this way, compounds that are not known can be identified by comparison with reference substances.

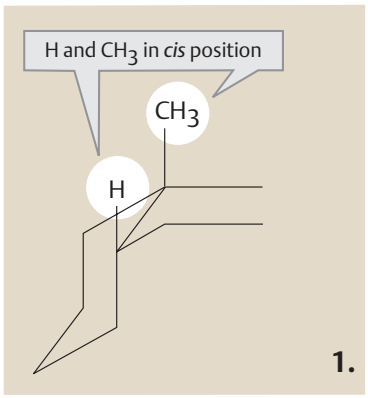
A process in which the polarity of the stationary and mobile phases is reversed—i. e., the stationary phase is apolar and the solvent is polar—is known as "reversed-phase thin-layer chromatography" (RP-TLC).

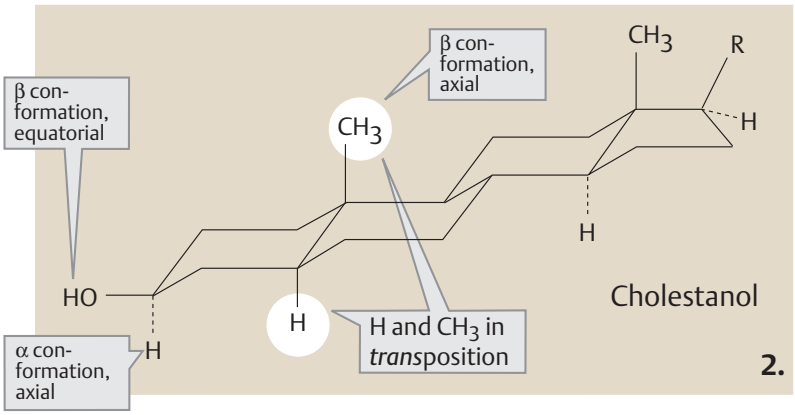
A. Steroid building blocks

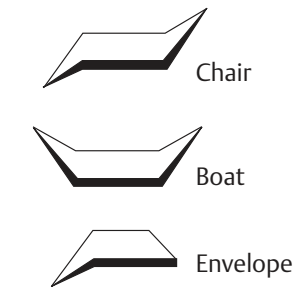


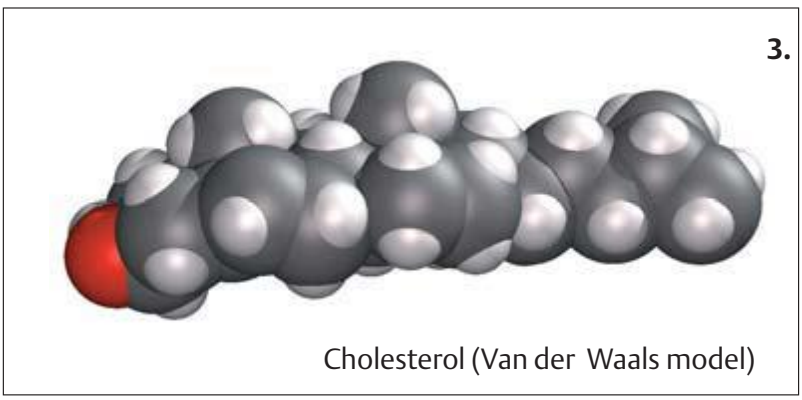
B. 3D structure



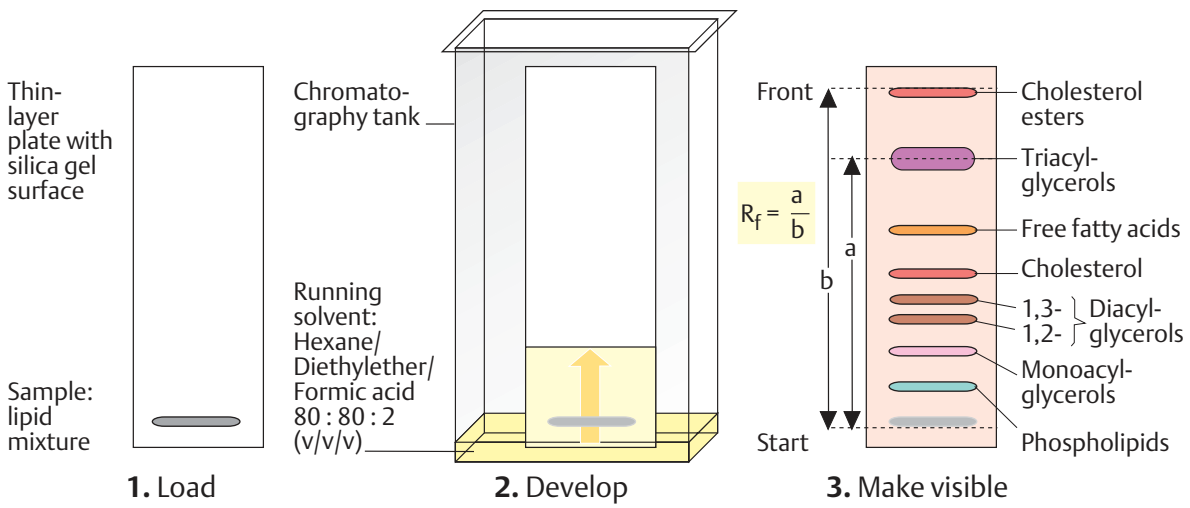








C. Thin-layer chromatography



Steroids: overview

The three most important groups of steroids are the *sterols*, *bile acids*, and *steroid hormones*. Particularly in plants, compounds with steroid structures are also found that are notable for their pharmacological effects—steroid alkaloids, digitalis glycosides, and saponins.

A. Sterols ◑

Sterols are *steroid alcohols*. They have a β-positioned hydroxyl group at C-3 and one or more double bonds in ring B and in the side chain. There are no further oxygen functions, as in the carbonyl and carboxyl groups.

The most important sterol in animals is **cholesterol**. Plants and microorganisms have a wide variety of closely related sterols instead of cholesterol—e.g., **ergosterol**, β-**sitosterol**, and **stigmasterol**.

Cholesterol is present in all animal tissues, and particularly in neural tissue. It is a major constituent of cellular membranes, in which it regulates fluidity (see p. 216). The storage and transport forms of cholesterol are its esters with fatty acids. In lipoproteins, cholesterol and its fatty acid esters are associated with other lipids (see p. 278). Cholesterol is a constituent of the bile and is therefore found in many gallstones. Its biosynthesis, metabolism, and transport are discussed elsewhere (see pp. 172, 312).

Cholesterol-rich lipoproteins of the LDL type are particularly important in the development of arteriosclerosis, in which the arterial walls are altered in connection with an excess plasma cholesterol level. In terms of dietary physiology, it is important that plant foodstuffs are low in cholesterol. By contrast, animal foods can contain large amounts of cholesterol—particularly butter, egg yolk, meat, liver, and brain.

B. Bile acids ◑

Bile acids are synthesized from cholesterol in the liver (see p. 314). Their structures can therefore be derived from that of cholesterol. Characteristic for the bile acids is a side chain shortened by three C atoms in which the last carbon atom is oxidized to a carboxyl group. The double bond in ring B is reduced and rings A and B are in *cis* position relative to each other (see p. 54). One to three hydroxyl groups (in α position) are found in the steroid core at positions 3, 7, and 12. Bile acids keep bile cholesterol in a soluble state as micelles and promote the digestion of lipids in the intestine (see p. 270). **Cholic acid** and **chenodeoxycholic acid** are *primary bile acids* that are formed by the liver. Their dehydroxylation at C-7 by microorganisms from the intestinal flora gives rise to the *secondary bile acids* **lithocholic acid** and **deoxycholic acid**.

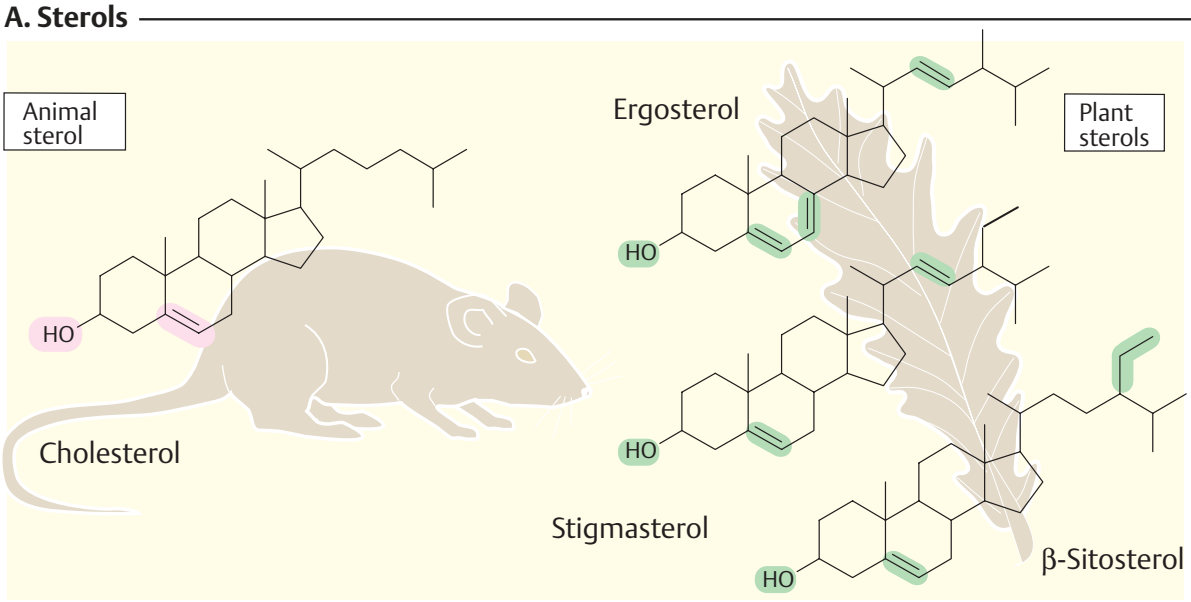
C. Steroid hormones ◑

The conversion of cholesterol to *steroid hormones* (see p. 376) is of minor importance quantitatively, but of major importance in terms of physiology. The steroid hormones are a group of lipophilic signal substances that regulate metabolism, growth, and reproduction (see p. 374).

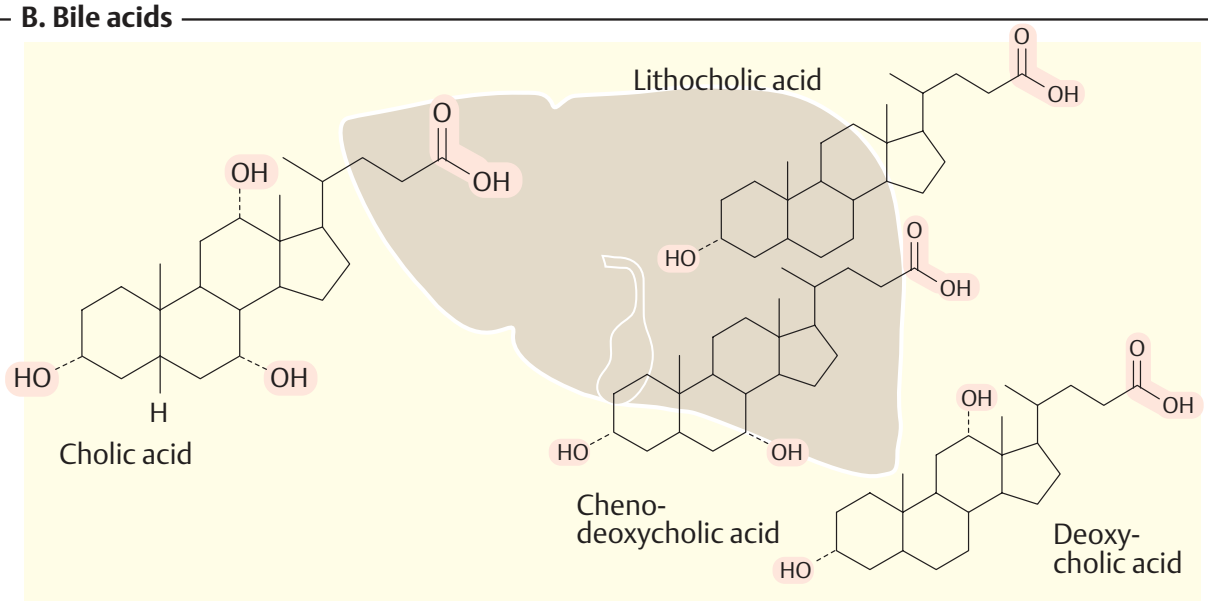
Humans have six steroid hormones: **progesterone**, **cortisol**, **aldosterone**, **testosterone**, **estradiol**, and **calcitriol**. With the exception of calcitriol, these steroids have either no side chain or only a short side one consisting of two carbons. Characteristic for most of them is an oxo group at C-3, conjugated with a double bond between C-4 and C-5 of ring A. Differences occur in rings C and D. Estradiol is aromatic in ring A, and its hydroxyl group at C-3 is therefore phenolic. Calcitriol differs from other vertebrate steroid hormones; it still contains the complete carbon framework of cholesterol, but lightdependent opening of ring B turns it into what is termed a "secosteroid" (a steroid with an open ring).

Ecdysone is the steroid hormone of the arthropods. It can be regarded as an early form of the steroid hormones. Steroid hormones with signaling functions also occur in plants.

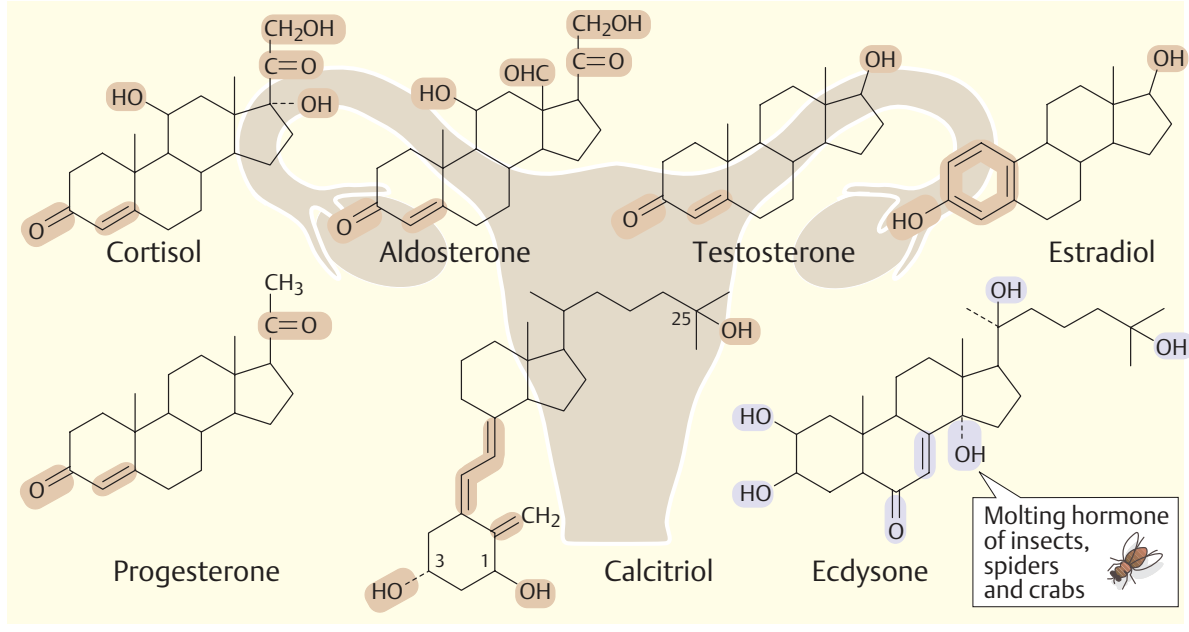
A. Sterols



B. Bile acids



C. Steroid hormones



Amino acids: chemistry and properties

A. Amino acids: functions ●

The amino acids (2-aminocarboxylic acids) fulfill various functions in the organism. Above all, they serve as the **components of peptides and proteins**. Only the 20 *proteinogenic amino acids* (see p. 60) are included in the genetic code and therefore regularly found in proteins. Some of these amino acids undergo further (post-translational) change following their incorporation into proteins (see p. 62). Amino acids or their derivatives are also form components of **lipids**—e. g., serine in phospholipids and glycine in bile salts. Several amino acids function as **neurotransmitters** themselves (see p. 352), while others are precursors of neurotransmitters, mediators, or hormones (see p. 380). Amino acids are important (and sometimes essential) components of food (see p. 360). Specific amino acids form **precursors** for other metabolites—e. g., for glucose in gluconeogenesis, for purine and pyrimidine bases, for heme, and for other molecules. Several nonproteinogenic amino acids function as intermediates in the synthesis and breakdown of proteinogenic amino acids (see p. 412) and in the urea cycle (see p. 182).

B. Optical activity ◑

The natural amino acids are mainly α-amino acids, in contrast to β-amino acids such as β-alanine and taurine. Most α-amino acids have four different substituents at C-2 (Cα). The α atom therefore represents a *chiral center*—i. e., there are two different **enantiomers** (L- and D-amino acids; see p. 8). Among the proteinogenic amino acids, only glycine is *not* chiral (R = H). In nature, it is almost exclusively **L-amino acids** that are found. **D**-Amino acids occur in bacteria—e. g., in murein (see p. 40)—and in peptide antibiotics. In animal metabolism, D-Amino acids would disturb the enzymatic reactions of L-amino acids and they are therefore broken down in the liver by the enzyme *D-amino acid oxidase*.

The **Fischer projection** (center) is used to present the formulas for chiral centers in biomolecules. It is derived from their three-dimensional structure as follows: firstly, the tetrahedron is rotated in such a way that the most oxidized group (the carboxylate group) is at the top. Rotation is then continued until the line connecting line COO^- and R (red) is level with the page. In L-amino acids, the NH_3^+ group is then on the left, while in **D**-amino acids it is on the right.

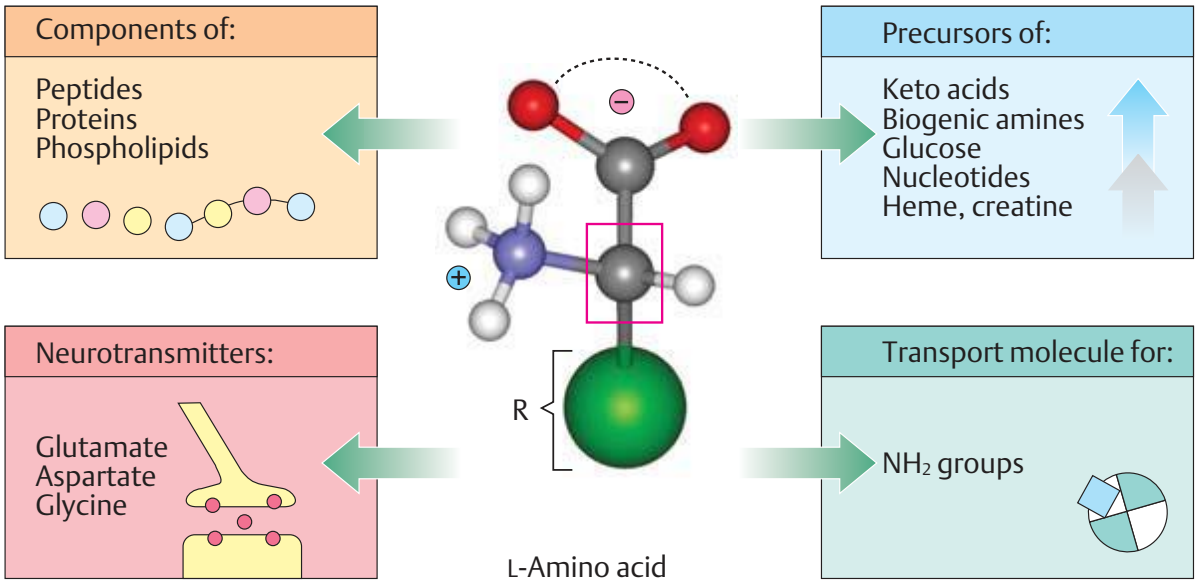
C. Dissociation curve of histidine ◑

All amino acids have at least two ionizable groups, and their net charge therefore depends on the pH value. The COOH groups at the α-C atom have pK_a values of between 1.8 and 2.8 and are therefore more acidic than simple monocarboxylic acids. The basicity of the α-amino function also varies, with pK_a values of between 8.8 and 10.6, depending on the amino acid. Acidic and basic amino acids have additional ionizable groups in their side chain. The pK_a values of these side chains are listed on p. 60. The electrical charges of peptides and proteins are mainly determined by groups in the side chains, as most α-carboxyl and α-amino functions are linked to peptide bonds (see p. 66).

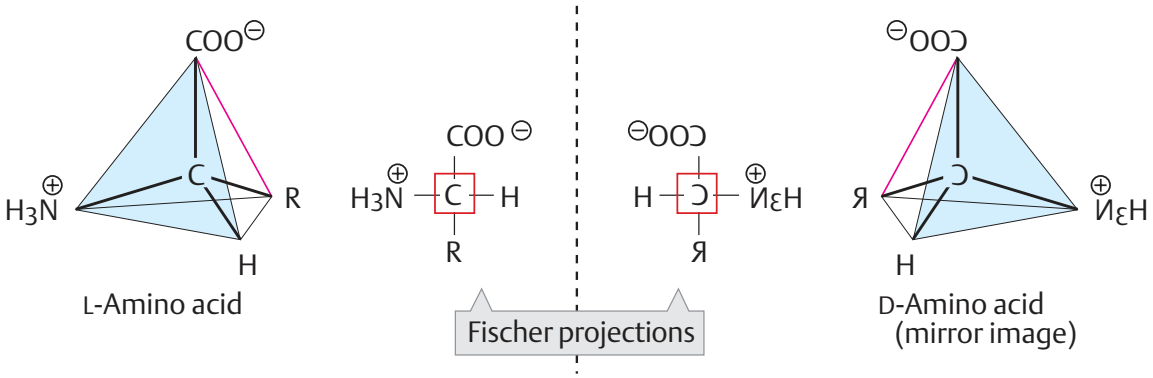
Histidine can be used here as an example of the pH-dependence of the net charge of an amino acid. In addition to the carboxyl group and the amino group at the α-C atom with pK_a values of 1.8 and 9.2, respectively, histidine also has an imidazole residue in its side chain with a pK_a value of 6.0. As the pH increases, the net charge (the sum of the positive and negative charges) therefore changes from +2 to –1. At pH 7.6, the net charge is zero, even though the molecule contains two almost completely ionized groups in these conditions. This pH value is called the **isoelectric point**.

At its isoelectric point, histidine is said to be **zwitterionic**, as it has both anionic *and* cationic properties. Most other amino acids are also zwitterionic at neutral pH. Peptides and proteins also have isoelectric points, which can vary widely depending on the composition of the amino acids.

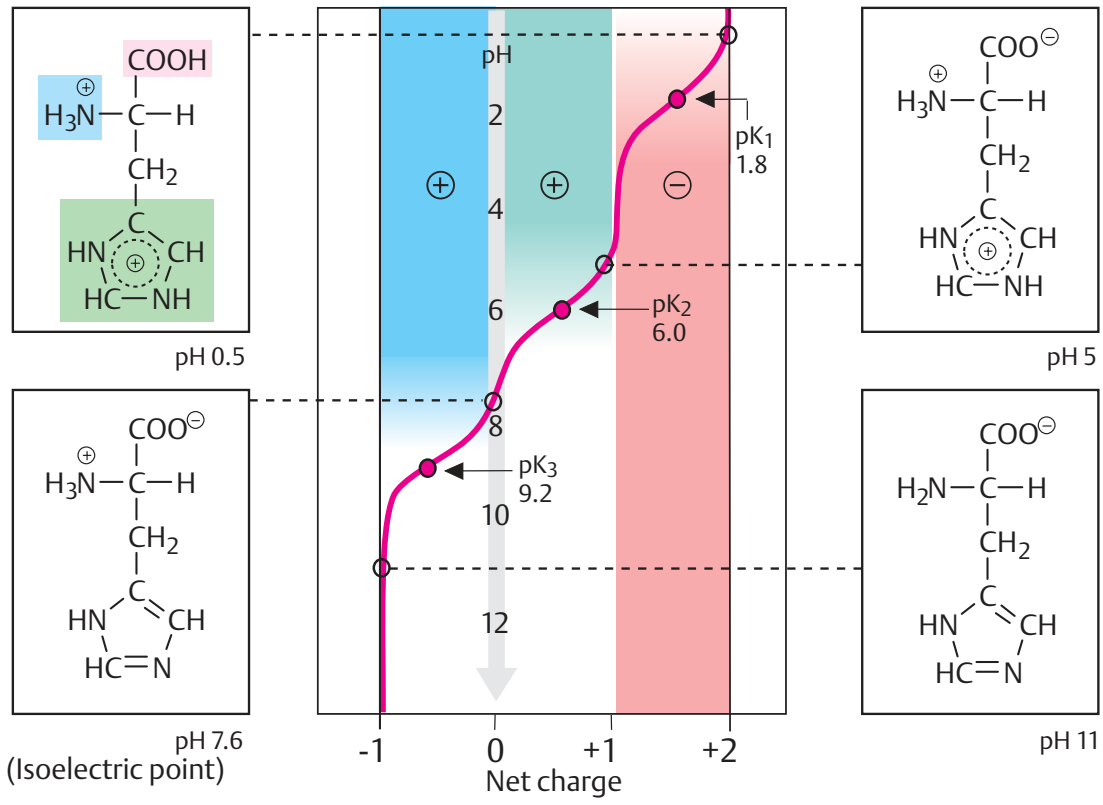
A. Amino acids: functions



B. Optical activity



C. Dissociation curve of histidine



Proteinogenic amino acids

A. The proteinogenic amino acids ◑

The amino acids that are included in the genetic code (see p. 248) are described as "proteinogenic." With a few exceptions (see p. 58), only these amino acids can be incorporated into proteins through *translation*. Only the side chains of the 20 proteinogenic amino acids are shown here. Their classification is based on the chemical structure of the side chains, on the one hand, and on their polarity on the other (see p. 6). The literature includes several slightly different systems for classifying amino acids, and details may differ from those in the system used here.

For each amino acid, the illustration names:
- *Membership of structural classes* I–VII (see below; e. g., III and VI for histidine)
- Name and abbreviation, formed from the first three letters of the name (e. g., histidine, His)
- The *one-letter symbol* introduced to save space in the electronic processing of sequence data (H for histidine)
- A quantitative *value for the polarity* of the side chain (bottom left; 10.3 for histidine). The more positive this value is, the *more polar* the amino acid is.

In addition, the polarity of the side chains is indicated by color. It increases from yellow, through light and dark green, to bluish green. For ionizing side chains, the corresponding pK_a values are also given (red numbers).

The **aliphatic** amino acids (class I) include *glycine, alanine, valine, leucine,* and *isoleucine*. These amino acids do not contain heteroatoms (N, O, or S) in their side chains and do not contain a ring system. Their side chains are markedly apolar. Together with threonine (see below), valine, leucine, and isoleucine form the group of *branched-chain amino acids*. The **sulfurcontaining amino acids** *cysteine* and *methionine* (class II), are also apolar. However, in the case of cysteine, this only applies to the undissociated state. Due to its ability to form disulfide bonds, cysteine plays an important role in the stabilization of proteins (see p. 72). Two cysteine residues linked by a disulfide bridge are referred to as *cystine* (not shown).

The **aromatic amino acids** (class III) contain resonancestabilized rings. In this group, only *phenylalanine* has strongly apolar properties. *Tyrosine* and *tryptophan* are moderately polar, and *histidine* is even strongly polar. The imidazole ring of histidine is already protonated at weakly acidic pH values. Histidine, which is only aromatic in protonated form (see p. 58), can therefore also be classified as a basic amino acid. Tyrosine and tryptophan show strong light absorption at wavelengths of 250–300 nm.

The **neutral** amino acids (class IV) have hydroxyl groups (*serine, threonine*) or amide groups (*asparagine, glutamine*). Despite their nonionic nature, the amide groups of asparagine and glutamine are markedly polar.

The carboxyl groups in the side chains of the **acidic** amino acids *aspartic acid* and *glutamic acid* (class V) are almost completely ionized at physiological pH values. The side chains of the **basic** amino acids *lysine* and *arginine* are also fully ionized—i. e., positively charged—at neutral pH. Arginine, with its positively charge guanidinium group, is particularly strongly basic, and therefore extremely polar.

Proline (VII) is a special case. Together with the α-C atom and the α-NH_2 group, its side chain forms a fivemembered ring. Its nitrogen atom is only weakly basic and is not protonated at physiological pH. Due to its ring structure, proline causes *bending of the peptide chain* in proteins (this is important in collagen, for example; see p. 70).

Several proteinogenic amino acids cannot be synthesized by the human organism, and therefore have to be supplied from the diet. These **essential amino acids** (see p. 360) are marked with a star in the illustration. Histidine and possibly also arginine are essential for infants and small children.

A. The proteinogenic amino acids

Aliphatic

Glycine (Gly, G)	Alanine (Ala, A)	Valine (Val, V) ☆	Leucine (Leu, L) ☆	Isoleucine (Ile, I) ☆
H	CH_3	$H_3C-CH(CH_3)$	$CH_2(H_3C-CH)(CH_3)$	$H_3C-C(H)(CH_2CH_3)$
Polarity				
−2.4	−1.9	−2.0	−2.3	−2.2

Sulfur-containing

Cysteine (Cys, C)	Methionine (Met, M) ☆
CH_2SH (8.3)	$CH_2CH_2SCH_3$
pK_a value	
−1.2	−1.5

Aromatic

Phenylalanine (Phe, F) ☆	Tyrosine (Tyr, Y)	Tryptophan (Trp, W) ☆
CH_2-phenyl	CH_2-phenyl-OH (10.1)	CH_2-indole
		Indole ring
+0.8	+6.1	+5.9

Cyclic

Proline (Pro, P)
Pyrrolidine ring
+6.0

Neutral

Serine (Ser, S)	Threonine (Thr, T) ☆
CH_2OH	$H_3C-C(H)(OH)$
+5.1	+4.9

☆ Essential amino acids

□ Chiral center

Neutral

Asparagine (Asn,N)	Glutamine (Gln, Q)
CH_2CONH_2	$CH_2CH_2CONH_2$
+9.7	+9.4

Acidic

Aspartic acid (Asp, D)	Glutamic acid (Glu, E)
CH_2COO^- (4.0)	$CH_2CH_2COO^-$ (4.3)
+11.0	+10.2

Basic

Histidine (His, H)	Lysine (Lys,K) ☆	Arginine (Arg, R)
Imidazole ring (6.0)	$CH_2CH_2CH_2CH_2NH_3^+$ (10.8)	$CH_2CH_2CH_2NH$-C(=NH_2^+)-NH_2 (12.5)
+10.3	+15.0	+20.0

Non-proteinogenic amino acids

In addition to the 20 proteinogenic amino acids (see p. 60), there are also many more compounds of the same type in nature. These arise during metabolic reactions (**A**) or as a result of enzymatic modifications of amino acid residues in peptides or proteins (**B**). The "biogenic amines" (**C**) are synthesized from α-amino acids by decarboxylation.

A. Rare amino acids ○

Only a few important representatives of the non-proteinogenic amino acids are mentioned here. The basic amino acid **ornithine** is an analogue of lysine with a shortened side chain. Transfer of a carbamoyl residue to ornithine yields **citrulline**. Both of these amino acids are intermediates in the urea cycle (see p. 182). **Dopa** (an acronym of 3,4-**d**ihydr**o**xy-**p**henyl**a**lanine) is synthesized by hydroxylation of tyrosine. It is an intermediate in the biosynthesis of catecholamines (see p. 352) and of melanin. It is in clinical use in the treatment of *Parkinson's disease*. **Selenocysteine**, a cysteine analogue, occurs as a component of a few proteins—e. g., in the enzyme glutathione peroxidase (see p. 284).

B. Post-translational protein modification ◑

Subsequent alteration of amino acid residues in finished peptides and proteins is referred to as *post-translational modification.* These reactions usually only involve polar amino acid residues, and they serve various purposes.

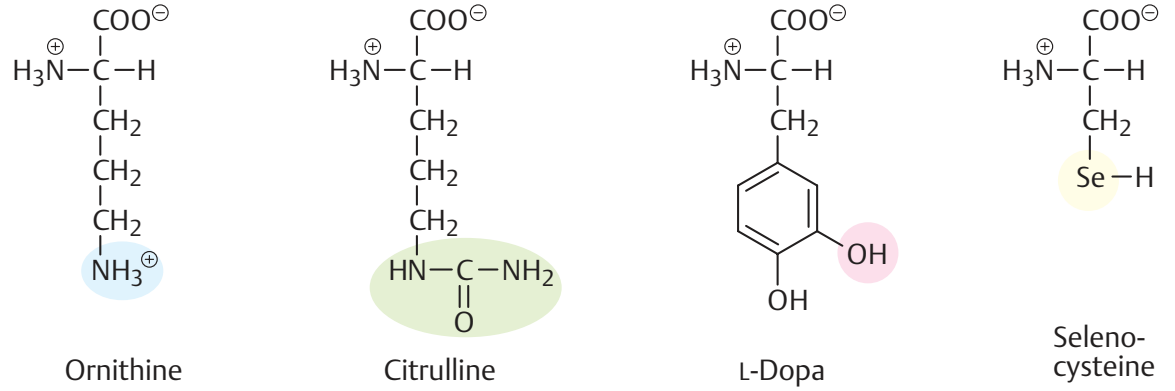
The free α-amino group at the *N*-terminus is blocked in many proteins by an acetyl residue or a longer acyl residue (**acylation**). *N*-terminal glutamate can cyclize into a pyroglutamate residue, while the *C*-terminal carboxylate group can be present in an amidated form (see TSH, p. 380). The side chains of serine and asparagine residues are often linked to oligosaccharides (**glycosylation**, see p. 230). **Phosphorylation** of proteins mainly affects serine and tyrosine residues. These reactions have mainly regulatory functions (see p. 114). Aspartate and histidine residues of enzymes are sometimes phosphorylated, too. A special modification of glutamate residues, **γ-carboxylation**, is found in coagulation factors. It is essential for blood coagulation (see p. 290). The ε-amino group of lysine residues is subject to a particularly large number of modifications. Its **acetylation** (or deacetylation) is an important mechanism for controlling genetic activity (see p. 244). Many coenzymes and cofactors are covalently linked to lysine residues. These include biotin (see p. 108), lipoic acid (see p. 106), and pyridoxal phosphate (see p. 108), as well as retinal (see p. 358). Covalent modification with **ubiquitin** marks proteins for breakdown (see p. 176). In collagen, lysine and proline residues are modified by **hydroxylation** to prepare for the formation of stable fibrils (see p. 70). Cysteine residues form **disulfide bonds** with one another (see p. 72). Cysteine **prenylation** serves to anchor proteins in membranes (see p. 214). Covalent bonding of a cysteine residue with heme occurs in cytochrome *c*. Flavins are sometimes covalently bound to cysteine or histidine residues of enzymes. Among the modifications of tyrosine residues, conversion into iodinated **thyroxine** (see p. 374) is particularly interesting.

C. Biogenic amines ◑

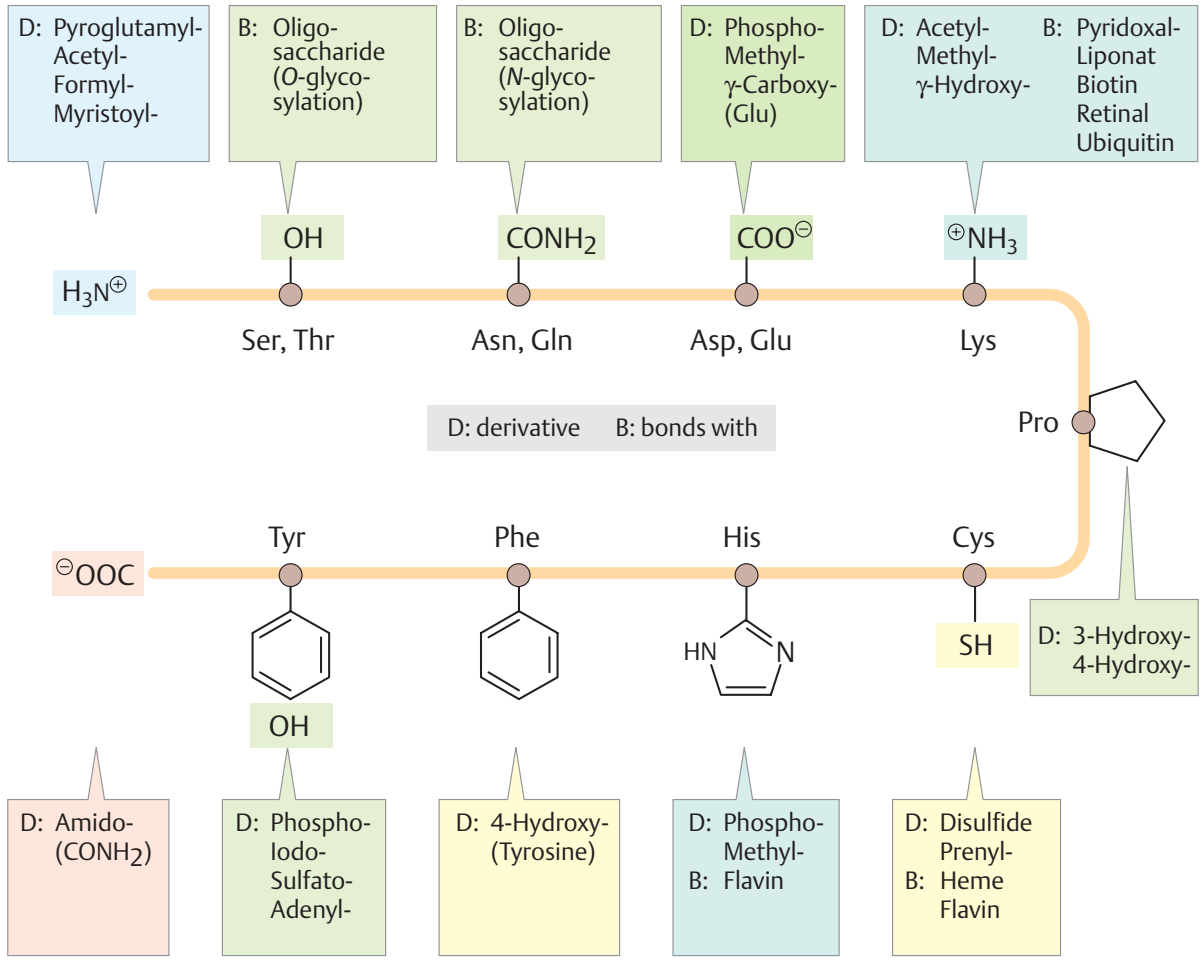
Several amino acids are broken down by *decarboxylation.* This reaction gives rise to what are known as biogenic amines, which have various functions. Some of them are **components of biomolecules**, such as *ethanolamine* in phospholipids (see p. 50). *Cysteamine* and *γ-alanine* are components of coenzyme A (see p. 12) and of pantetheine (see pp. 108, 168). Other amines function as signaling substances. An important **neurotransmitter** derived from glutamate is γ-aminobutyrate (GABA, see p. 356). The transmitter *dopamine* is also a precursor for the catecholamines epinephrine and norepinephrine (see p. 352). The biogenic amine *serotonin*, a substance that has many effects, is synthesized from tryptophan via the intermediate 5-hydroxytryptophan.

Monamines are inactivated into aldehydes by *amine oxidase* (monoamine oxidase, "MAO") with deamination and simultaneous oxidation. MAO inhibitors therefore play an important role in pharmacological interventions in neurotransmitter metabolism.

A. Rare amino acids



B. Post-translational protein modification



C. Biogenic amines

Amino acid	Amine	Function
Serine	Ethanol-amine	Glutamate
Cysteine	Cysteamine	Component of coenzyme A
Threonine	Amino-propanol	Component of vitamin B_{12}
Aspartate	β-Alanine	Component of coenzyme A

Amino acid	Amine	Function
Glutamate	γ-Amino-butyrate	Neurotrans-mitter (GABA)
Histidine	Histamine	Mediator, neuro-transmitter
Dopa	Dopamine	Neurotransmitter
5-Hydroxy-tryptophan	Serotonin	Mediator, neuro-transmitter

Peptides and proteins: overview

A. Proteins ◑

When amino acids are linked together by acid–amide bonds, linear macromolecules (peptides) are produced. Those containing more than ca. 100 amino acid residues are described as **proteins** (polypeptides). Every organism contains thousands of different proteins, which have a variety of functions. At a magnification of ca. 1.5 million, the semischematic illustration shows the structures of a few intra and extracellular proteins, giving an impression of their variety. The functions of proteins can be classified as follows.

Establishment and maintenance of structure. Structural proteins are responsible for the *shape and stability* of cells and tissues. A small part of a **collagen** molecule is shown as an example (right; see p. 70). The complete molecule is 1.5 300 nm in size, and at the magnification used here it would be as long as three pages of the book. **Histones** are also structural proteins. They organize the arrangement of DNA in chromatin. The basic components of chromatin, the *nucleosomes* (top right; see p. 218) consist of an octameric complex of histones, around which the DNA is coiled.

Transport. A wellknown transport protein is **hemoglobin** in the erythrocytes (bottom left). It is responsible for the transport of oxygen and carbon dioxide between the lungs and tissues (see p. 282). The blood plasma also contains many other proteins with transport functions. **Prealbumin** (transthyretin; middle), for example, transports the thyroid hormones thyroxin and triiodothyronine. **Ion channels** and other integral membrane proteins (see p. 220) facilitate the transport of ions and metabolites across biological membranes.

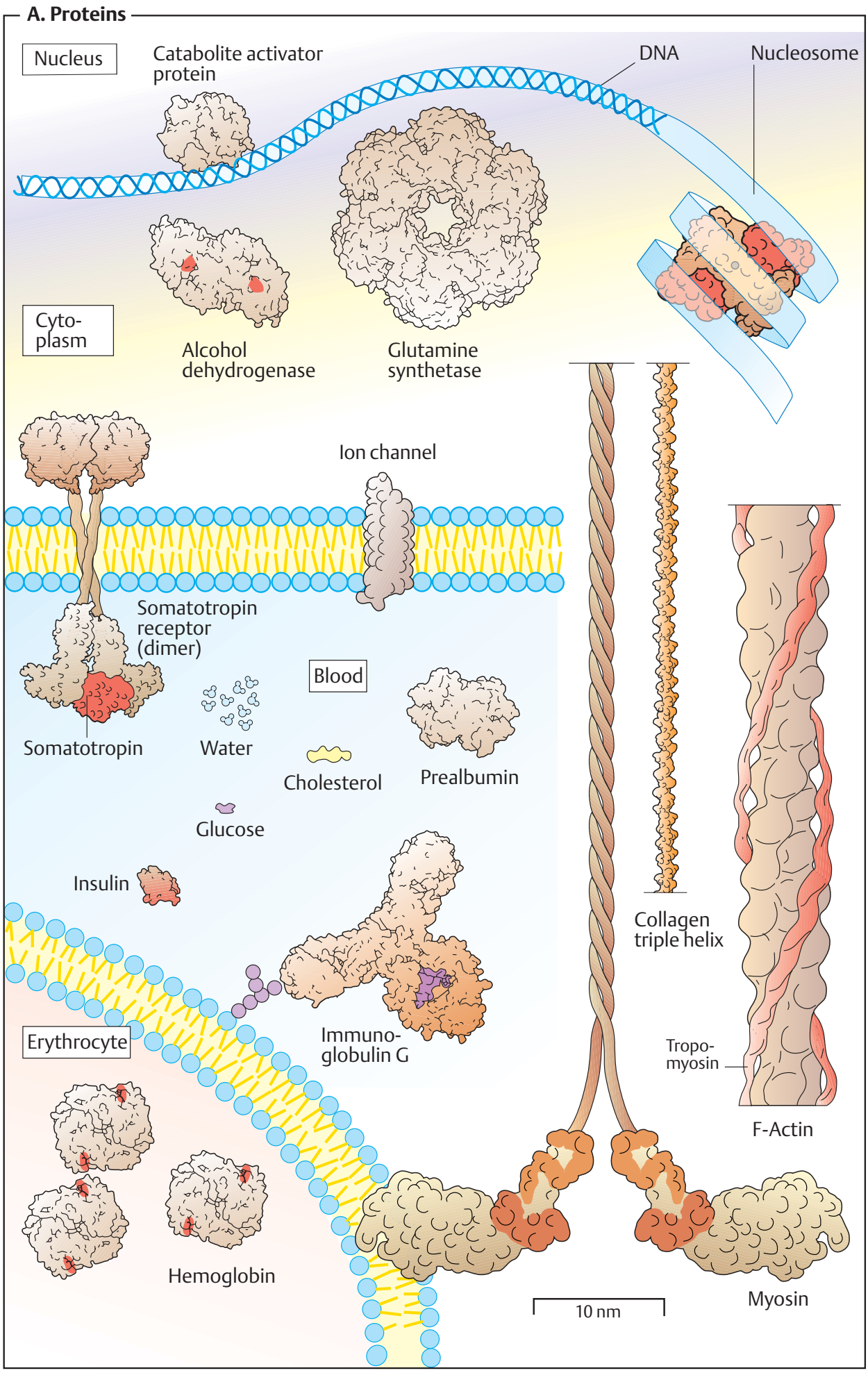
Protection and defense. The immune system protects the body from pathogens and foreign substances. An important component of this system is **immunoglobulin G** (bottom left; see p. 300). The molecule shown here is bound to an erythrocyte by complex formation with surface glycolipids (see p. 292).

Control and regulation. In biochemical signal chains, proteins function as signaling substances (hormones) and as hormone receptors. The complex between the growth hormone **somatotropin** and its **receptor** is shown here as an example (middle). Here, the extracellular domains of two receptor molecules here bind one molecule of the hormone. This binding activates the cytoplasmic domains of the complex, leading to further conduction of the signal to the interior of the cell (see p. 384). The small peptide hormone **insulin** is discussed in detail elsewhere (see pp. 76, 160). DNA-binding proteins (*transcription factors;* see p. 118) are decisively involved in regulating the metabolism and in differentiation processes. The structure and function of the **catabolite activator protein** (top left) and similar bacterial transcription factors have been particularly well investigated.

Catalysis. *Enzymes,* with more than 2000 known representatives, are the largest group of proteins in terms of numbers (see p. 88). The smallest enzymes have molecular masses of 10–15 kDa. Intermediatesized enzymes, such as **alcohol dehydrogenase** (top left) are around 100–200 kDa, and the largest—including **glutamine synthetase** with its 12 monomers (top right)—can reach more than 500 kDa.

Movement. The interaction between actin and myosin is responsible for muscle contraction and cell movement (see p. 332). **Myosin** (right), with a length of over 150 nm, is among the largest proteins there are. Actin filaments (**F-actin**) arise due to the polymerization of relatively small protein subunits (G-actin). Along with other proteins, **tropomyosin,** which is associated with F-actin, controls contraction.

Storage. Plants contain special **storage proteins**, which are also important for human nutrition (not shown). In animals, *muscle proteins* constitute a nutrient reserve that can be mobilized in emergencies.



Peptide bonds

A. Peptide bond ◑

The amino acid components of peptides and proteins are linked together by *amide* bonds (see p. 60) between α-carboxyl and α-amino groups. This type of bonding is therefore also known as **peptide bonding**. In the **dipeptide** shown here, the serine residue has a free ammonium group, while the carboxylate group in alanine is free. Since the amino acid with the free NH_3^+ group is named first, the peptide is known as **seryl alanine**, or in abbreviated form Ser-Ala or SA.

B. Resonance ◑

Like all acid–amide bonds, the peptide bond is **stabilized by resonance** (see p. 4). In the conventional notation (top right) it is represented as a combination of a C=O double bond with a C–N single bond. However, a C=N double bond with charges at O and N could also be written (middle). Both of these are only extreme cases of electron distribution, known as *resonance structures*. In reality, the π electrons are *delocalized* throughout all the atoms (bottom). As a mesomeric system, the peptide bond is *planar*. Rotation around the C–N bond would only be possible at the expense of large amounts of energy, and the bond is therefore *not freely rotatable*. Rotations are only possible around the single bonds marked with arrows. The state of these is expressed using the angles ϕ and ψ (see **D**). The plane in which the atoms of the peptide bond lie is highlighted in light blue here and on the following pages.

C. Peptide nomenclature ◑

Peptide chains have a *direction* and therefore two different ends. The amino terminus (**N terminus**) of a peptide has a free ammonium group, while the carboxy terminus (**C terminus**) is formed by the carboxylate group of the last amino acid. In peptides and proteins, the amino acid components are usually linked in linear fashion. To express the **sequence** of a peptide, it is therefore suf cient to combine the three-letter or single-letter abbreviations for the amino acid residues (see p. 60). This sequence always starts at the N terminus. For example, the peptide hormone *angiotensin II* (see p. 330) has the sequence Asp-Arg-Val-Tyr-Ile-His-Pro-Phe, or DRVYIHPF.

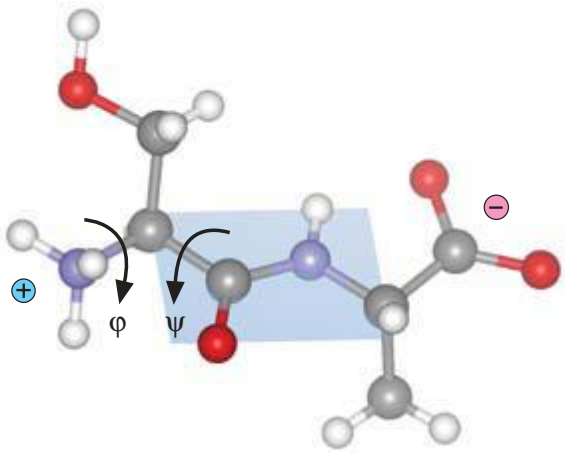
D. Conformational space of the peptide chain ○

With the exception of the terminal residues, every amino acid in a peptide is involved in *two* peptide bonds (one with the preceding residue and one with the following one). Due to the restricted rotation around the C–N bond, rotations are only possible around the N–C_α and C_α–C bonds (**2**). As mentioned above, these rotations are described by the dihedral angles ϕ (phi) and ψ (psi). The angle describes rotation around the N–C_α bond; ψ describes rotation around C_α–C—i. e., the position of the subsequent bond.

For steric reasons, only specific combinations of the dihedral angles are possible. These relationships can be illustrated clearly by a so-called ϕ/ψ *diagram* (**1**). Most combinations of ϕ and ψ are sterically "forbidden" (red areas). For example, the combination ϕ = 0° and ψ = 180° (**4**) would place the two carbonyl oxygen atoms less than 115 pm apart—i. e., at a distance much smaller than the sum of their van der Waals radii (see p. 6). Similarly, in the case of ϕ = 180° and ψ = 0° (**5**), the two NH hydrogen atoms would collide. The combinations located within the green areas are the only ones that are sterically feasible (e. g., **2** and **3**). The important secondary structures that are discussed in the following pages are also located in these areas. The conformations located in the yellow areas are energetically less favorable, but still possible.

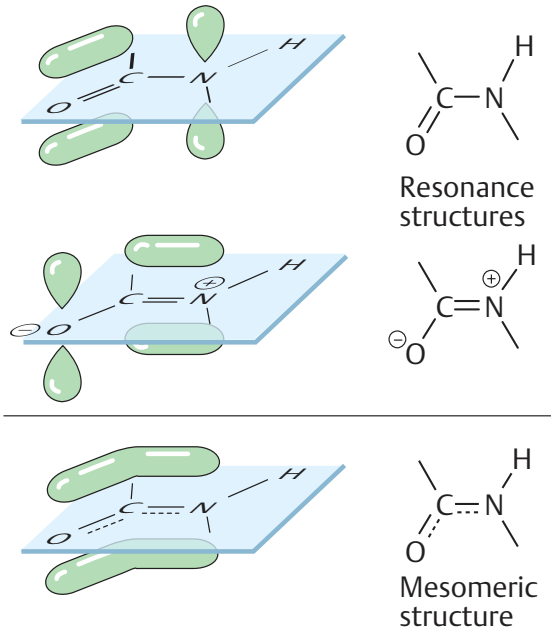
The ϕ/ψ diagram (also known as a **Ramachandran plot**) was developed from modeling studies of small peptides. However, the conformations of most of the amino acids in proteins are also located in the permitted areas. The corresponding data for the small protein, insulin (see p. 76), are represented by black dots in **1**.

A. Peptide bonds

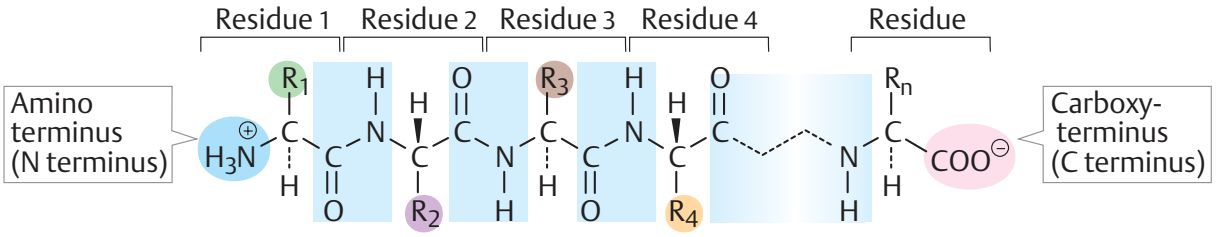


Seryl alanine
(Ser-Ala, $^{\oplus}H_3N$-Ser-Ala-$COO^{\ominus}$, SA)

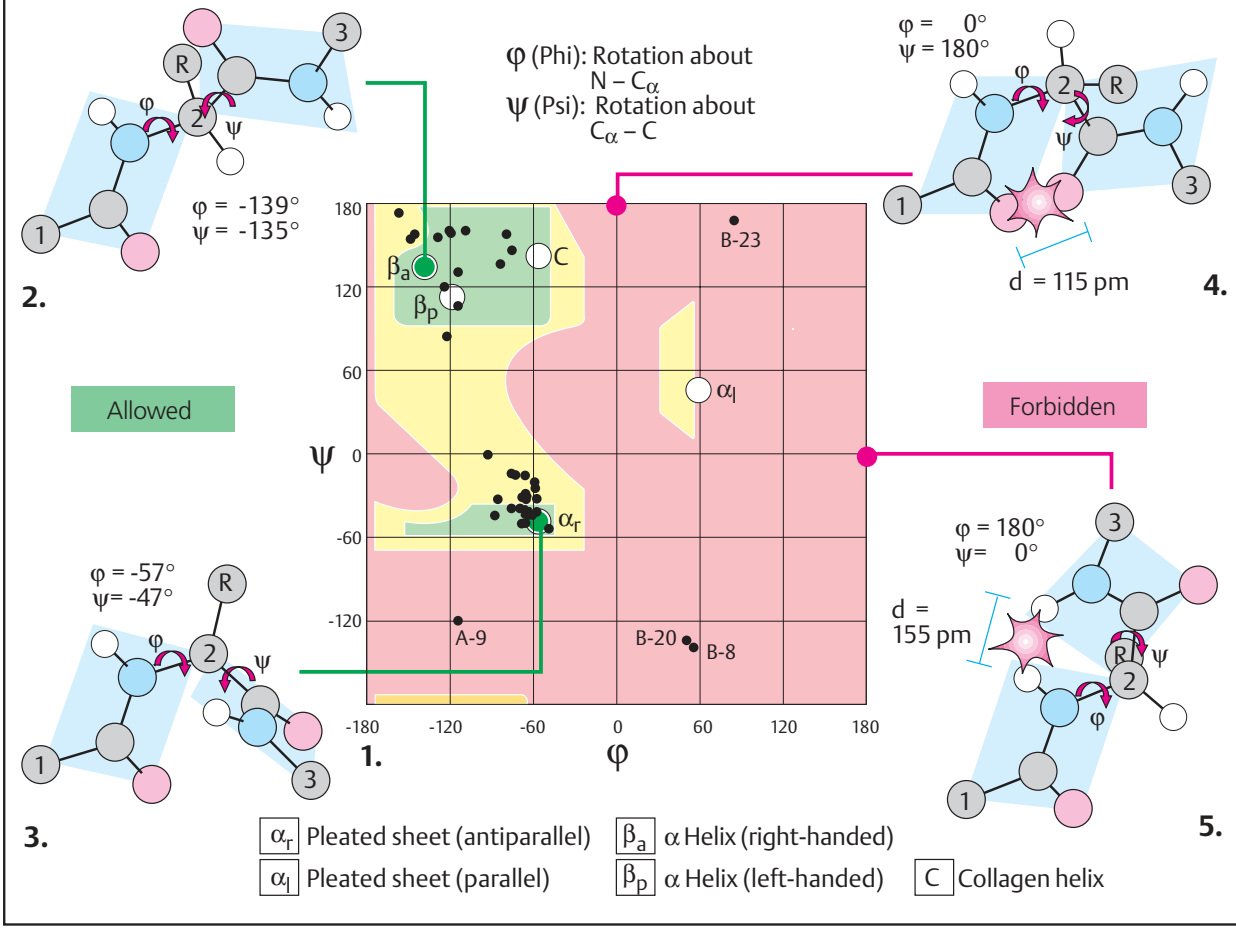
B. Resonance



C. Peptide nomenclature



D. Conformation space of the peptide chain



Secondary structures

In proteins, specific combinations of the dihedral angles ϕ and ψ (see p. 66) are much more common than others. When several successive residues adopt one of these conformations, defined **secondary structures** arise, which are stabilized by hydrogen bonds either within the peptide chain or between neighboring chains. When a large part of a protein takes on a defined secondary structure, the protein often forms mechanically stable filaments or fibers. **Structural proteins** of this type (see p. 70) usually have characteristic amino acid compositions.

The most important secondary structural elements of proteins are discussed here first. The illustrations only show the course of the peptide chain; the side chains are omitted. To make the course of the chains clearer, the levels of the peptide bonds are shown as blue planes. The dihedral angles of the structures shown here are also marked in diagram D1 on p. 67.

A. α-Helix ◑

The **right-handed** α-helix (α_R) is one of the most common secondary structures. In this conformation, the peptide chain is wound like a screw. Each turn of the screw (the screw axis in shown in orange) covers approximately 3.6 amino acid residues. The *pitch* of the screw (i. e., the smallest distance between two equivalent points) is 0.54 nm. α-Helices are stabilized by almost linear *hydrogen bonds* between the NH and CO groups of residues, which are four positions apart from each another in the sequence (indicated by red dots; see p. 6). In longer helices, most amino acid residues thus enter into *two* H bonds. Apolar or amphipathic α-helices with five to seven turns often serve to anchor proteins in biological membranes (*transmembrane helices;* see p. 214).

The mirror image of the α_R helix, the **left-handed α-helix** (α_L), is rarely found in nature, although it would be energetically "permissible."

B. Collagen helix ◑

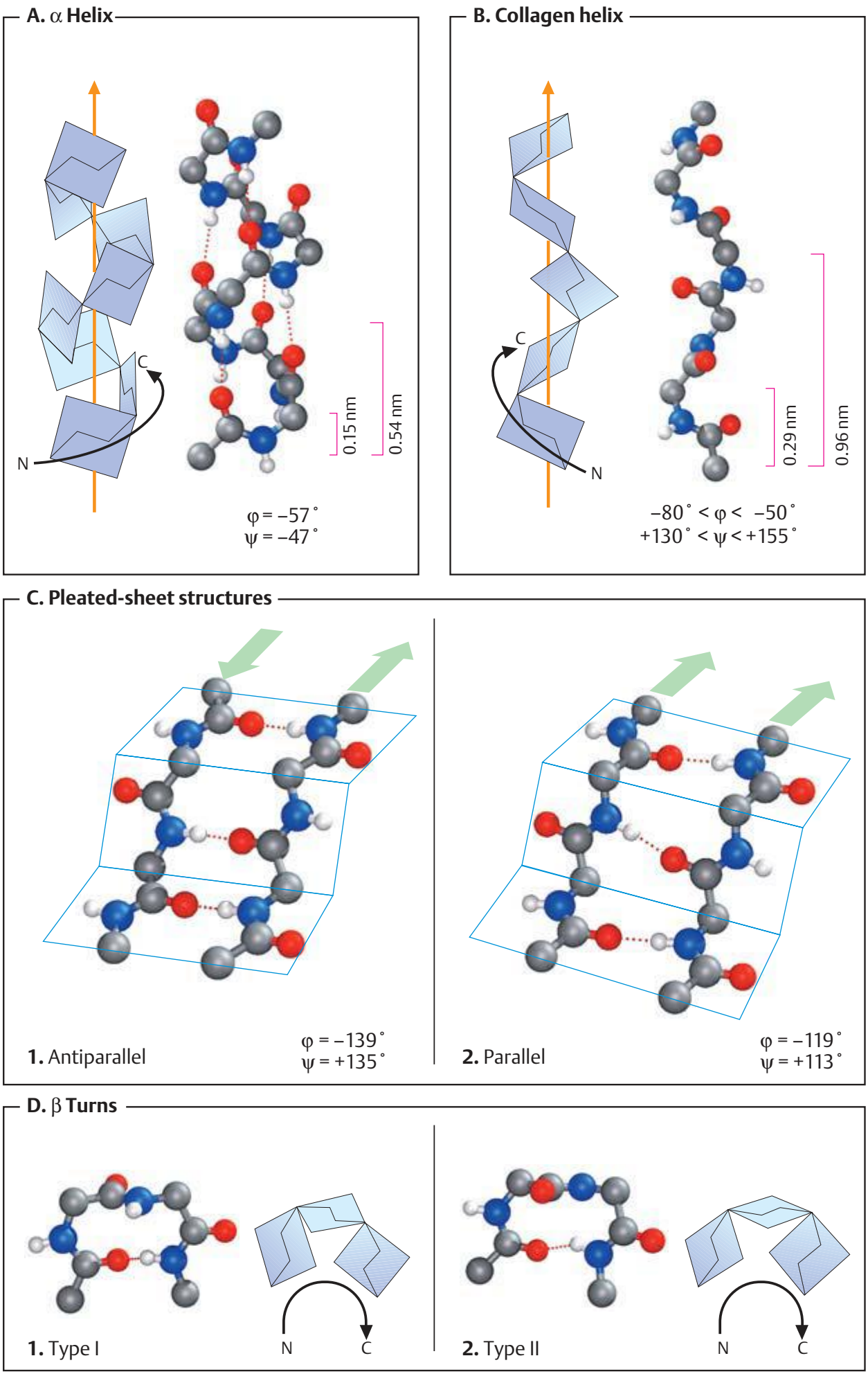
Another type of helix occurs in the collagens, which are important constituents of the connectivetissue matrix (see pp. 70, 344). The **collagen helix** is **left-handed**, and with a pitch of 0.96 nm and 3.3 residues per turn, it is steeper than the α-helix. In contrast to the α-helix, H bonds are not possible *within* the collagen helix. However, the conformation is stabilized by the association of three helices to form a righthanded **collagen triple helix** (see p. 70).

C. Pleated-sheet structures ◑

Two additional, almost stretched, conformations of the peptide chain are known as **β pleated sheets**, as the peptide planes are arranged like a regularly folded sheet of paper. Again, H bonds can only form between *neighboring chains* ("strands") in pleated sheets. When the two strands run in opposite directions (**1**), the structure is referred to as an **antiparallel pleated sheet** (β_a). When they run in the same direction (**2**), it is a **parallel pleated sheet** (β_p). In both cases, the α-C atoms occupy the highest and lowest points in the structure, and the side chains point alternately straight up or straight down (see p. 71 C). The β_a structure, with its almost linear H bonds, is energetically more favorable. In extended pleated sheets, the individual strands of the sheet are usually not parallel, but twisted relative to one another (see p. 74).

D. β Turns ◑

β Turns are often found at sites where the peptide chain changes direction. These are sections in which four amino acid residues are arranged in such a way that the course of the chain reverses by about 180° into the opposite direction. The two turns shown (types I and II) are particularly frequent. Both are stabilized by hydrogen bonds between residues 1 and 4. β Turns are often located between the individual strands of antiparallel pleated sheets, or between strands of pleated sheets and α helices.



Structural proteins

The **structural proteins** give extracellular structures mechanical stability, and are involved in the structure of the cytoskeleton (see p. 204). Most of these proteins contain a high percentage of specific secondary structures (see p. 68). For this reason, the amino acid composition of many structural proteins is also characteristic (see below).

A. α Keratin ○

α-Keratin is a structural protein that predominantly consists of α helices. Hair (wool), feathers, nails, claws and the hooves of animals consist largely of keratin. It is also an important component of the cytoskeleton (cytokeratin), where it appears in intermediate filaments (see p. 204).

In the keratins, large parts of the peptide chain show right-handed α-helical coiling. Two chains each form a left-handed **superhelix**, as is also seen in myosin (see p. 65). The superhelical keratin dimers join to form tetramers, and these aggregate further to form **protofilaments**, with a diameter of 3 nm. Finally, eight protofilaments then form an **intermediate filament**, with a diameter of 10 nm (see p. 204).

Similar keratin filaments are found in **hair**. In a single wool fiber with a diameter of about 20 μm, millions of filaments are bundled together within dead cells. The individual keratin helices are cross-linked and stabilized by numerous disulfide bonds (see p. 72). This fact is exploited in the *perming* of hair. Initially, the disulfide bonds of hair keratin are disrupted by reduction with thiol compounds (see p. 8). The hair is then styled in the desired shape and heat-dried. In the process, new disulfide bonds are formed by oxidation, which maintain the hairstyle for some time.

B. Collagen ◑

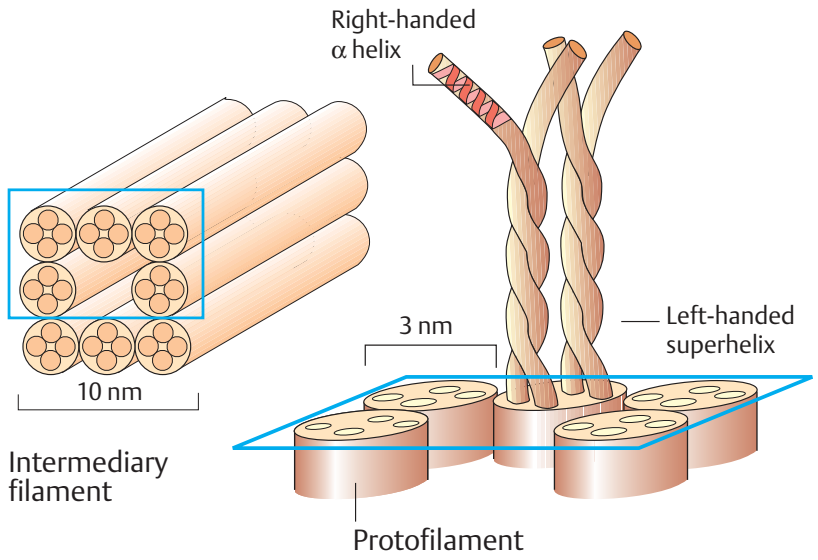
Collagen is the quantitatively most important protein in mammals, making up about 25% of the total protein. There are many different types of collagen, particularly in connective tissue. Collagen has an unusual amino acid composition. Approximately one-third of the amino acids are *glycine* (Gly), about 10% *proline* (Pro), and 10% *hydroxyproline* (Hyp). The two latter amino acids are only formed during collagen biosynthesis as a result of *posttranslational modification* (see p. 344).

The triplet Gly-X-Y (**2**) is constantly repeated in the sequence of collagen, with the X position often being occupied by Pro and the Y position by Hyp. The reason for this is that collagen is largely present as a **triple helix** made up of three individual collagen helices (**1**). In triple helices, every third residue lies on the inside of the molecule, where for steric reasons there is only room for glycine residues (**3**; the glycine residues are shown in yellow). Only a small section of a triple helix is illustrated here. The complete collagen molecule is approximately 300 nm long.

C. Silk fibroin ○

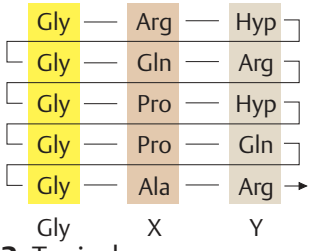
Silk is produced from the spun threads from silkworms (the larvae of the moth *Bombyx mori* and related species). The main protein in silk, **fibroin**, consists of antiparallel *pleated sheet structures* arranged one on top of the other in numerous layers (**1**). Since the amino acid side chains in pleated sheets point either straight up or straight down (see p. 68), only compact side chains fit between the layers. In fact, more than 80% of fibroin consists of glycine, alanine, and serine, the three amino acids with the shortest side chains. A typical repetitive amino acid sequence is *(Gly-Ala-Gly-Ala-Gly-Ser)*. The individual pleated sheet layers in fibroin are found to lie alternately 0.35 nm and 0.57 nm apart. In the first case, only glycine residues (R = H) are opposed to one another. The slightly greater distance of 0.57 nm results from repulsion forces between the side chains of alanine and serine residues (**2**).

A. α-Keratin

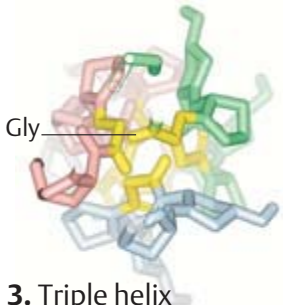


B. Collagen

1. Triple helix (section)

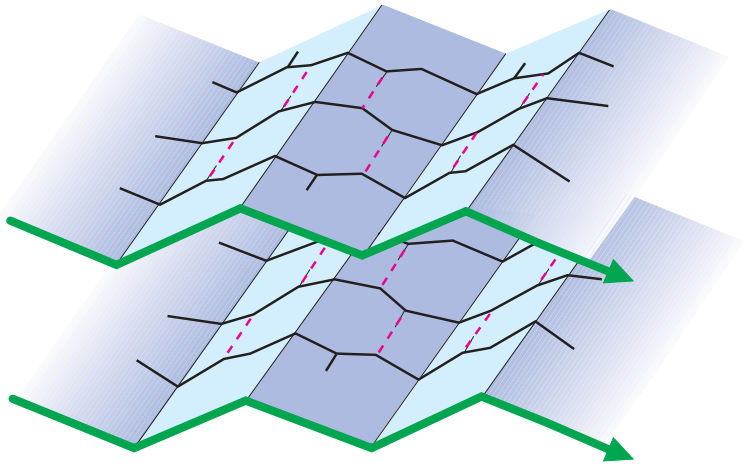


2. Typical sequence

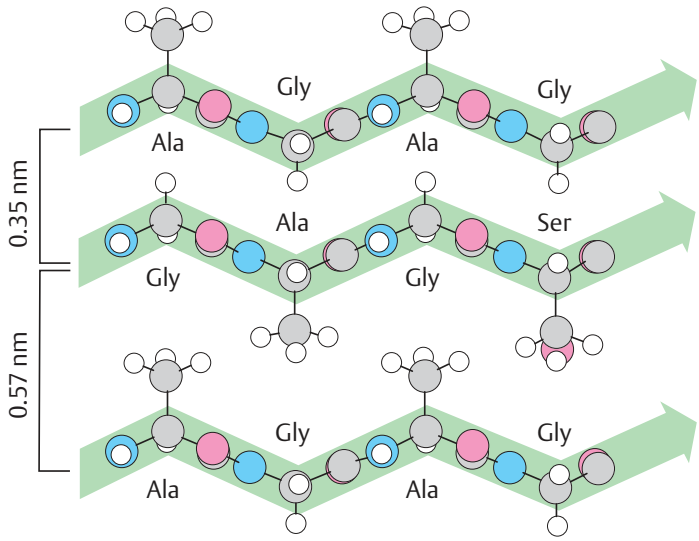


3. Triple helix (view from above)

C. Silk fibroin



1. Spatial illustration



2. Front view

Globular proteins

Soluble proteins have a more complex structure than the fibrous, completely insoluble structural proteins. The shape of soluble proteins is more or less spherical (globular). In their biologically active form, **globular proteins** have a defined spatial structure (the **native conformation**). If this structure is destroyed (**denaturation**; see p. 74), not only does the biological effect disappear, but the protein also usually precipitates in insoluble form. This happens, for example, when eggs are boiled; the proteins dissolved in the egg white are denatured by the heat and produce the solid egg white.

To illustrate protein conformations in a clear (but extremely simplified) way, *Richardson diagrams* are often used. In these diagrams, α-helices are symbolized by red cylinders or spirals and strands of pleated sheets by green arrows. Less structured areas of the chain, including the β-turns, are shown as sections of gray tubing.

A. Conformation-stabilizing interactions ◑

The native conformation of proteins is stabilized by a number of different interactions. Among these, only the **disulfide bonds** (**B**) represent covalent bonds. **Hydrogen bonds**, which can form inside secondary structures, as well as between more distant residues, are involved in all proteins (see p. 6). Many proteins are also stabilized by **complex formation** with metal ions (see pp. 76, 342, and 378, for example). The **hydrophobic effect** is particularly important for protein stability. In globular proteins, most hydrophobic amino acid residues are arranged in the interior of the structure in the native conformation, while the polar amino acids are mainly found on the surface (see pp. 28, 76).

B. Disulfide bonds ◑

Disulfide bonds arise when the SH groups of two cysteine residues are covalently linked as a dithiol by oxidation. Bonds of this type are only found (with a few exceptions) in extracellular proteins, because in the interior of the cell *glutathione* (see p. 284) and other reducing compounds are present in such high concentrations that disulfides would be reductively cleaved again. The small plant protein *crambin* (46 amino acids) contains three disulfide bonds and is therefore very stable. The high degree of stability of insulin (see p. 76) has a similar reason.

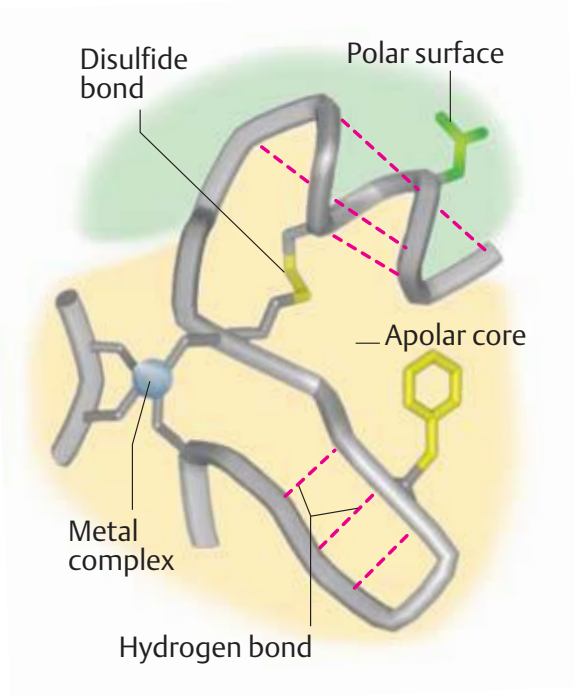
C. Protein dynamics ◑

The conformations of globular proteins are not rigid, but can change dramatically on binding of ligands or in contact with other proteins. For example, the enzyme *adenylate kinase* (see p. 336) has a mobile domain (domain = independently folded partial structure), which folds shut after binding of the substrate (yellow). The larger domain (bottom) also markedly alters its conformation. There are large numbers of **allosteric proteins** of this type. This group includes, for example, *hemoglobin* (see p. 280), *calmodulin* (see p. 386), and many allosteric enzymes such as *aspartate carbamoyltransferase* (see p. 116).

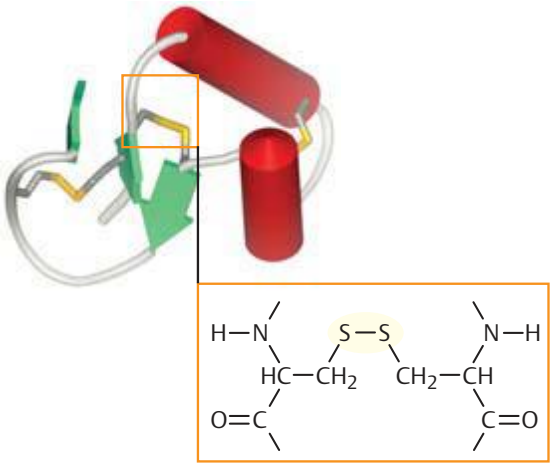
D. Folding patterns ○

The globular proteins show a high degree of variability in folding of their peptide chains. Only a few examples are shown here. Purely helically folded proteins such as *myoglobin* (**1**; see p. 74, heme yellow) are rare. In general, pleated sheet and helical elements exist alongside each other. In the hormone-binding domain of the *estrogen receptor* (**2**; see p. 378), a small, two-stranded pleated sheet functions as a "cover" for the hormone binding site (estradiol yellow). In *flavodoxin,* a small flavoprotein with a redox function (**3**; FMN yellow), a fan-shaped, pleated sheet made up of five parallel strands forms the core of the molecule. The conformation of the β subunit of the G-protein *transducin* (**4**; see pp. 224, 358) is very unusual. Seven pleated sheets form a large, symmetrical "β propeller." The *N*-terminal section of the protein contains one long and one short helix.

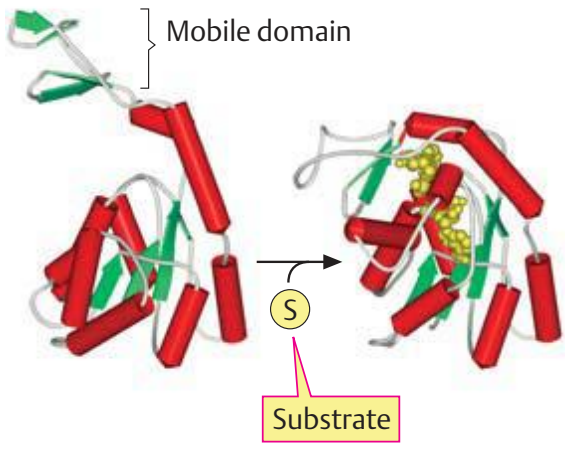
A. Conformation-stabilizing interactions



B. Disulfide bonds



C. Protein dynamics

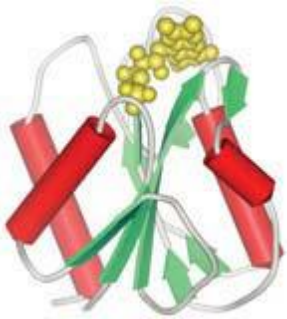


Adenylate kinase

D. Folding patterns

1. Myoglobin

2. Estrogen receptor (domain)



3. Flavodoxin

4. Transducin (β subunit)

Protein folding

Information about the biologically active (*native*) conformation of proteins is already encoded in their amino acid sequences. The native forms of many proteins arise spontaneously in the test tube and within a few minutes. Nevertheless, there are special auxiliary proteins (chaperonines) that support the folding of other proteins in the conditions present within the cell (see p. 232). An important goal of biochemistry is to understand the laws governing **protein folding**. This would make it possible to predict the conformation of a protein from the easily accessible DNA sequence (see p. 260).

A. Folding and denaturation of ribonuclease A ◑

The **folding** of proteins to the native form is favored under physiological conditions. The native conformation is lost, as the result of **denaturation**, at extreme pH values, at high temperatures, and in the presence of organic solvents, detergents, and other denaturing substances, such as urea.

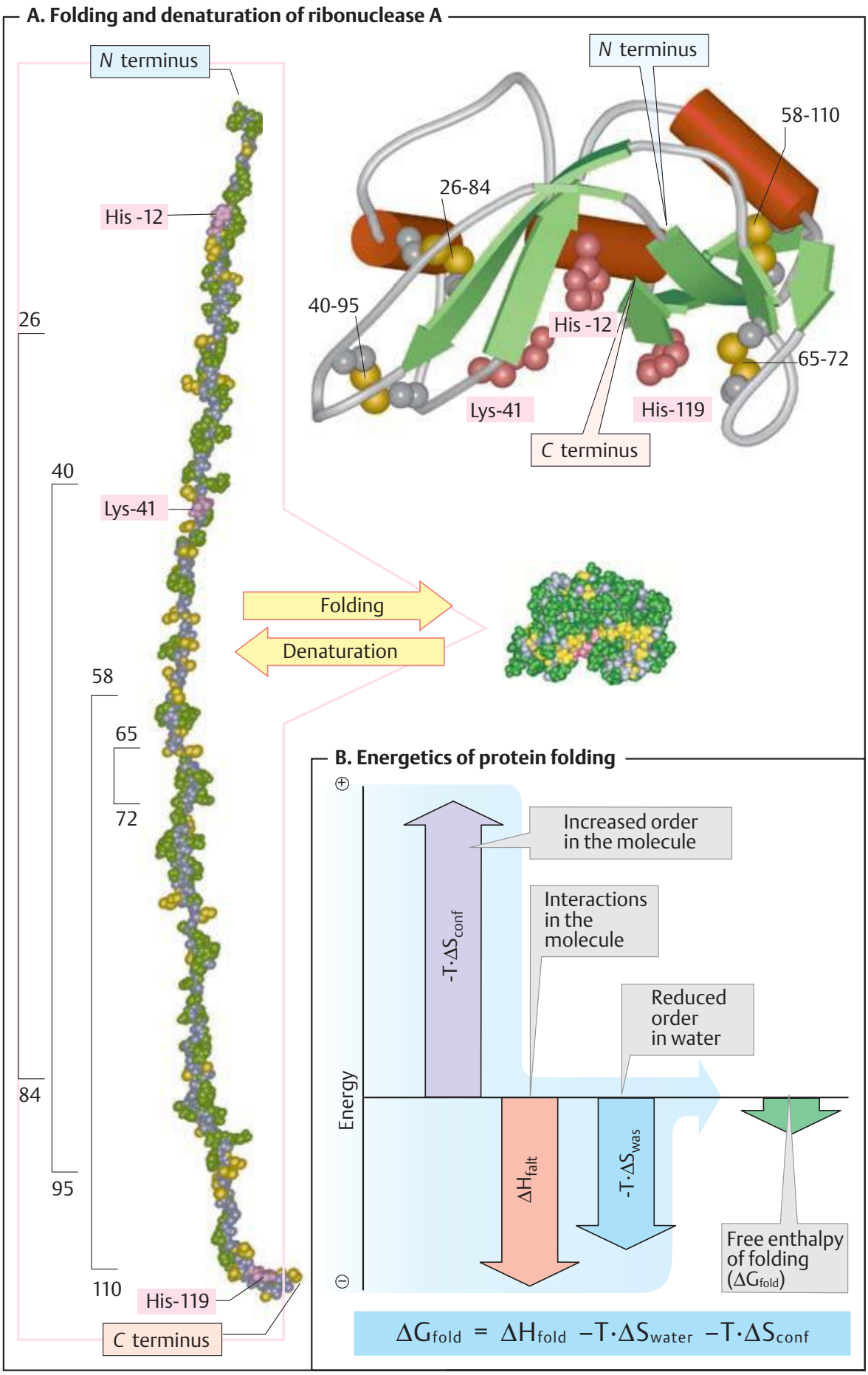
The fact that a denatured protein can spontaneously return to its native conformation was demonstrated for the first time with **ribonuclease**, a digestive enzyme (see p. 266) consisting of 124 amino acids. In the native form (top right), there are extensive pleated sheet structures and three α helices. The eight cysteine residues of the protein are forming four disulfide bonds. Residues His-12, Lys-41 and His-119 (pink) are particularly important for catalysis. Together with additional amino acids, they form the enzyme's *active center.*

The disulfide bonds can be reductively cleaved by *thiols* (e.g., mercaptoethanol, $HO\text{-}CH_2\text{-}CH_2\text{-}SH$). If *urea* at a high concentration is also added, the protein unfolds completely. In this form (left), it is up to 35 nm long. Polar (green) and apolar (yellow) side chains are distributed randomly. The denatured enzyme is completely inactive, because the catalytically important amino acids (pink) are too far away from each other to be able to interact with each other and with the substrate.

When the urea and thiol are removed by dialysis (see p. 78), secondary and tertiary structures develop again spontaneously. The cysteine residues thus return to a suf ciently close spatial vicinity that disulfide bonds can once again form under the oxidative effect of atmospheric oxygen. The active center also reestablishes itself. In comparison with the denatured protein, the native form is astonishingly compact, at 4.5 2.5 nm. In this state, the apolar side chains (yellow) predominate in the interior of the protein, while the polar residues are mainly found on the surface. This distribution is due to the "hydrophobic effect" (see p. 28), and it makes a vital contribution to the stability of the native conformation (**B**).

B. Energetics of protein folding ○

The **energetics** of protein folding are not at present satisfactorily understood. Only a simplified model is discussed here. The conformation of a molecule is stable in any given conditions if the change in its free enthalpy during folding (ΔG_{fold}) is negative (see p. 16). The magnitude of the folding enthalpy is affected by several factors. The main factor working *against* folding is the strong increase in the ordering of the molecule involved. As discussed on p. 20, this leads to a negative change in entropy of ΔS_{conf} and therefore to a strongly positive entropy term $-T\ \Delta S$ (violet arrow). By contrast, the covalent and noncovalent bonds in the interior of the protein have a *stabilizing* influence. For this reason, the change in folding enthalpy ΔH_{fold} is negative (red arrow). A third factor is the change in the system's entropy due to the hydrophobic effect. During folding, the degree of order in the *surrounding water* decreases—i.e., ΔS_{water} is positive and therefore $-T\ \Delta S$ is negative (blue arrow). When the sum of these effects is negative (green arrow), the protein folds spontaneously into its native conformation.



Molecular models: insulin

The opposite page presents models of insulin, a small protein. The biosynthesis and function of this important hormone are discussed elsewhere in this book (pp. 160, 388). Monomeric insulin consists of 51 amino acids, and with a molecular mass of 5.5 kDa it is only half the size of the smallest enzymes. Nevertheless, it has the typical properties of a globular protein.

Large quantities of pure insulin are required for the treatment of *diabetes mellitus* (see p. 160). The annual requirement for insulin is over 500 kg in a country the size of Germany. Formerly, the hormone had to be obtained from the pancreas of slaughtered animals in a complicated and expensive procedure. **Human insulin**, which is produced by *overexpression* in genetically engineered bacteria, is now mainly used (see p. 262).

A. Structure of insulin ○

There are various different structural levels in proteins, and these can be briefly discussed again here using the example of insulin.

The **primary structure** of a protein is its amino acid sequence. During the biosynthesis of insulin in the pancreas, a continuous peptide chain with 84 residues is first synthesized—*proinsulin* (see p. 160). After folding of the molecule, the three disulfide bonds are first formed, and residues 31 to 63 are then proteolytically cleaved releasing the so-called *C peptide*. The molecule that is left over (**1**) now consists of two peptide chains, the *A chain* (21 residues, shown in yellow) and the *B chain* (30 residues, orange). One of the disulfide bonds is located inside the A chain, and the two others link the two chains together.

Secondary structures are regions of the peptide chain with a defined conformation (see p. 68) that are stabilized by H-bonds. In insulin (**2**), the α-helical areas are predominant, making up 57% of the molecule; 6% consists of β-pleated-sheet structures, and 10% of β-turns, while the remainder (27%) cannot be assigned to any of the secondary structures.

The three-dimensional conformation of a protein, made up of secondary structural elements and unordered sections, is referred to as the **tertiary structure**. In insulin, it is compact and wedge-shaped (**B**). The tip of the wedge is formed by the B chain, which changes its direction at this point.

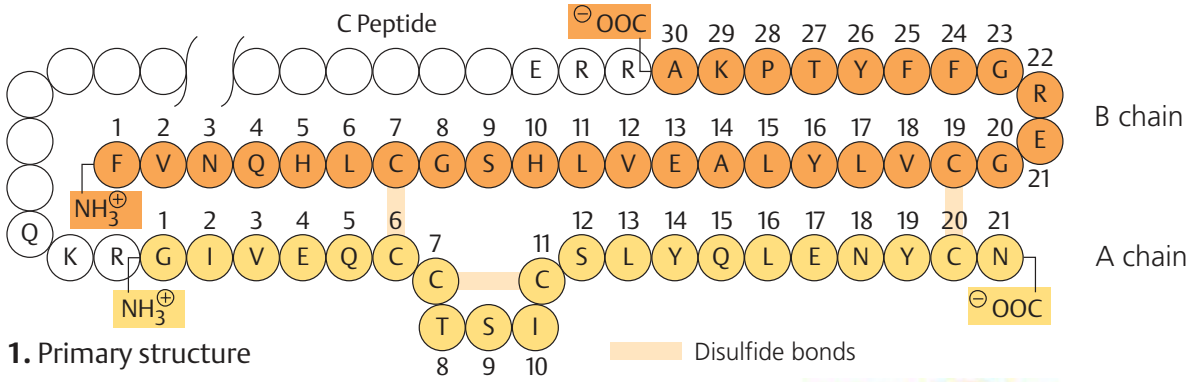
Quaternary structure. Due to non-covalent interactions, many proteins assemble to form symmetrical complexes (oligomers). The individual components of oligomeric proteins (usually 2–12) are termed *subunits* or *monomers*. Insulin also forms quaternary structures. In the blood, it is partly present as a dimer. In addition, there are also hexamers stabilized by Zn^{2+} ions (light blue) (**3**), which represent the form in which insulin is stored in the pancreas (see p. 160).

B. Insulin (monomer) ○

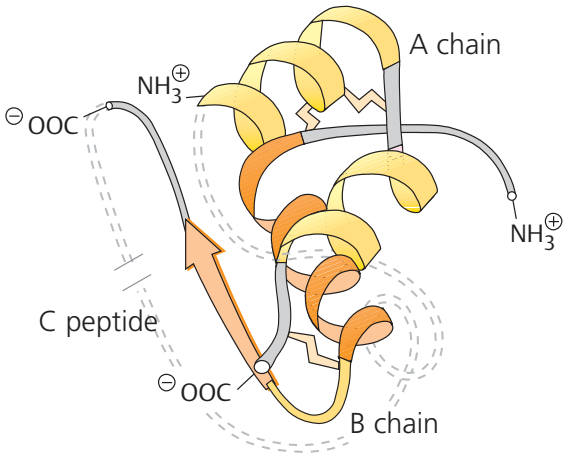
The van der Waals model of monomeric insulin (**1**) once again shows the wedge-shaped tertiary structure formed by the two chains together. In the second model (**3**, bottom), the side chains of polar amino acids are shown in blue, while apolar residues are yellow or pink. This model emphasizes the importance of the "hydrophobic effect" for protein folding (see p. 74). In insulin as well, most hydrophobic side chains are located on the inside of the molecule, while the hydrophilic residues are located on the surface. Apparently in contradiction to this rule, several apolar side chains (pink) are found on the surface. However, all of these residues are involved in hydrophobic interactions that stabilize the dimeric and hexameric forms of insulin.

In the third model (**2**, right), the colored residues are those that are located on the surface and occur *invariably* (red) or *almost invariably* (orange) in all known insulins. It is assumed that amino acid residues that are not replaced by other residues during the course of evolution are essential for the protein's function. In the case of insulin, almost all of these residues are located on one side of the molecule. They are probably involved in the binding of the hormone to its receptor (see p. 224).

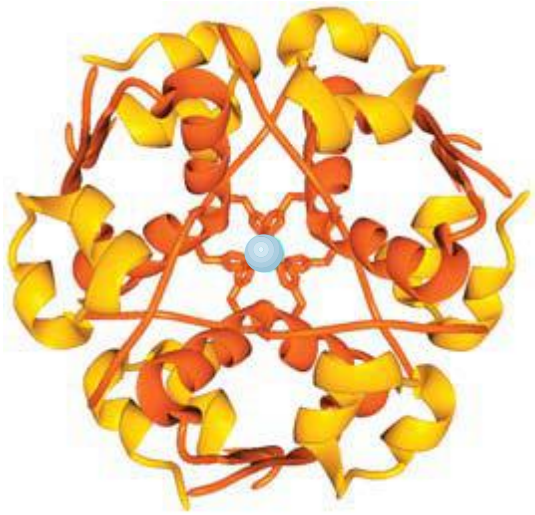
A. Structure of insulin



1. Primary structure

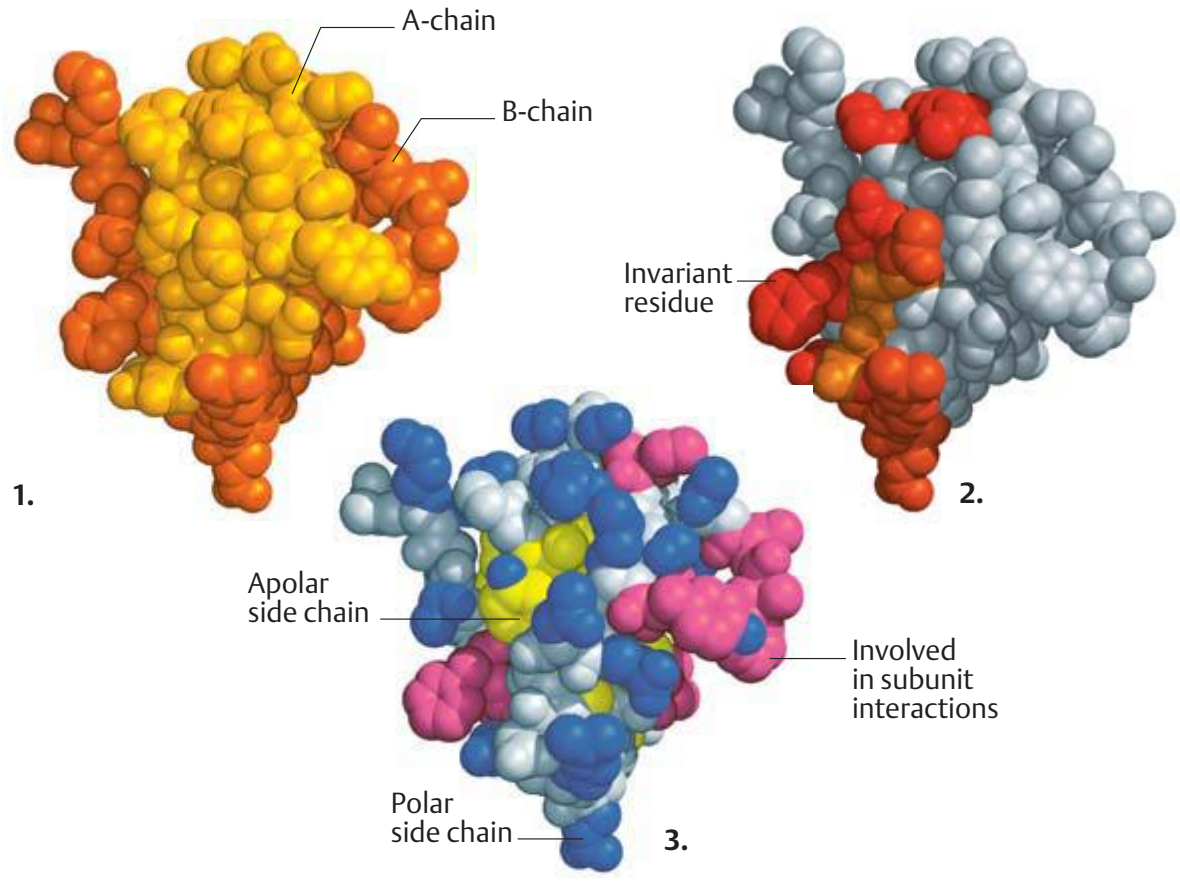


2. Secondary and tertiary structure



3. Quaternary structure

B. Insulin (monomer)



Isolation and analysis of proteins

Purified proteins are nowadays required for a wide variety of applications in research, medicine, and biotechnology. Since the globular proteins in particular are very unstable (see p. 72), purification is carried out at low temperatures (0–5 °C) and particularly gentle separation processes are used. A few of the methods of purifying and characterizing proteins are discussed on this page.

A. Salt precipitation ○

The solubility of proteins is strongly dependent on the salt concentration (*ionic strength*) of the medium. Proteins are usually poorly soluble in pure water. Their solubility increases as the ionic strength increases, because more and more of the well-hydrated anorganic ions (blue circles) are bound to the protein's surface, preventing aggregation of the molecules (**salting in**). At very high ionic strengths, the salt withdraws the hydrate water from the proteins and thus leads to aggregation and precipitation of the molecules (**salting out**). For this reason, adding salts such as ammonium sulfate $(NH_4)_2SO_4$ makes it possible to separate proteins from a mixture according to their degree of solubility (fractionation).

B. Dialysis ○

Dialysis is used to remove lower-molecular components from protein solutions, or to exchange the medium. Dialysis is based on the fact that due to their size, protein molecules are unable to pass through the pores of a **semipermeable membrane**, while lower-molecular substances distribute themselves evenly between the inner and outer spaces over time. After repeated exchanging of the external solution, the conditions inside the *dialysis tube* (salt concentration, pH, etc.) will be the same as in the surrounding solution.

C. Gel filtration ○

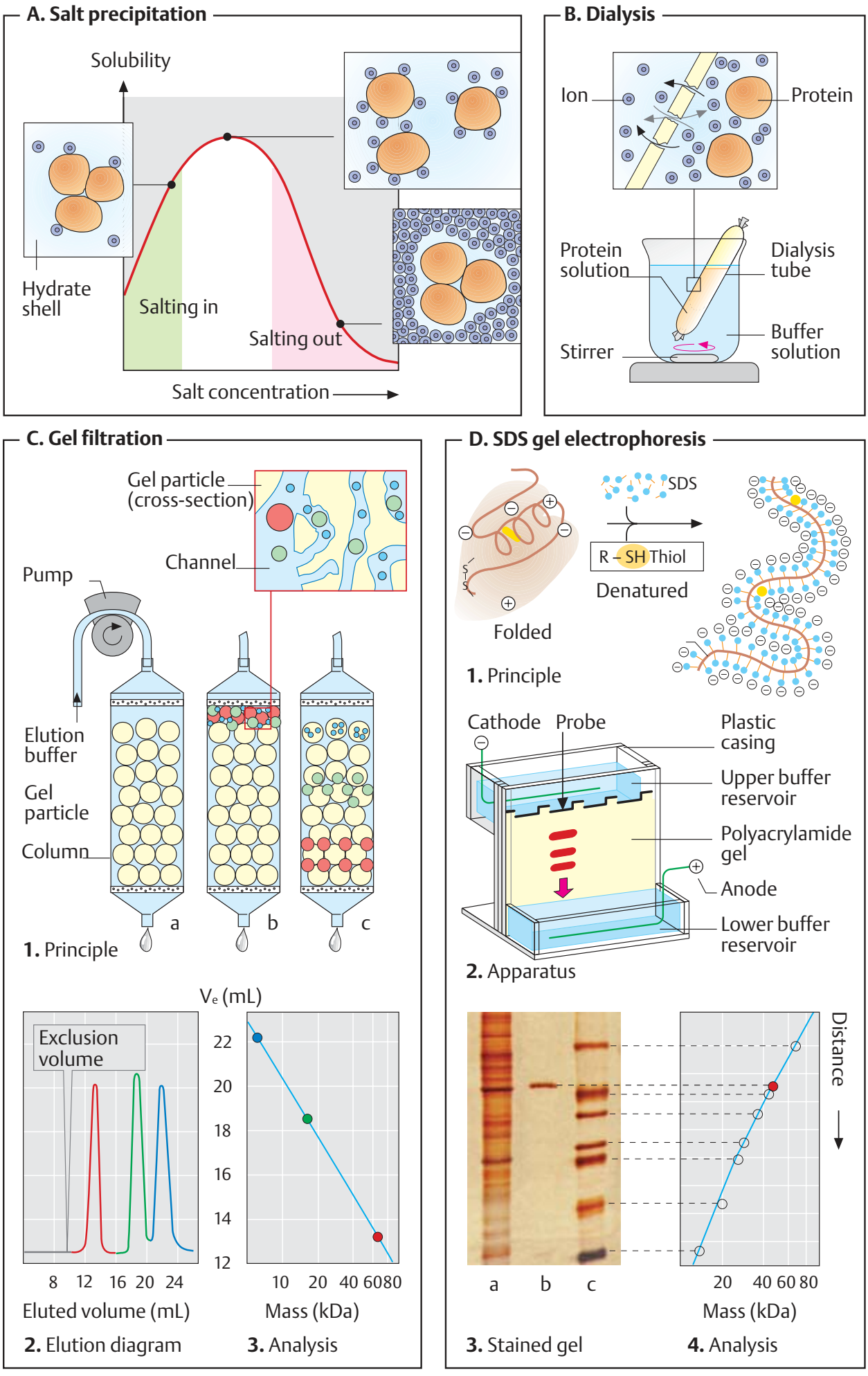
Gel permeation chromatography ("gel filtration") separates proteins according to their size and shape. This is done using a *chromatography column,* which is filled with spherical *gel particles* (diameter 10–500 μm) of polymeric material (shown schematically in **1a**). The insides of the particles are traversed by channels that have defined diameters. A protein mixture is then introduced at the upper end of the column (**1b**) and *elution* is carried out by passing a buffer solution through the column. Large protein molecules (red) are unable to penetrate the particles, and therefore pass through the column quickly. Medium-sized (green) and small particles (blue) are delayed for longer or shorter periods (**1c**). The proteins can be collected separately from the ef uent (*eluate*) (**2**). Their elution volume V_e depends mainly on their molecular mass (**3**).

D. SDS gel electrophoresis ○

The most commonly used procedure for checking the purity of proteins is sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). In electrophoresis, molecules move in an electrical field (see p. 276). Normally, the speed of their movement depends on three factors—their size, their shape, and their electrical charge.

In SDS-PAGE, the protein mixture is treated in such a way that only the molecules' mass affects their movement. This is achieved by adding *sodium dodecyl sulfate* ($C_{12}H_{25}$-OSO_3Na), the sulfuric acid ester of lauryl alcohol (dodecyl alcohol). SDS is a *detergent* with strongly amphipathic properties (see p. 28). It separates oligomeric proteins into their subunits and denatures them. SDS molecules bind to the unfolded peptide chains (ca. 0.4 g SDS / g protein) and give them a strongly negative charge. To achieve complete denaturation, thiols are also added in order to cleave the disulfide bonds (**1**).

Following electrophoresis, which is carried out in a vertically arranged gel of polymeric acrylamide (**2**), the separated proteins are made visible by staining. In example (**3**), the following were separated: **a**) a cell extract with hundreds of different proteins, **b**) a protein purified from this, and **c**) a mixture of proteins with known masses.



Bases and nucleotides

The nucleic acids play a central role in the storage and expression of genetic information (see p. 236). They are divided into two major classes: **deoxyribonucleic acid (DNA)** functions solely in information storage, while **ribonucleic acids (RNAs)** are involved in most steps of gene expression and protein biosynthesis. All nucleic acids are made up from **nucleotide components**, which in turn consist of a *base*, a *sugar*, and a *phosphate residue*. DNA and RNA differ from one another in the type of the sugar and in one of the bases that they contain.

A. Nucleic acid bases ◑

The bases that occur in nucleic acids are *aromatic* heterocyclic compounds derived from either **pyrimidine** or **purine**. Five of these bases are the main components of nucleic acids in all living creatures. The purine bases **adenine** (abbreviation Ade, *not* "A"!) and **guanine** (Gua) and the pyrimidine base **cytosine** (Cyt) are present in both RNA *and* DNA. In contrast, **uracil** (Ura) is only found in RNA. In DNA, **uracil** is replaced by thymine (Thy), the 5-methyl derivative of uracil. 5-methylcytosine also occurs in small amounts in the DNA of the higher animals. A large number of other modified bases occur in tRNA (see p. 82) and in other types of RNA.

B. Nucleosides, nucleotides ◑

When a nucleic acid base is N-glycosidically linked to ribose or 2-deoxyribose (see p. 38), it yields a **nucleoside**. The nucleoside **adenosine** (abbreviation: A) is formed in this way from adenine and ribose, for example. The corresponding derivatives of the other bases are called *guanosine* (G), *uridine* (U), *thymidine* (T) and *cytidine* (C). When the sugar component is 2-deoxyribose, the product is a **deoxyribonucleoside**—e. g., 2′-deoxyadenosine (dA, not shown). In the cell, the 5′OH group of the sugar component of the nucleoside is usually esterified with phosphoric acid. 2′-Deoxythymidine (dT) therefore gives rise to **2′-deoxythymidine-5′-monophosphate (dTMP)**, one of the components of DNA (**2**). If the 5′phosphate residue is linked via an acid–anhydride bond to additional phosphate residues, it yields nucleoside diphosphates and triphosphates—e. g., ADP and ATP, which are important coenzymes in energy metabolism (see p. 106). All of these nucleoside phosphates are classified as **nucleotides**.

In nucleosides and nucleotides, the pentose residues are present in the furanose form (see p. 34). The sugars and bases are linked by an *N*-**glycosidic bond** between the C-1 of the sugar and either the N-9 of the purine ring or N-1 of the pyrimidine ring. This bond always adopts the β-configuration.

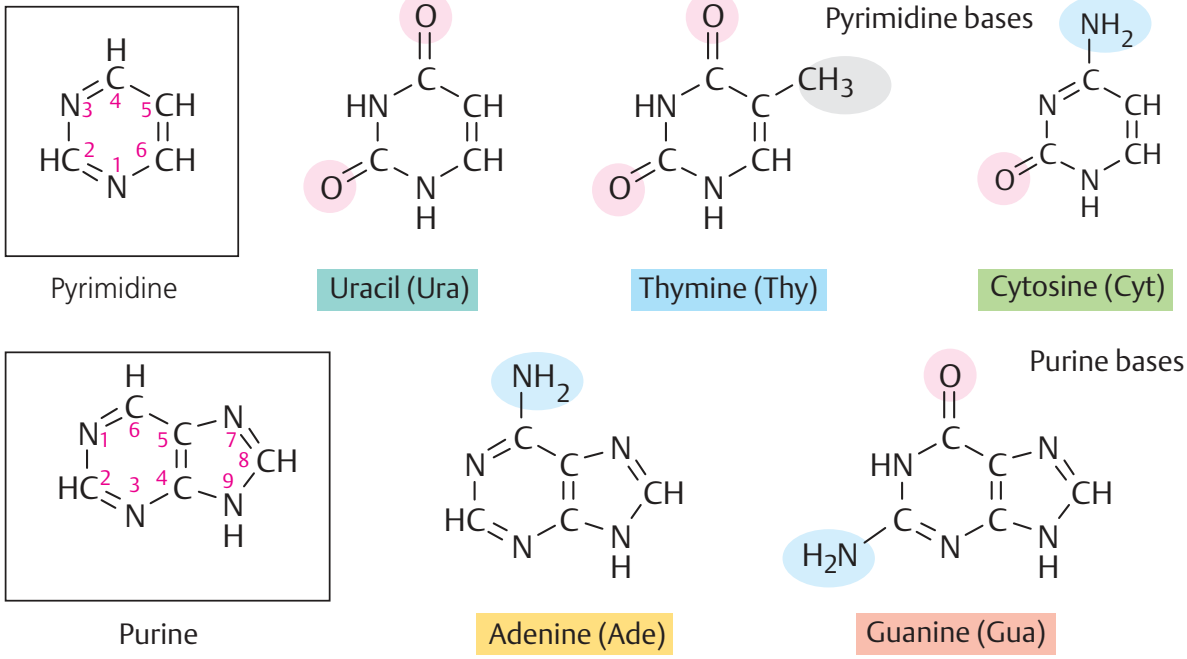
C. Oligonucleotides, polynucleotides ◑

Phosphoric acid molecules can form acid–anhydride bonds with each other. It is therefore possible for two nucleotides to be linked via the phosphate residues. This gives rise to *dinucleotides with a phosphoric acid–anhydride* structure. This group includes the coenzymes $NAD(P)^+$ and CoA, as well as the flavin derivative **FAD** (**1**; see p. 104).

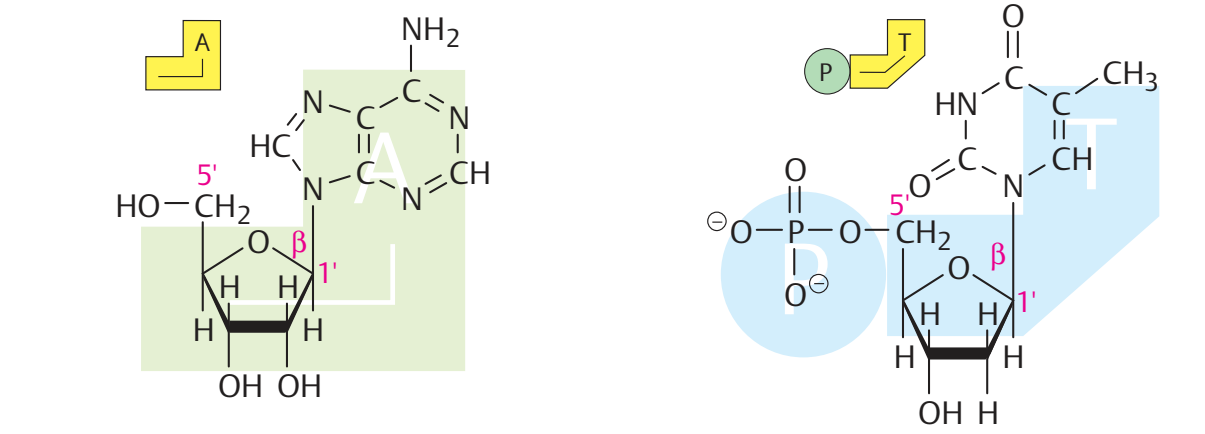
If the phosphate residue of a nucleotide reacts with the 3′-OH group of a second nucleotide, the result is a *dinucleotide with a phosphoric acid diester structure*. Dinucleotides of this type have a free phosphate residue at the *5′end* and a free OH group at the *3′ end*. They can therefore be extended with additional mononucleotides by adding further phosphoric acid diester bonds. This is the way in which **oligonucleotides**, and ultimately **polynucleotides**, are synthesized.

Polynucleotides consisting of ribonucleotide components are called **ribonucleic acid (RNA)**, while those consisting of deoxyribonucleotide monomers are called **deoxyribonucleic acid (DNA**; see p. 84). To describe the structure of polynucleotides, the abbreviations for the *nucleoside* components are written from left to right *in the 5′→3′* direction. The position of the phosphate residue is also sometimes indicated by a "p". In this way, the structure of the RNA segment shown Fig. **2** can be abbreviated as ..pUpG.. or simply as $..^{UG}..$.

A. Nucleic acid bases



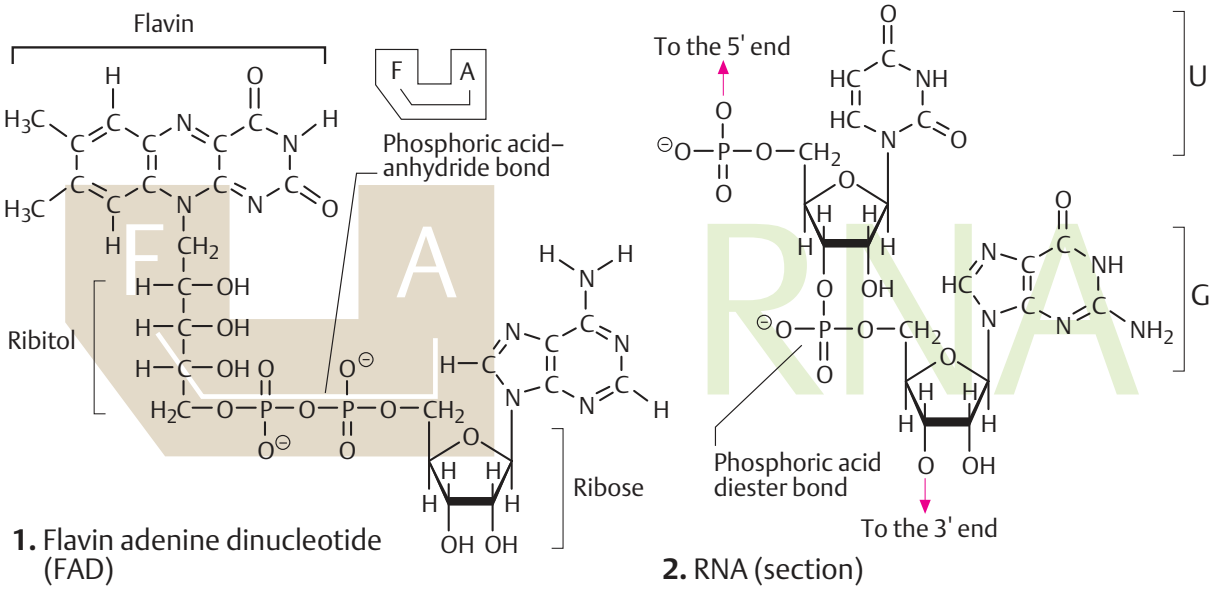
B. Nucleosides, nucleotides



1. Adenosine (Ado)

2. 2'-Deoxythymidine 5'-monophosphate (dtMP)

C. Oligonucleotides, polynucleotides



1. Flavin adenine dinucleotide (FAD)

2. RNA (section)

RNA

Ribonucleic acids (RNAs) are polymers consisting of nucleoside phosphate components that are linked by phosphoric acid diester bonds (see p. 80). The bases the contain are mainly uracil, cytosine, adenine, and guanine, but many unusual and modified bases are also found in RNAs (**B**).

A. Ribonucleic acids (RNAs) ◑

RNAs are involved in all the individual steps of gene expression and protein biosynthesis (see pp. 242–253). The properties of the most important forms of RNA are summarized in the table. The schematic diagram also gives an idea of the secondary structure of these molecules.

In contrast to DNA, RNAs do not form extended double helices. In RNAs, the base pairs (see p. 84) usually only extend over a few residues. For this reason, substructures often arise that have a finger shape or clover-leaf shape in two-dimensional representations. In these, the paired stem regions are linked by loops. Large RNAs such as ribosomal 16S-rRNA (center) contain numerous "stem and loop" regions of this type. These sections are again folded three-dimensionally—i. e., like proteins, RNAs have a tertiary structure (see p. 86). However, tertiary structures are only known of small RNAs, mainly tRNAs. The diagrams in Fig. **B** and on p. 86 show that the "clover-leaf" structure is not recognizable in a three-dimensional representation.

Cellular RNAs vary widely in their size, structure, and lifespan. The great majority of them are ribosomal RNA **(rRNA)**, which in several forms is a structural and functional component of *ribosomes* (see p. 250). Ribosomal RNA is produced from DNA by transcription in the nucleolus, and it is processed there and assembled with proteins to form ribosome subunits (see pp. 208, 242). The bacterial 16S-rRNA shown in Fig. **A**, with 1542 nucleotides (nt), is a component of the small ribosomae subunit, while the much smaller 5S-rRNA (118 nt) is located in the large subunit.

Messenger RNAs **(mRNAs)** transfer genetic information from the cell nucleus to the cytoplasm. The primary transcripts are substantially modified while still in the nucleus (mRNA maturation; see p. 246). Since mRNAs have to be read codon by codon in the ribosome, they must not form a stable tertiary structure. This is ensured in part by the attachment of *RNA-binding proteins,* which prevent base pairing. Due to the varying amounts of information that they carry, the lengths of mRNAs also vary widely. Their lifespan is usually short, as they are quickly broken down after translation.

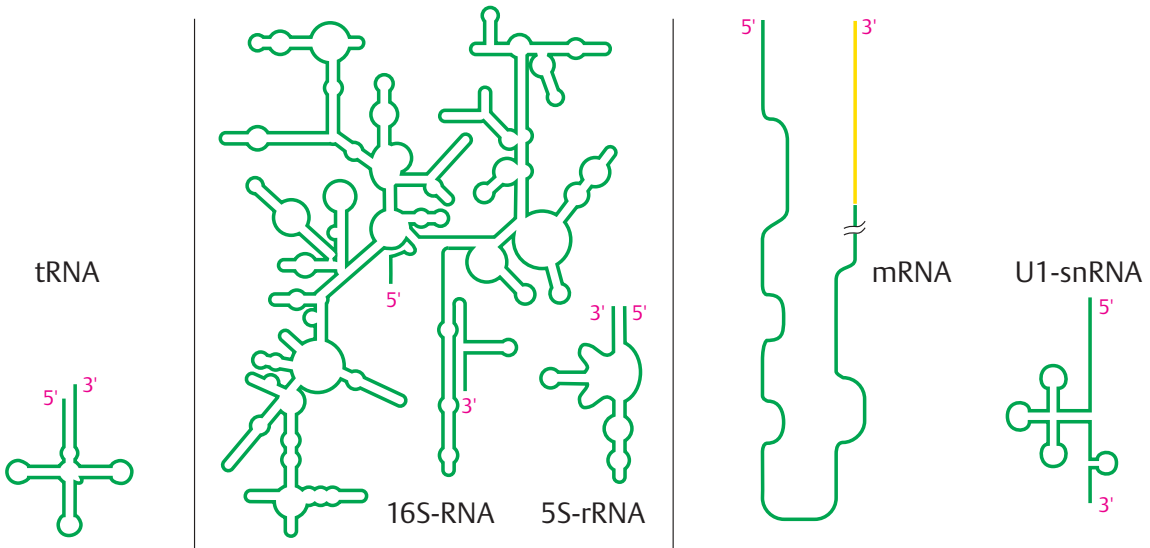
Small nuclear RNAs **(snRNAs)** are involved in the splicing of mRNA precursors (see p. 246). They associate with numerous proteins to form "spliceosomes."

B. Transfer RNA ($tRNA^{Phe}$) ◑

The transfer RNAs **(tRNAs)** function during translation (see p. 250) as links between the nucleic acids and proteins. They are small RNA molecules consisting of 70–90 nucleotides, which "recognize" specific mRNA codons with their *anticodons* through base pairing. At the same time, at their 3′ end (sequence .. CCA-3′) they carry the amino acid that is assigned to the relevant mRNA codon according to the genetic code (see p. 248).

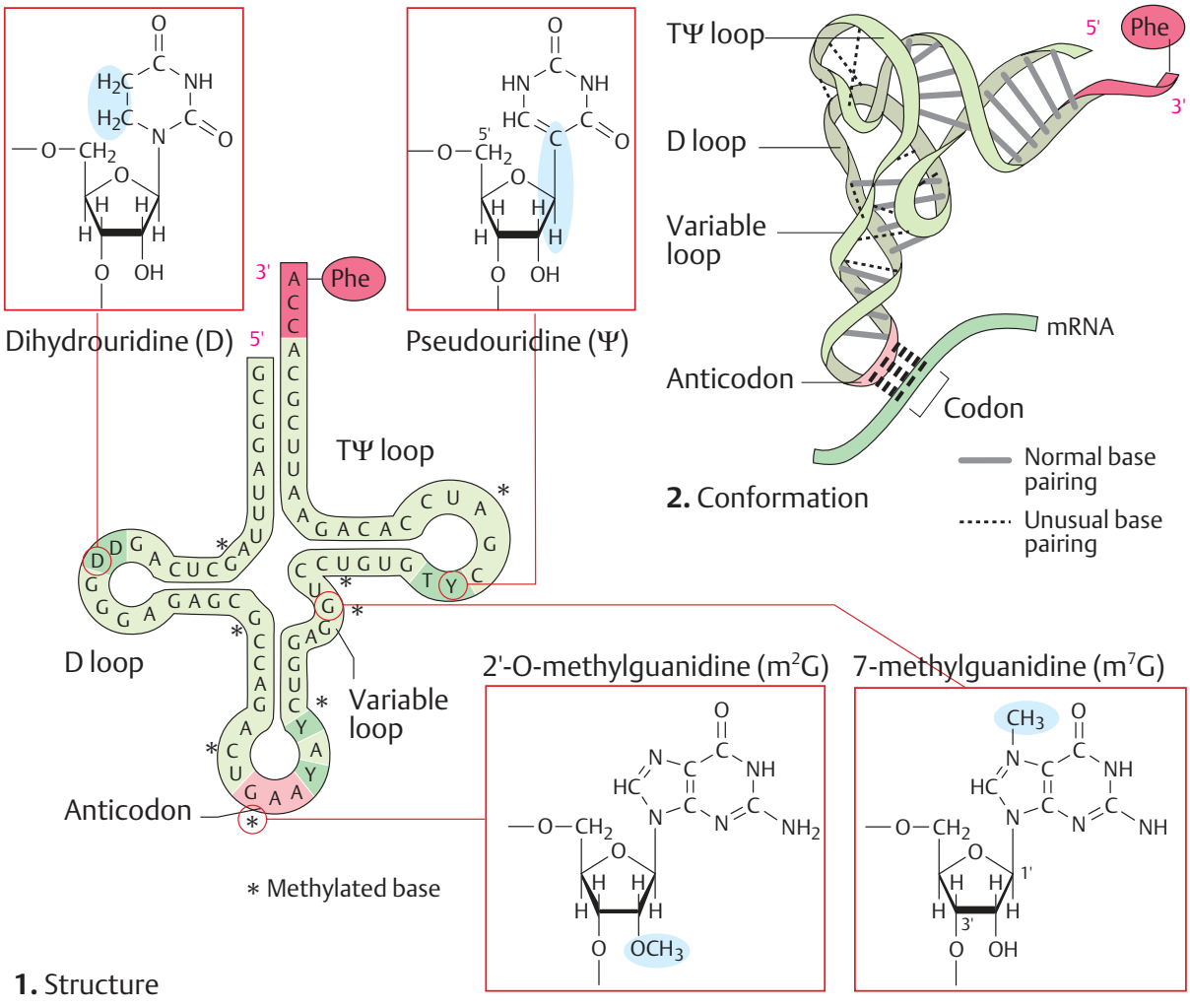
The base sequence and the tertiary structure of the yeast tRNA specific for phenylalanine ($tRNA^{Phe}$) is typical of all tRNAs. The molecule (see also p. 86) contains a high proportion of unusual and modified components (shaded in dark green in Fig. **1**). These include *pseudouridine* (Ψ), *dihydrouridine* (D), *thymidine* (T), which otherwise only occurs in DNA, and many methylated nucleotides such as *7-methylguanidine* (m^7G) and—in the anticodon—*2′-O-methylguanidine* (m^2G). Numerous base pairs, sometimes deviating from the usual pattern, stabilize the molecule's conformation (**2**).

A. Ribonucleic acids (RNAs)



tRNA	rRNA	Type	mRNA	snRNA
>50	4	Species per cell	> 1000	~ 10
74 - 95	120 - 5000	Length (b)	400 - 6000	100 - 300
10-20%	80%	Proportion	5%	< 1%
Long	Long	Lifespan	Short	Long
Translation	Translation	Function	Translation	Splicing

B. Transfer RNA (tRNA^{Phe})



DNA

A. DNA: structure ●

Like RNAs (see p. 82), deoxyribonucleic acids (DNAs) are polymeric molecules consisting of nucleotide building blocks. Instead of ribose, however, DNA contains 2′-deoxyribose, and the *uracil* base in RNA is replaced by *thymine*. The spatial structure of the two molecules also differs (see p. 86).

The first evidence of the special structure of DNA was the observation that the amounts of adenine and thymine are almost equal in every type of DNA. The same applies to guanine and cytosine. The model of DNA structure formulated in 1953 explains these *constant base ratios:* intact DNA consists of *two* polydeoxynucleotide molecules ("strands"). Each base in one strand is linked to a *complementary* base in the other strand by H-bonds. Adenine is complementary to thymine, and guanine is complementary to cytosine. One purine base and one pyrimidine base are thus involved in each **base pair**.

The complementarity of A with T and of G with C can be understood by considering the H bonds that are possible between the different bases. Potential donors (see p. 6) are amino groups (Ade, Cyt, Gua) and ring NH groups. Possible acceptors are carbonyl oxygen atoms (Thy, Cyt, Gua) and ring nitrogen atoms. *Two* linear and therefore highly stable bonds can thus be formed in A–T pairs, and *three* in G–C pairs.

Base pairings of this type are only possible, however, when the *polarity* of the two strands differs—i. e., when they run in opposite directions (see p. 80). In addition, the two strands have to be intertwined to form a **double helix**. Due to steric hindrance by the 2′-OH groups of the ribose residues, RNA is unable to form a double helix. The structure of RNA is therefore less regular than that of DNA (see p. 82).

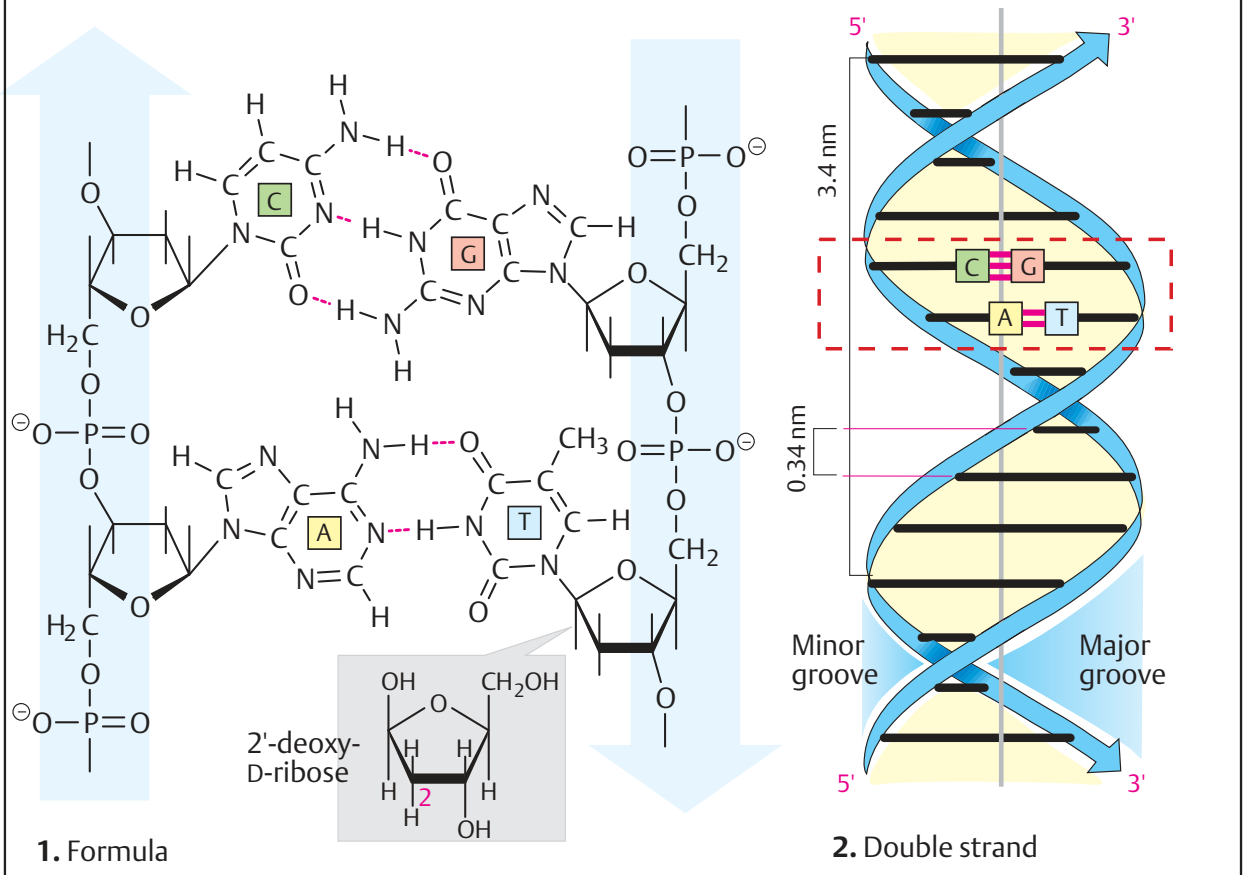
The conformation of DNA that predominates within the cell (known as **B-DNA**) is shown schematically in Fig. **A2** and as a van der Waals model in Fig. **B1**. In the schematic diagram (**A2**), the deoxyribose–phosphate "backbone" is shown as a ribbon. The bases (indicated by lines) are located on the inside of the **double helix**. This area of DNA is therefore apolar. By contrast, the molecule's surface is polar and negatively charged, due to the sugar and phosphate residues in the backbone. Along the whole length of the DNA molecule, there are two depressions—referred to as the "minor groove" and the "major groove"—that lie between the strands.

B. Coding of genetic information ●

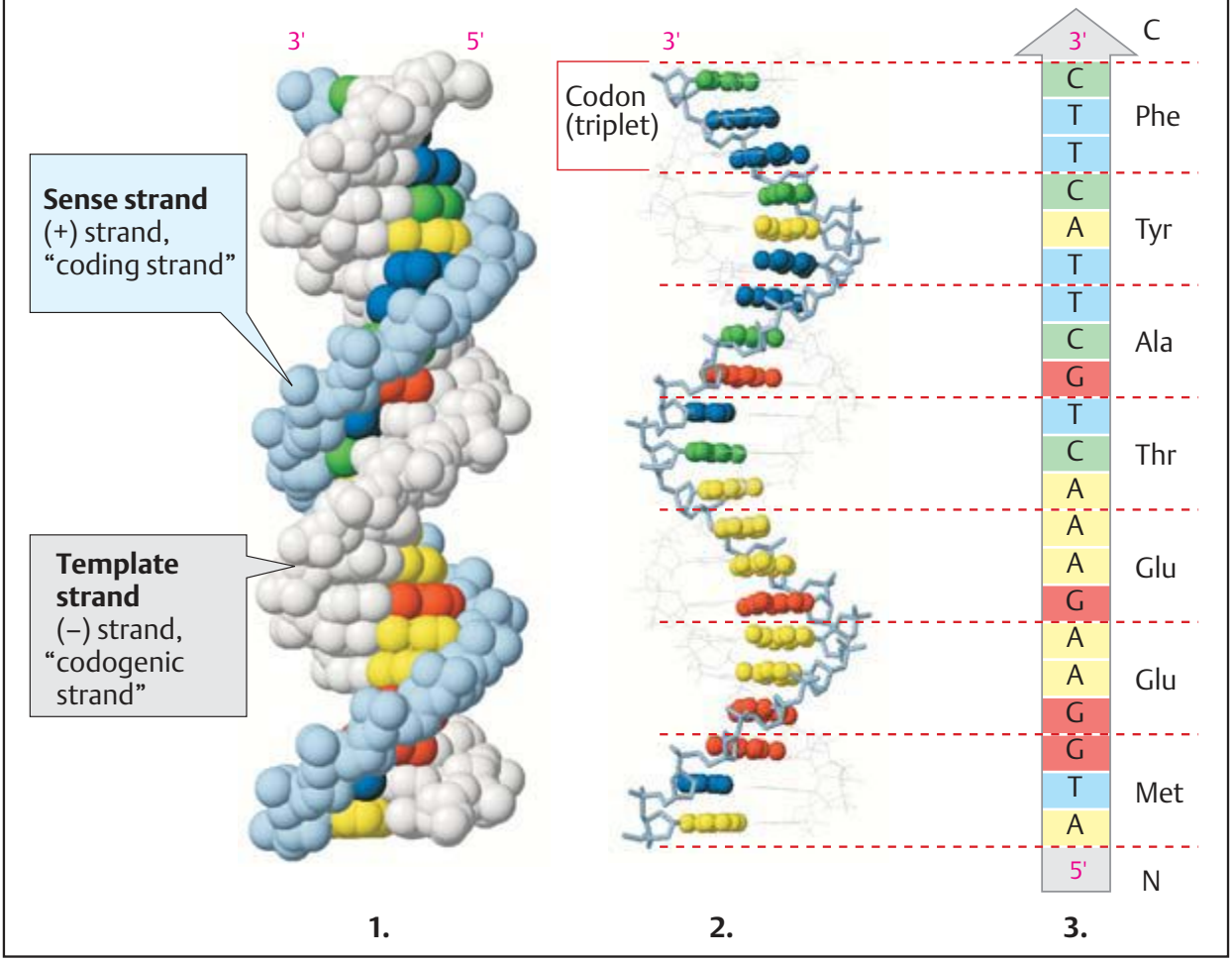
In all living cells, DNA serves to **store genetic information**. Specific segments of DNA ("genes") are transcribed as needed into RNAs, which either carry out structural or catalytic tasks themselves or provide the basis for synthesizing proteins (see p. 82). In the latter case, the DNA codes for the primary structure of proteins. The "language" used in this process has four letters (A, G, C, and T). All of the words ("codons") contain three letters ("triplets"), and each triplet stands for one of the 20 proteinogenic amino acids.

The two strands of DNA are not functionally equivalent. The **template strand** (the (–) strand or "codogenic strand," shown in light gray in Fig. **1**) is the one that is read during the synthesis of RNA (transcription; see p. 242). Its sequence is complementary to the RNA formed. The **sense strand** (the (+) strand or "coding strand," shown in color in Figs. **1** and **2**) has the *same sequence as the RNA,* except that T is exchanged for U. By convention, it is agreed that gene sequences are expressed by reading the sequence of the sense strand in the 5′→3′ direction. Using the genetic code (see p. 248), in this case the protein sequence (**3**) is obtained directly in the reading direction usual for proteins—i. e., from the *N* terminus to the *C* terminus.

A. DNA: structure



B. Coding of genetic information



Molecular models: DNA and RNA

The illustration opposite shows selected nucleic acid molecules. Fig. **A** shows various conformations of DNA, and Fig. **B** shows the spatial structures of two small RNA molecules. In both, the van der Waals models (see p. 6) are accompanied by ribbon diagrams that make the course of the chains clear. In all of the models, the polynucleotide "backbone" of the molecule is shown in a darker color, while the bases are lighter.

A. DNA: conformation ○

Investigations of synthetic DNA molecules have shown that DNA can adopt several different conformations. All of the DNA segments shown consist of 21 base pairs (bp) and have the same sequence.

By far the most common form is **B-DNA** (**2**). As discussed on p. 84, this consists of two antiparallel polydeoxynucleotide strands intertwined with one another to form a **right-handed double helix**. The "backbone" of these strands is formed by deoxyribose and phosphate residues linked by phosphoric acid diester bonds.

In the B conformation, the aromatic rings of the nucleobases are stacked at a distance of 0.34 nm almost at right angles to the axis of the helix. Each base is rotated relative to the preceding one by an angle of 35°. A complete turn of the double helix (360°) therefore contains around 10 base pairs (abbreviation: bp), i. e., the *pitch* of the helix is 3.4 nm. Between the backbones of the two individual strands there are two grooves with different widths. The *major groove* is visible at the top and bottom, while the narrower *minor groove* is seen in the middle. DNA-binding proteins and transcription factors (see pp. 118, 244) usually enter into interactions in the area of the major groove, with its more easily accessible bases.

In certain conditions, DNA can adopt the **A conformation** (**1**). In this arrangement, the double helix is still right-handed, but the bases are no longer arranged at right angles to the axis of the helix, as in the B form. As can be seen, the A conformation is more compact than the other two conformations. The minor groove almost completely disappears, and the major groove is narrower than in the B form. A-DNA arises when B-DNA is dehydrated. It probably does not occur in the cell.

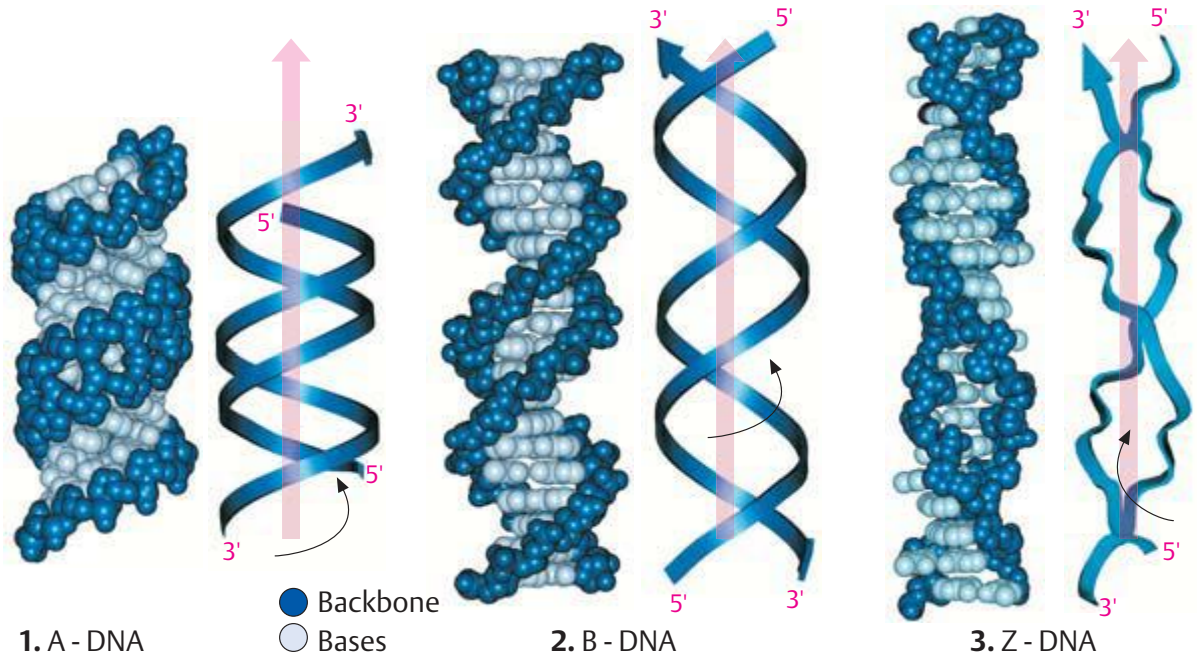
In the **Z-conformation** (**3**), which can occur within GC-rich regions of B-DNA, the organization of the nucleotides is completely different. In this case, the helix is *left-handed,* and the backbone adopts a characteristic *zig-zag* conformation (hence "Z-DNA"). The Z double helix has a smaller pitch than B-DNA. DNA segments in the Z conformation probably have physiological significance, but details are not yet known.

B. RNA ○

RNA molecules are unable to form extended double helices, and are therefore less highly ordered than DNA molecules. Nevertheless, they have defined secondary and tertiary structures, and a large proportion of the nucleotide components enter into base pairings with other nucleotides. The examples shown here are **5S-rRNA** (see p. 242), which occurs as a structural component in ribosomes, and a **tRNA** molecule from yeast (see p. 82) that is specific for phenylalanine.

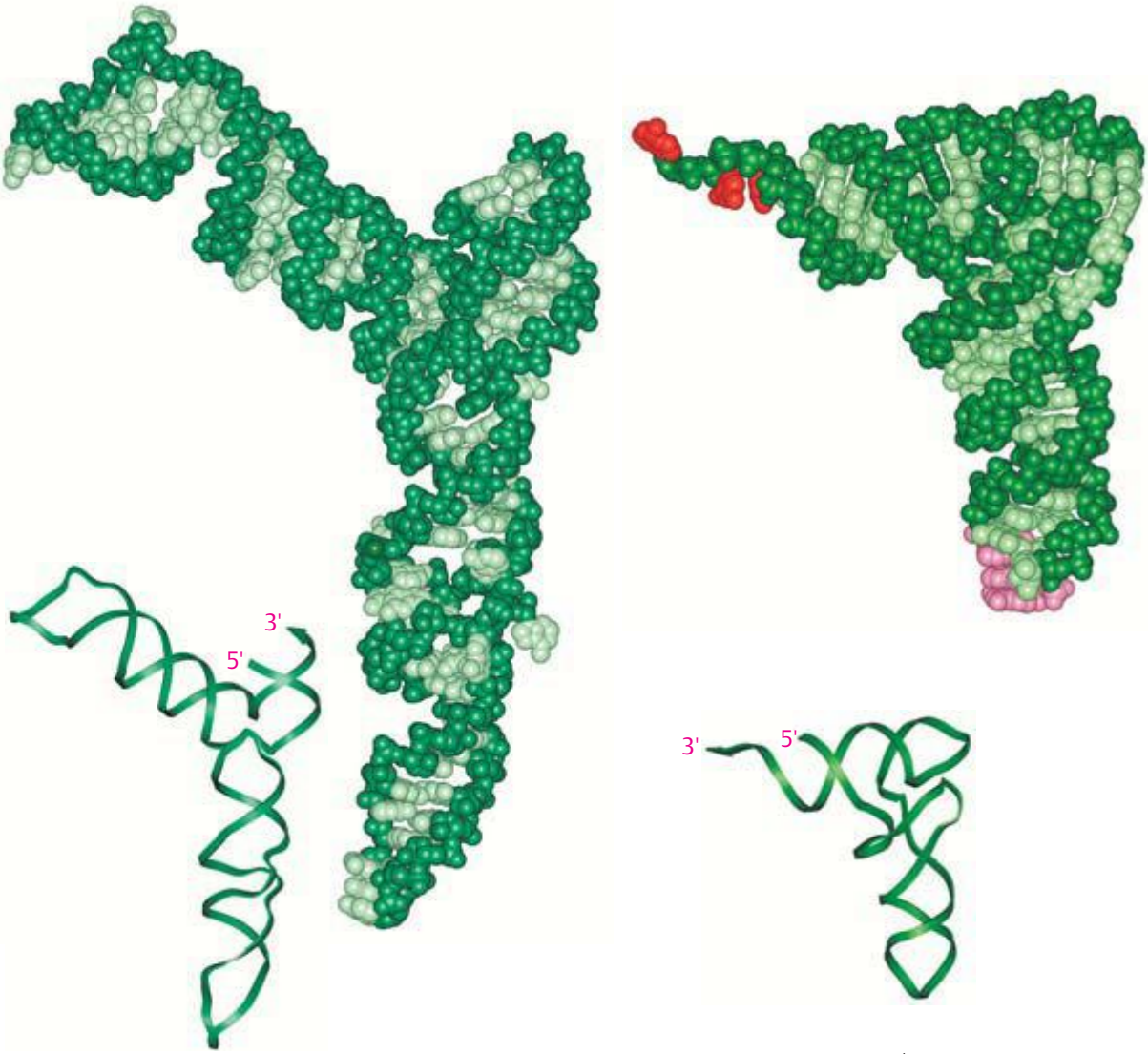
Both molecules are folded in such a way that the 3′ end and the 5′ end are close together. As in DNA, most of the bases are located in the inside of the structures, while the much more polar "backbone" is turned outwards. An exception to this is seen in the three bases of the *anticodon* of the tRNA (pink), which have to interact with mRNA and therefore lie on the surface of the molecule. The bases of the conserved CCA triplet at the 3′ end (red) also jut outward. During amino acid activation (see p. 248), they are recognized and bound by the ligases.

A. DNA: conformation



1. A - DNA **2.** B - DNA **3.** Z - DNA

B. RNA



1. 5S-rRNA (118 nucleotides)

2. Phe-tRNA^{Phe} (77 nucleotides)