1 Introduction to biomolecules

Bioanalytical chemistry relies on the identification and characterization of *particles* and *compounds*, particularly those involved with life and health processes. Living matter comprises certain key *elements*, and in mammals the most abundant of these, representing around 97% of dry weight of humans, are: carbon (C), nitrogen (N), oxygen (O), hydrogen (H), calcium (Ca), phosphorus (P) and sulfur (S). However, other elements such as sodium (Na), potassium (K), magnesium (Mg) and chlorine (Cl), although less abundant, nevertheless play a very significant role in organ function. In addition, miniscule amounts of so-called *trace elements*, including iron (Fe), play vital roles, regulating biochemical pathways and biological function. By definition, *biomolecules* are naturally occurring chemical compounds found in living organisms that are constructed from various combinations of key chemical elements. Not surprisingly there are fundamental similarities in the way organisms use such biomolecules to perform diverse tasks such as propagating the species and genetic information, and maintaining energy production and utilization. From this it is evident that much can be learned about the functionality of life processes in higher mammals through the study of micro-organisms and single cells. Indeed, the study of yeast and bacteria allowed genetic mapping before the Human Genome Project. This chapter provides an introduction to significant biomolecules of importance in the life and health sciences, covering their major properties and basic characteristics.

Learning Objectives

- To be aware of important chemical and physical characteristics of biomolecules and their components.

Understanding Bioanalytical Chemistry: Principles and applications Victor A. Gault and Neville H. McClenaghan

- To recognize different classifications of biomolecules.
- To understand and be able to demonstrate knowledge of key features and characteristics of major biomolecules.
- To identify and relate structure–function relationships of biomolecules.
- To illustrate and exemplify the impact of biomolecules in nature and science.

1.1 Overview of chemical and physical attributes of biomolecules

Atoms and elements

Chemical elements are constructed from *atoms*, which are small particles or units that retain the chemical properties of that particular element. Atoms comprise a number of different *sub-atomic particles*, primarily *electrons*, *protons* and *neutrons*. The *nucleus* of an atom contains positively charged protons and uncharged neutrons, and a cloud of negatively charged electrons surrounds this region. Electrons are particularly interesting as they allow atoms to interact (in bonding), and elements to become ions (through loss or gain of electrons). Further topics in atomic theory relevant to bioanalysis will be discussed throughout this book, and an overview of *atomic bonding* is given below.

Bonding

The physical processes underlying attractive interactions between atoms, elements and molecules are termed *chemical bonding*. Strong chemical bonds are associated with the sharing or transfer of electrons between bonding atoms, and such bonds hold biomolecules together. *Bond strength* depends on certain factors, and so-called *covalent bonds* and *ionic bonds* are generally categorized as ‘strong bonds’, while *hydrogen bonds* and *van der Waal’s forces of attraction* within molecules are examples of ‘weak bonds’. These terms are, however, quite subjective, as the strongest ‘weak bonds’ may well be stronger than the weakest ‘strong bonds’. Chemical bonds also help dictate the structure of matter. In essence, covalent bonding (electron sharing) relies on the fact that opposite forces attract, and negatively charged electrons *orbiting* one atomic nucleus may be attracted to the positively charged nucleus of a neighbouring atom. Ionic bonding involves *electrostatic attraction* between two neighbouring atoms, where one positively charged nucleus 'forces' the other to become negatively charged (through electron transfer) and, as opposites attract, they bond. Historically, bonding was first considered in the twelfth century, and in the eighteenth century English all-round scientist, Isaac Newton, proposed that a 'force' attached atoms. All bonds can be explained by quantum theory (in very large textbooks), encompassing the *octet rule* (where eight is the magic number when so-called *valence electrons* combine), the *valence shell electron pair repulsion theory* (where valence electrons repel each other in such a way as to determine geometrical shape), *valence bond theory* (including orbital hybridization and resonance) and *molecular orbital theory* (as electrons are found in discrete orbitals, the position of an electron will dictate whether or not, and how, it will participate in bonding). When considering bonding, some important terms are *bond length* (separation distance where molecule is most stable), *bond energy* (energy dependent on separation distance), *non-bonding electrons* (valence electrons that do not participate in bonding), *electronegativity* (measure of attraction of bound electrons in polar bonds, where the greater the difference in electronegativity, the more polar the bond). *Electron-dot structures* or *Lewis structures* (named after American chemist Gilbert N. Lewis) are helpful ways of conceptualizing simple atomic bonding involving electrons on outer valence shells (see Figure 1.1).

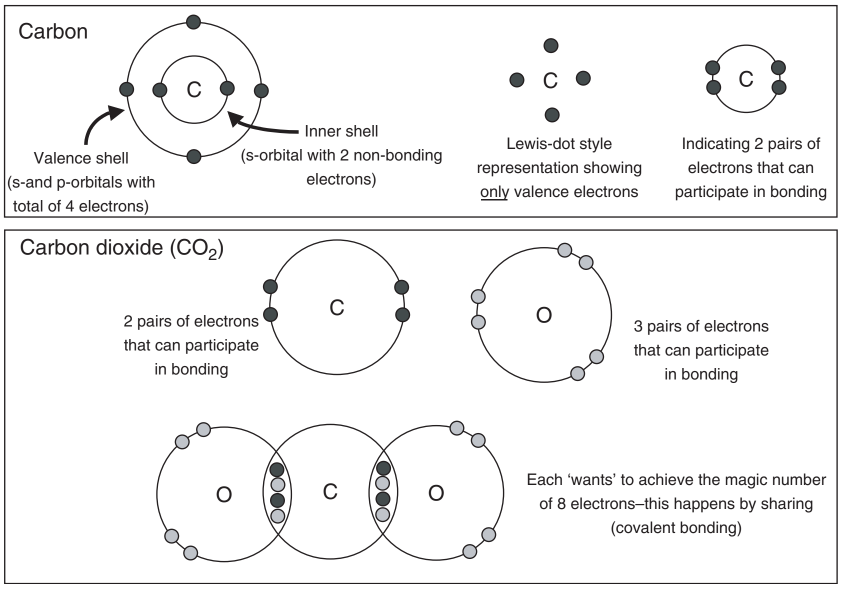


Figure 1.1 **Lewis structures illustrating covalent bonding in carbon dioxide.**

Phases of matter

Matter is loosely defined as anything having mass and taking up space, and is the basic building block of everything. There are three basic *phases of matter*, namely *gas*, *liquid* and *solid*, with different physical and chemical properties. Matter is maintained in these phases by pressure and temperature, and as conditions change matter can change from one phase to another, for example, solid ice converts to liquid water with rise in temperature. These changes are referred to as *phase transitions* inherently requiring energy, following the *Laws of Thermodynamics*. When referring to matter, the word *states* is sometimes used interchangeably with that of *phases*, which can cause confusion as, for example, gases may be in different thermodynamic states but the same state of matter. This has led to a decrease in the popularity of the traditional term *state of matter*. While the general term *thermodynamics* refers to the effects of heat, pressure and volume on physical systems, chemical thermodynamics studies the relationship of heat to chemical reactions or physical state following the basic Laws of Thermodynamics. Importantly, as energy can neither be created nor destroyed, but rather exchanged or emitted (for example as heat) or stored (for example in chemical bonds), this helps define the physical state of matter.

Physical and chemical properties

Matter comprising biomolecules has distinct physical and chemical properties, which can be measured or observed. However, it is important to note that physical properties are distinct from chemical properties. Whereas physical properties can be directly observed without the need for a change in the chemical composition, the study of chemical properties actually requires a change in chemical composition, which results from so-called chemical reactions. Chemical reactions encompass processes that involve the rearrangement, removal, replacement or addition of atoms to produce a new substance(s). Properties of matter may be dependent (*extensive*) or independent (*intensive*) on the quantity of a substance, for example mass and volume are extensive properties of a substance.

Studying physical and chemical properties of biomolecules

A diverse range of bioanalytical techniques have been used to study the basic composition and characteristics of biomolecules. Typically these techniques focus on measures of distinct physical and/or chemical attributes, to identify and determine the presence of different biomolecules in biological samples. This has been important from a diagnostic and scientific standpoint, and some of the major technologies are described in this book. Examples of physical and chemical properties and primary methods used to study that particular property are as follows:

Physical properties: Charge (see ion-exchange chromatography; Chapter 7); Density (see centrifugation; Chapter 6); Mass (see mass spectrometry; Chapter 9); and Shape (see spectroscopy; Chapter 5).

Chemical properties: Bonding (see spectroscopy and electrophoresis; Chapters 5 and 8); Solubility (see precipitation and chromatography; Chapters 6 and 7); Structure (see spectroscopy; Chapter 5).

1.2 Classification of biomolecules

It is important to note that whilst biomolecules are also referred to by more generic terms such as molecules, chemical compounds, substances, and the like, not all molecules, chemical compounds and substances are actually biomolecules. As noted earlier, the term *biomolecule* is used exclusively to describe naturally occurring chemical compounds found in living organisms, virtually all of which contain carbon. The study of carbon-containing molecules is a specific discipline within chemistry called *organic chemistry*. Organic chemistry involves the study of attributes and reactions of chemical compounds that primarily consist of carbon and hydrogen, but may also contain other chemical elements. Importantly, the field of organic chemistry emerged with the misconception by nineteenth century chemists that all organic molecules were related to life processes and that a 'vital force' was necessary to make such molecules. This archaic way of thinking was blown out of the water when organic molecules such as soaps (Michel Chevreul, 1816) and urea (Friedrich Wöhler, 1828) were created in the laboratory without this magical 'vital force'. However, despite being one of the greatest thinkers in the field of chemistry, the German chemist Wöhler was pretty smart not to make too much out of his work, even though it obviously obliterated the vital force concept and the doctrine of vitalism. So from this it is important to remember that not all organic molecules are biomolecules.

Life processes also depend on *inorganic molecules*, and a classic example includes the so-called 'transition metals', key to the function of many molecules (e.g. enzymes). As such, when considering biomolecules it is imperative to understand fundamental features of transition metals and their interaction with biomolecules. Indeed, transition metal chemistry is an effective means of learning basic aspects of inorganic chemistry, its interface with organic chemistry, and how these two fields of study impact on health and disease, and a whole chapter of this book is devoted to this important subject (Chapter 3). There are very many ways of classifying molecules and biomolecules, which often causes some confusion. The simplest division of biomolecules is on the basis of their size, that is, small (*micromolecules*) or large (*macromolecules*). However, while the umbrella term *macromolecule* is widely used, smaller molecules are most often referred to by their actual names (e.g. amino acid) or the more popular term *small molecule*. Yet even the subjective term *macromolecule* and its use are very confused. Historically, this term was coined in the early 1900s by the German chemist Hermann Staudinger, who in 1953 was awarded a Nobel Prize in Chemistry for his work on the characterization of polymers. Given this, the word macromolecule is often used interchangeably with the word *polymer* (or polymer molecule). For the purposes of this book the authors will use the following three categories to classify biomolecules:

Small molecules: The term *small molecule* refers to a diverse range of substances including: lipids and derivatives; vitamins; hormones and neurotransmitters; and carbohydrates.

Monomers: The term *monomer* refers to compounds which act as building blocks to construct larger molecules called *polymers* and includes: amino acids; nucleotides; and monosaccharides.

Polymers: Constructed of repeating linked structural units or monomers, polymers (derived from the Greek words *polys* meaning many and *meros* meaning parts) include: peptides/oligopeptides/polypetides/proteins; nucleic acids; and oligosaccharides/polysaccharides.

1.3 Features and characteristics of major biomolecules

Differences in the properties of biomolecules are dictated by their components, design and construction, giving the inherent key features and characteristics of each biomolecule that enable its specific function(s). There are a number of classes of more abundant biomolecules that participate in life processes and are the subject of study by bioanalytical chemists using a plethora of fundamental and state-of-the-art technologies in order to increase knowledge and understanding at the forefront of life and health sciences. Before considering important biomolecules it is first necessary to examine their key components and construction.

Building biomolecules

Biomolecules primarily consist of carbon (C) and hydrogen (H) as well as oxygen (O), nitrogen (N), phosphorus (P) and sulfur (S), but also have other chemical components (including trace elements such as iron). For now, focus will be placed on the core components carbon, hydrogen, and oxygen, and simple combinations (see also Table 1.1).

Carbon: The basis of the chemistry of all life centres on carbon and carbon-containing biomolecules, and it is the same carbon that comprises coal and diamonds that forms the basis of amino acids and other biomolecules. In other words, carbon is carbon is carbon, irrespective of the product material, which may be hard (diamond) or soft (graphite). Carbon is a versatile constituent with a great affinity for bonding other atoms through *single bonds* or *multiple bonds*, adding to complexity and forming around 10 million different compounds (Figure 1.2). As chemical elements very rarely convert into other elements, the amount of carbon on Earth remains almost totally constant, and thus life processes that use carbon must obtain it somewhere and get rid of it somehow. The flow of carbon in the environment is termed the *carbon cycle*, and the most simple relevant example lies in the fact that plants utilize (or recycle) the gas carbon dioxide (CO_2), in a process called *carbon respiration*, to grow and develop. These plants may then be consumed by humans and with digestion and other processes there is the ultimate generation of CO_2, some of which is exhaled and available again for plants to take up, and so the cycle continues. Being crude, in essence humans and other animals act as vehicles for carbon cycling, being *d*esigned for life in the womb, *d*evouring food and fluids, *d*eveloping, *d*efecating, *d*ying and *d*ecaying, the ‘6 D’s of life’.

Hydrogen: This is the most abundant (and lightest) chemical element, which naturally forms a highly flammable, odourless and colourless diatomic gas (H_2). The Swiss scientist Paracelsus, who pioneered the use of chemicals and minerals in medical practice, is the first credited with making hydrogen gas by mixing metals with strong acids. At the time Paracelsus didn’t know this gas was a new chemical element, an intuition attributed to British scientist Henry Cavendish, who described hydrogen gas in 1766 as ‘inflammable air’, later named by French nobleman and aspiring scientist, Antoine-Laurent Lavoisier, who co-discovered, recognized and named hydrogen (and oxygen), and invented the first *Periodic Table*.

Gaseous hydrogen can be burned (producing by-product water) and thus historically was used as a fuel. For obvious safety reasons helium (He), rather than

Table 1.1 Examples of simple combinations of carbon, hydrogen and oxygen

Compound	Chemical formula	Notes
Acetaldehyde	CH_3CHO (MeCHO)	Flammable liquid, fruity smell, found in ripe fruit, and metabolic product of plant metabolism. Chemical associated with the 'hangover' following overindulgence in alcohol.
Acetic acid	CH_3COOH	Hygroscopic liquid, gives vinegar its characteristic taste and smell. Corrosive weak acid.
Acetylene (ethyne)	C_2H_2	Gas containing C to C triple bond. Unsaturated chemical compound which can volatilize carbon in radiocarbon dating.
Benzyl acetate	$C_6H_5CH_2OCOCH_3$	Solid, sweet smelling ester, found naturally in many flowers (e.g. jasmine). Used in perfumes, cosmetics and flavourings.
Carbon dioxide	CO_2	Colourless, odourless and potentially toxic gas which can also exist in solid state (dry ice). Important component of the carbon cycle, a 'greenhouse gas', and contributes to the 'carbon footprint'.
Carbon monoxide	CO	Colourless, odourless and extremely toxic gas, produced by incomplete combustion of carbon-containing compounds (e.g. in internal combustion engines–exhaust fumes).
Ethanol	C_2H_5OH	Flammable, colourless, slightly toxic liquid, found in alcoholic beverages.
Methane	CH_4	Simplest alkane. Gaseous and principal component of natural gas. When burned in O_2 produces CO_2 and H_2O.
Water	H_2O	Normally odourless, colourless and tasteless liquid, but can also exist in solid (ice) or gas (water vapour) states. Non-inert common universal solvent.

hydrogen, was the gas of choice for floatation of Zeppelin airships. Indeed, the now famous Zeppelin airship 'The Hindenburg' was to be filled with He, but because of a US military embargo, the Germans modified the design of the airship to use flammable H_2 gas; an accident waiting to happen, and the rest is history.

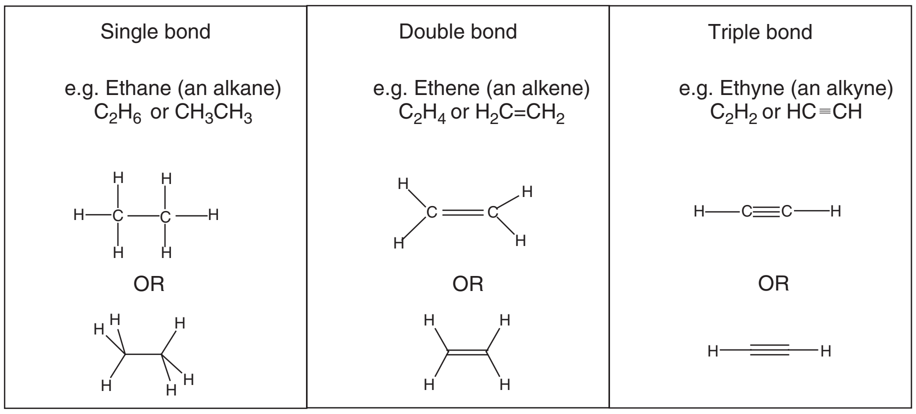


Figure 1.2 Illustration of carbon single, double and triple bonds.

In terms of biomolecules, hydrogen atoms usually outnumber both carbon and oxygen atoms.

Oxygen: As Lavoisier first generated oxygen from acidic reactions, he falsely believed that it was a component of all acids, deriving the name from the Greek words *oxys* (acid) and *genēs* (forming). Oxygen is usually bonded covalently or ionically to other elements such as carbon and hydrogen, and dioxygen gas (O_2) is a major component of air. Plants produce O_2 during the process of photosynthesis, and all species relying on *aerobic respiration* inherently depend on it for survival. Oxygen also forms a triatomic form (O_3) called *ozone* in the upper layers of the Earth's atmosphere, famously shielding us from UV radiation emitted from the

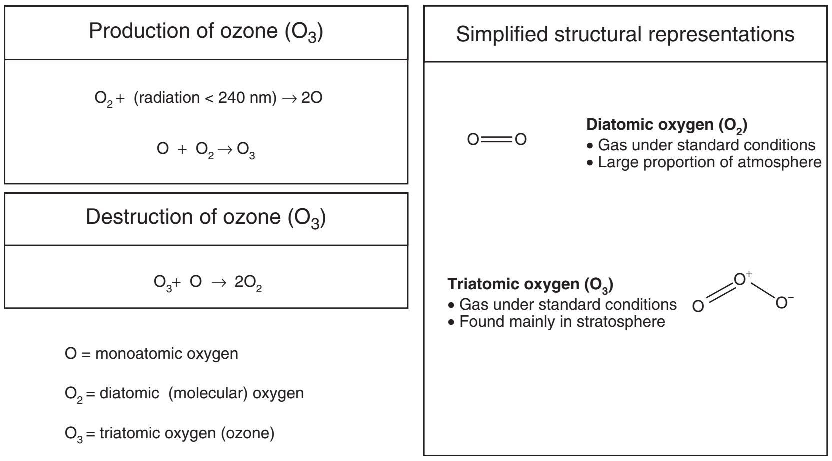


Figure 1.3 Chemical reactions involved in the production and destruction of ozone.

Sun (Figure 1.3). From a physiological and biochemical perspective, oxygen is both friend and foe; without it vital metabolic processes stop (friend) but exposure to oxygen in the form of certain oxygen-containing species (e.g. free radicals such as singlet oxygen) can be harmful (foe), and in extreme cases toxic, to body tissues, by exerting damaging actions on biomolecules regulating cellular and functional integrity.

Constructing complex biomolecules

As indicated above, C, H, O and other elements (such as N or P) can bind in a range of combinations to make simple compounds such as those given in Table 1.1. However, the same elements can also bind together to form much more complex structural and functional compounds (or biomolecules) which play vital roles in physiological processes. Major classes of these complex biomolecules are outlined in the boxes below.

Nucleotides

- Nucleotides consist of three components: a heterocyclic nitrogenous base, a sugar, and one or more phosphate groups.
- Nitrogenous bases of nucleotides are derivatives of either purine (adenine, A; or guanine, G) or pyrimidine (cytosine, C; thymine, T; or uracil, U) (see Figure 1.4).
- Nucleotides may be termed *ribonucleotides* (where component sugar is ribose) or *deoxyribonucleotides* (where component sugar is 2-deoxyribose).
- The bases bind to the sugar through glycosidic linkages.
- Also, one or more phosphate groups can bind to either the third carbon (C3) of the sugar of the nucleotide (so-called 3’ end) or the fifth carbon (C5) of the sugar (so-called 5’ end).
- Nucleotides are structural units of deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and cofactors such as coenzyme A (CoA), flavin adenine dinucleotide (FAD), nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP), with important roles in energy transfer, metabolism and intracellular signalling.

- Notably, polynucleotides are acidic at physiological pH due to the phosphate group (PO_4^-); this negatively charged anion is important for bioanalysis.

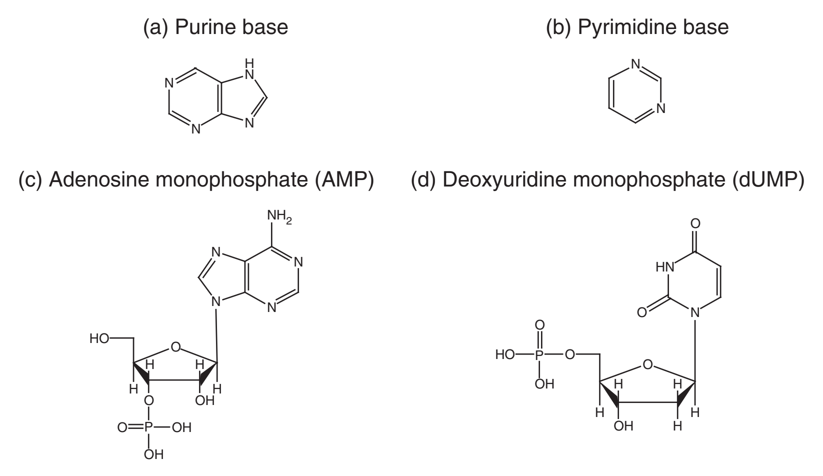


Figure 1.4 Diagrammatic representations of (a) a purine base, (b) a pyrimidine base, (c) a ribonucleotide, adenosine monophosphate (AMP) and (d) a deoxyribonucleotide, deoxyuridine monophosphate (dUMP).

Nucleic acids (e.g. RNA and DNA)

- Nucleic acids are polymers constructed from nucleotides (monomers) and found in cell nuclei.
- RNA comprises ribonucleotides while DNA contains deoxyribonucleotides.
- RNA can comprise the bases adenine (A), cytosine (C), guanine (G), and uracil (U).
- DNA can comprise the bases adenine (A), cytosine (C), guanine (G), and thymine (T).
- A nucleotide comprising a nucleic acid joins with another nucleotide through a so-called phosphodiester bond.
- Polymers of nucleic acids typically have different properties from individual units (nucleic acid monomers).
- There are also structural differences; RNA is usually single-stranded (alpha helix) while DNA is usually double-stranded (double helix). (Figure 1.5)

- DNA can replicate by separation of the two strands of the helix, which act as a template for synthesis of complementary strands.

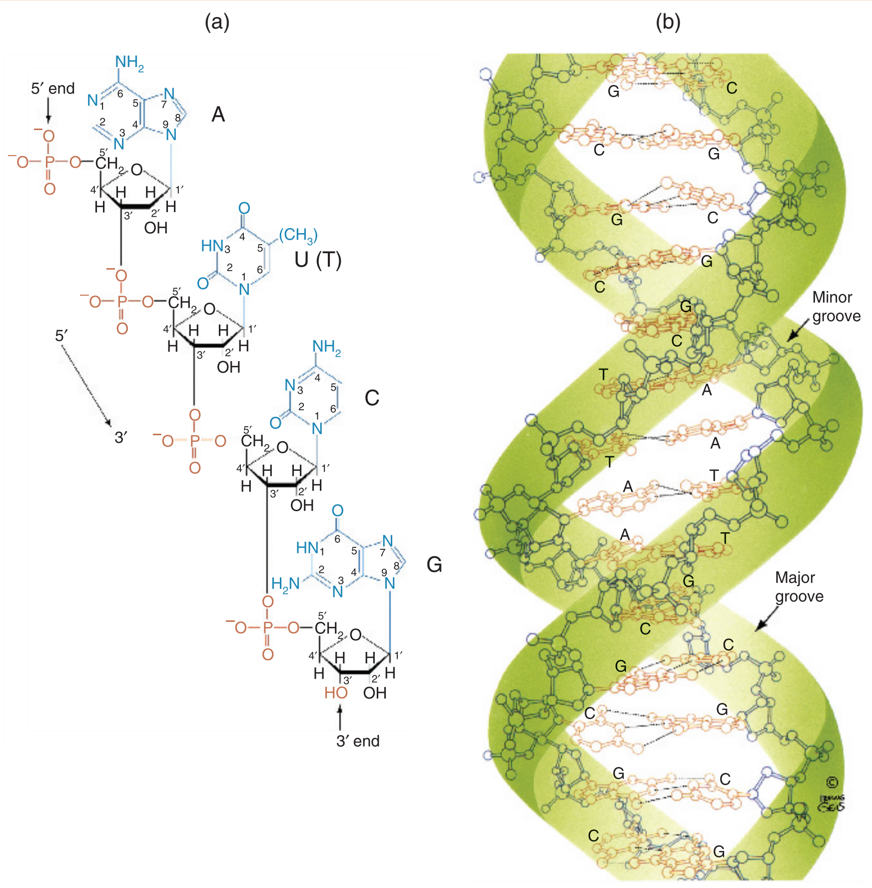
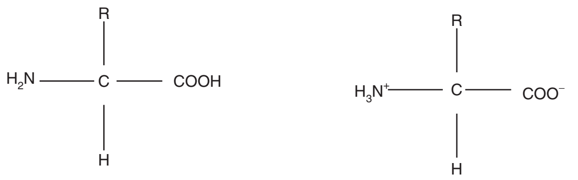


Figure 1.5 Diagrammatic representation of (a) a nucleic acid and (b) double helix structure of DNA. Illustrations, Irving Geiss. Rights owned by Howard Hughes Medical Institute. Reproduction by permission only.

Amino acids

- Molecules that contain a central carbon atom (alpha-carbon) attached to a carboxyl group (COOH), an amine group (NH_2), hydrogen atom (H), and a side chain (R group). (Figure 1.6)



The general structure of an α-amino acid – where each amino acid has a different R group

The structure of a zwitterion – where again each amino acid has a different R group

Figure 1.6 General representative chemical structure of an amino acid.

- The R group essentially defines the structure and function of an amino acid; these can generally be classified under three main groups: non-polar, uncharged polar, or charged polar amino acids.
- Amino acids (exceptions include Gly and Cys) are so-called chiral molecules (four different groups attached to alpha-carbon), which means they can exist as two different optical isomers called D (e.g. D-Ala) or more abundant L (e.g. L-Ala).
- There are 20 standard proteinogenic amino acids, of which 10 are essential amino acids that cannot be synthesized in the body so must be derived from the diet.
- Essential amino acids include: Iso, Leu, Lys, Met, Phe, Thr, Trp, Val, Arg and His, where the last two, Arg and His, are only actually essential under certain conditions.
- Amino (NH_2) and carboxylic acid (COOH) groups of the amino acid can readily ionize (to NH_3^+ and COO^-) at certain pHs to form an acid or base.
- The pH at which an amino acid is not in its ionized form (i.e. bears no electric charge) is known as its *isoelectric point*.
- When amino acids contain both positive and negative charges and are electrically neutral they fulfil the criteria of being a zwitterion (dipolar ion), which are highly water-soluble.
- Amino acids can be polymerized to form chains through condensation reactions, joining together by so-called peptide bonds, and they are often referred to as the building blocks of peptides and proteins.

Table 1.2 Classification of essential amino acids

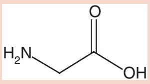
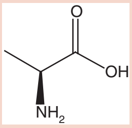
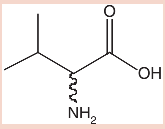
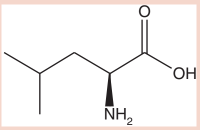
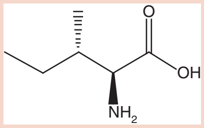
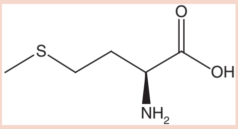
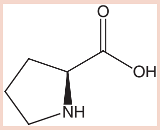
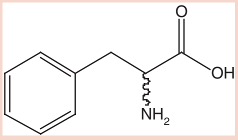
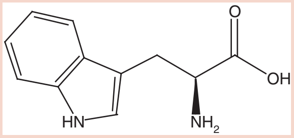
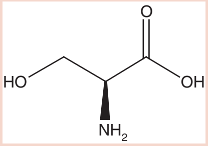
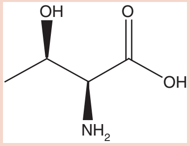
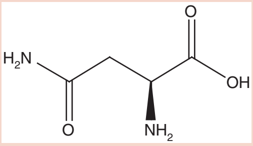
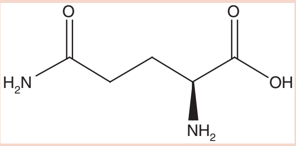
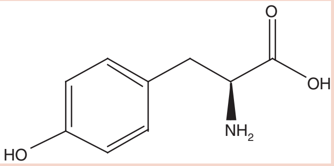
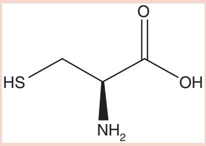
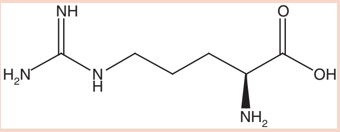
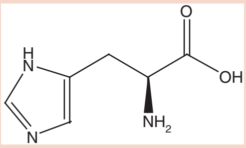
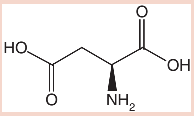
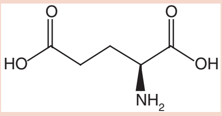
Classification	Name	3-letter code	1-letter code	Structure
Non-polar	Glycine	Gly	G	
	Alanine	Ala	A	
	Valine	Val	V	
	Leucine	Leu	L	
	Isoleucine	Ile	I	
	Methionine	Met	M	
	Proline	Pro	P	
	Phenylalanine	Phe	F	

Table 1.2 (*continued*)

Classification	Name	3-letter code	1-letter code	Structure
	Tryptophan	Trp	W	
Uncharged polar	Serine	Ser	S	
	Threonine	Thr	T	
	Asparagine	Asn	N	
	Glutamine	Gln	Q	
	Tyrosine	Tyr	Y	
	Cysteine	Cys	C	

(*continued overleaf*)

Table 1.2 (*continued*)

Classification	Name	3-letter code	1-letter code	Structure
Charged polar	Lysine	Lys	K	
	Arginine	Arg	R	
	Histidine	His	H	
	Aspartic acid	Asp	D	
	Glutamic acid	Glu	E	

Peptides and proteins

- Each peptide or protein is constructed as a string or chain of amino acids, creating huge numbers of variants, analogous to how letters of the alphabet make words.
- Importantly, peptides and proteins are assembled from amino acids on the basis of genetic coding (gene-nucleotide sequence), or can be synthesized in the laboratory.

- Peptides and proteins have individual and unique amino acid sequences (residues), giving them unique structural conformations and biological activity.
- A peptide is a short molecule formed by amino acids linked through amide (peptide) bonds. (Figure 1.7)

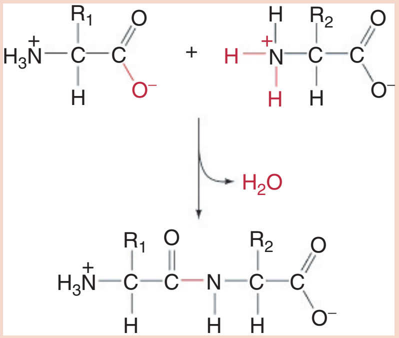


Figure 1.7 Formation of a peptide (amide) bond. From Voet, Voet & Pratt Fundamentals of Biochemistry, 2nd edn; © 2006 Voet, Voet & Pratt; reprinted with permission of John Wiley & Sons, Inc.

- When two amino acids join they form a dipeptide, three a tripeptide and so forth.
- Oligopeptides usually comprise between 3 and 10 amino acids and peptides with more than 10 are often referred to as *polypeptides*.
- Some peptides are called *peptide hormones* and as the name suggests these peptides have hormonal (endocrine) function.
- Proteins comprise one or more polypeptides and while peptides are short, polypeptide proteins are long.
- There are different ways of classifying proteins and one of the most important is on the basis of their structure.
- Biochemically there are four major classifications of protein structure: primary (amino acid sequence), secondary (local spatial arrangement), tertiary (overall 3D structure) and quaternary (protein complex structure). (Figure 1.8)

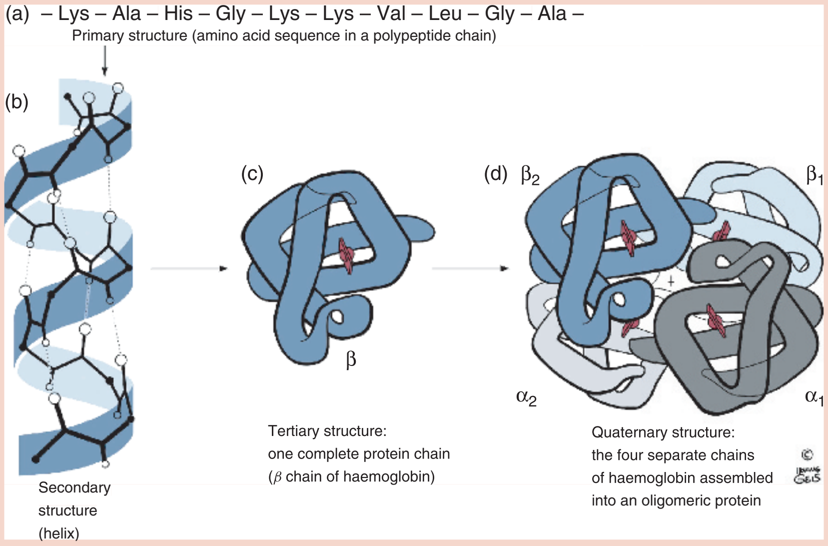
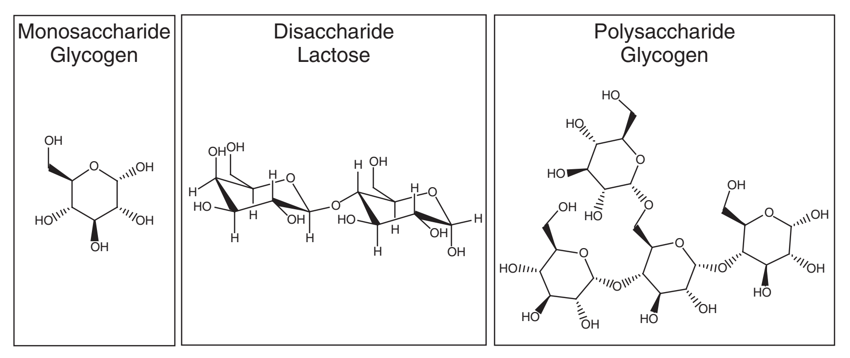


Figure 1.8 Diagrammatic representations of secondary, tertiary and quaternary protein structures. From Voet, Voet & Pratt Fundamentals of Biochemistry, 2nd edn; © 2006 Voet, Voet & Pratt; reprinted with permission of John Wiley & Sons, Inc.

Carbohydrates

- Simple, neutral biomolecules composed of C, H and O, often referred to as *saccharides*.
- All carbohydrates have an aldehyde (aldose) or ketone (ketose) functional group (containing C=O) and a hydroxyl group (–OH).
- Importantly, as carbohydrates contain aldehyde or ketone groups they undergo the same reactions as individual aldehyde or ketone molecules (e.g. oxidation and reduction reactions).
- Classification is based on the number of structural sugar units (and aldehyde or ketone group) in the chain, where 1 unit makes a monosaccharide (e.g. glucose), 2 units are disaccharides (e.g. lactose), 3–10 units are oligosaccharides (e.g. raffinose) and greater than 10 sugar units are polysaccharides (e.g. starch). (Figure 1.9)
- Sugar units are joined together through oxygen atoms, forming a so-called glycosidic bond.



Monosaccharide Glycogen

Disaccharide Lactose

Polysaccharide Glycogen

Figure 1.9 Examples of a monosaccharide, disaccharide and polysaccharide.

- Polysaccharides can reach many thousands of units in length, and carbohydrates can contain one or more modified monosaccharide units adding to complexity.
- Carbohydrates are abundant biomolecules in plants (produced by photosynthesis) and animals, with important roles in energy storage, and transport and structural components.
- Nutritionally, carbohydrates may generally be classified as simple sugars (e.g. glucose or fructose), polysaccharides (e.g. homoglycans such as starch or beta-glucans) or complex carbohydrates (e.g. glycoproteins).

Lipids

- Lipids have polar heads (are water-loving, hydrophilic) and long hydrocarbon tails (which are water-repelling, hydrophobic).
- Unlike other biomolecules, lipids are generally not made from repeating monomeric units.
- While the term *lipid* is often used interchangeably with the word fat, fat is actually a subgroup of lipids called *triglycerides* (triacylglycerols).
- Lipids are generally classified as: fatty acids (e.g. stearic acid), triglycerides (e.g. glycerol), phosphoglycerides (e.g. phosphatidylinositol), sphingolipids (e.g. ceramide), and steroids (e.g. cholesterol). (Figure 1.10)
- Lipids have important roles in energy storage, cell membrane structure and cell signalling.

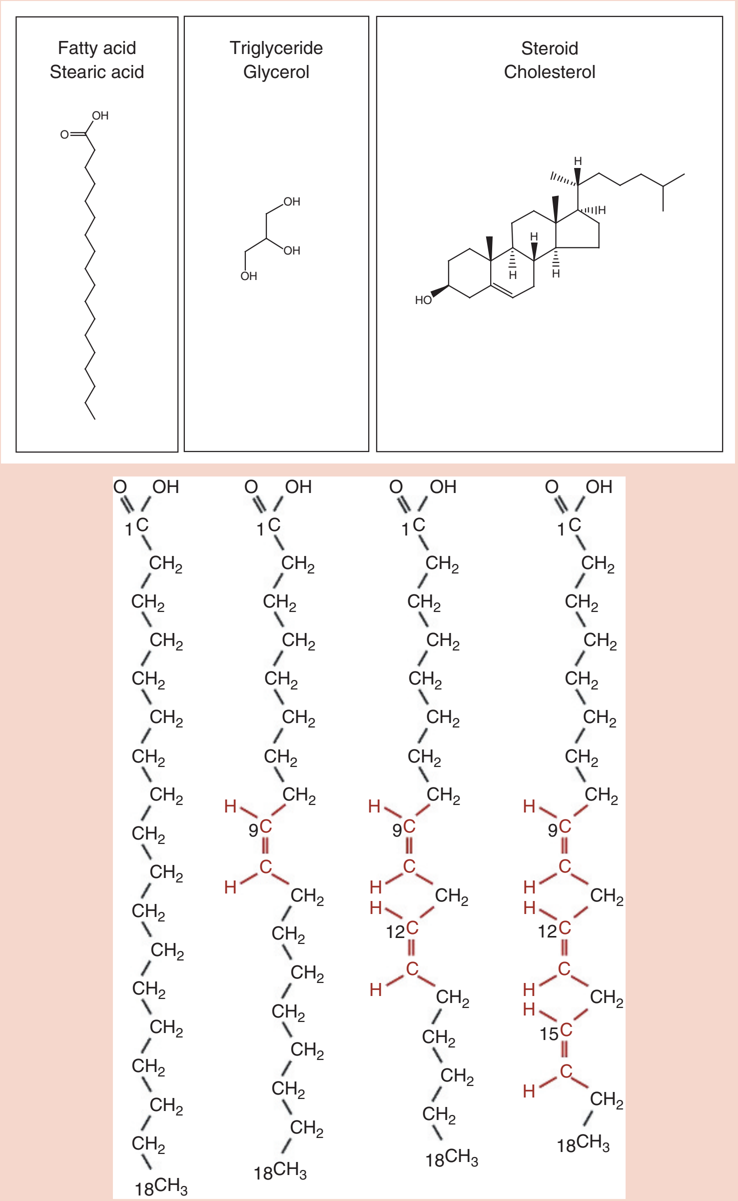


Figure 1.10 **Examples of a fatty acid, triglyceride and steroid. From Voet, Voet & Pratt Fundamentals of Biochemistry, 2nd edn; © 2006 Voet, Voet & Pratt; reprinted with permission of John Wiley & Sons, Inc.**

1.4 Structure–function relationships

The structure of a given biomolecule will confer certain functional attributes and, as such, is a key defining feature of that biomolecule. There are very many environmental influences that can impact on structure and/or function and, importantly, the stability of the biomolecule. Furthermore, this information is also important when considering bioanalysis, as the outcome of an analytical process or procedure, and indeed the stability of a biomolecule, is dependent on numerous physical and chemical factors including pH, temperature and solvent concentration/polarity. Changes in any or all of these parameters may result in general or specific structural and/or functional changes to a biomolecule that may be *reversible* (e.g. partial unfolding of a protein) or *irreversible* (protein denaturation and degradation). Simple visual examples include the reversible change to hair when it is straightened or curled ('hair perm') and irreversible protein denaturation of 'egg whites' (essentially egg albumins in water) that turn from a transparent liquid into an opaque white solid on cooking (temperature). On a more scientific vein, variations in pH can alter the ionization states of biomolecules such as amino acids in proteins, and phosphate groups in nucleotides. Alterations to *functional groups* can greatly alter the activity and properties of biomolecules, and this is why most physiological, biochemical and enzymatic processes require *homeostatic* conditions (i.e. maintenance of a relatively constant internal environment).

It is also important to map the chemical structure of biomolecules with certain well-defined processes, which may relate to biological activity or chemical reactivity. Medicinal chemists place particular importance on understanding *structure–activity relationships* (SARs) of biomolecules in order to bioengineer modified biomolecules with enhanced activity (*potency*), for example by changing amino acid composition of peptides or insertion/addition of chemical groups. This approach has given new and exciting insights into chemical groups that affect biological processes, and allowed complex mathematical modelling of *quantitative structure–function relationships* (QSARs). This has inherent difficulties, as certain features such as *post-translational modification* of proteins may depend on multiple factors, and thus not all related biomolecules have similar activities (so-called SAR paradox). Historically, one of the first simple examples of QSAR was to predict boiling points on the basis of the number of carbon atoms in organic compounds; more modern applications of QSAR are in drug design and discovery, discussed in more detail in Chapter 12.

1.5 Significance of biomolecules in nature and science

Biomolecules are the essence and currency of life and health processes lying at the heart of the simplest to the most complex system. Understanding the fundamental nature of biomolecules, their structure, location, behaviour and function, is critical to knowledge and understanding of health, the development of disease and appropriate therapeutic intervention. To this end, the ability to measure biomolecules in test samples and compare these with given ‘norms’, taken from healthy individuals (single cell to organism) in a population, is of paramount importance to the management of health and disease. When considering biomolecules it is thus important to include structural variants or anomalies that may arise either spontaneously or as a result of some interaction that can alter functionality. Using advanced bioanalytical tools (such as mass spectrometry, Chapter 9) it is possible to gain both qualitative and quantitative information on a given biomolecule or variant (synthetic or otherwise) which is of scientific and therapeutic importance. This is illustrated briefly below, considering what key classes of biomolecule normally do, what happens when things involving those biomolecules go wrong, and how understanding normal functionality and defects can give new insights into diseases and their treatment.

Nucleic acids (RNA and DNA)

There are various types of ribonucleic acid (RNA) molecule, and some confusion lies in the fact that not all RNA performs the characteristic general function of translating genetic information into proteins. Different RNA molecules have different biological functions: (i) *messenger ribonucleic acid* (mRNA) – carries information from deoxyribonucleic acid (DNA) to ribosomes (cellular protein synthetic ‘factories’); (ii) *transfer ribonucleic acid* (tRNA)–transfers specific amino acids to a growing polypeptide chain during protein synthesis (so-called *translation*); (iii) *ribosomal ribonucleic acid* (rRNA) – provides structural scaffolding within the ribosome and catalyses formation of peptide bonds; (iv) *non-coding RNA* (RNA genes) – genes encoding RNA that are not translated into protein; (v) *catalytic RNA*–which catalyses chemical reactions; (vi) *double-stranded ribonucleic acid* (dsRNA) – forms genetic material of some viruses; can initiate *ribonucleic acid interference* (RNAi) and is an intermediate step in *small interfering ribonucleic acid* (siRNA) formation; can induce gene expression at transcriptional level, where dsRNAs are referred to as *small activating RNA*. Problems with the functions of these different RNAs will obviously impact on processes critical to protein synthesis and while, at present, there is little that can be done to fix RNA-driven processes, the ability to detect such alterations is of diagnostic/therapeutic value. For example, as siRNA can knock down specific genes, it has proven experimentally useful in the study of gene function and their role in complex pathways, and also offers the exciting possibility of therapeutic silencing of specific genes mediating disease pathways.

Within DNA lies the genetic code (blueprint) of all living organisms that contains genetic instructions to make individual cells, tissues and organisms. DNA is organized within *chromosomes*, and a set of chromosomes in a cell makes up the cell's *genome*. Furthermore, DNA can replicate to make an identical copy, an important means of transferring genetic information into new cells. While *genes* may be defined in a number of ways, they are generally considered inheritable DNA sequences that both store and carry genetic information throughout the lifespan of an individual. The coding information of genes depends on the bases comprising the DNA, and the sequence of the four bases (i.e. A, T, G, and C) confers the *genetic code* that specifies the sequence of amino acids making up a particular protein within a cell. A process called *transcription* reads the genetic code, where the enzyme RNA polymerase allows transfer of genetic information from DNA into mRNA before the message is translated into protein (translation and protein synthesis). Given the importance of DNA, it is perhaps not surprising that cells inherently have a restricted ability to repair and protect DNA. However, the failure to correct DNA lesions can cause disease, and if *mutated DNA* is *heritable* then it may pass down to offspring. In humans, inherited mutations affecting DNA repair genes have been associated with cancer risk, for example the famous BRCA1 and BRCA2 (which stands for breast cancer 1 and 2, respectively) mutations. Notably, cancer therapy also primarily acts to overwhelm the capacity of cells to repair DNA damage, resulting in preferential death of the most rapidly growing cells, which include the target cancer cells.

Peptides and proteins

Peptides and proteins are often grouped into distinct families according to various criteria, such as structure and primary function. Given that peptides and proteins are major regulators of very many different biological processes, there is an incredibly wide and diverse range of peptides and proteins in nature. Indeed, some peptides/proteins not found in man may still have biological or medical applications in the regulation of human processes (e.g. cell signalling) and related therapeutic applications. For example, a peptide called *exendin* was originally isolated from the saliva of the large, slow moving, venomous lizard, the 'Gila monster' (*Heloderma suspectum*), found in Arizona and other parts of the United States/Mexico. The venom, secreted into the lizard's saliva, contained a rich 'cocktail' of different biological active molecules, including exendin, which was subsequently found to demonstrate antidiabetic properties. While some scientists were initially sceptical about commercial success of this peptide as a pharmaceutical product (under names exenatide or Byetta), it has proven a winner, with >$500 million in sales in its first year! This therapeutic is the first in a new class of medicines which is used to control blood glucose levels in human Type 2 diabetes, and indeed other peptides derived from the human gut peptides glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP) also hold great promise for the future treatment of the ‘diabetes epidemic’.

When considering how changes to protein structure can alter function and thus contribute to disease processes, focus should be directed to protein misfolding. This can occur for different reasons, but largely arises due to problems resulting from genetic mutations that can cause defective protein folding, incorrect assembly and processing. Indeed, incorrect folding is associated with defective cellular transport and/or loss of functional activity, which is the molecular basis of a number of diseases. For example, changes in secondary and tertiary protein structure can lead to neurodegenerative disorders. Alterations of so-called prion proteins are closely associated with Creutzfeld–Jakob disease (CJD) (and variant CJD) and transmissible spongiform encephalopathy (TSE). While these diseases have different origins they are related to each other and amyloidoses, as they involve an aberrant accumulation/deposition of proteins as amyloid fibrils or plaques. There are many other examples of diseases arising from protein folding defects in humans, including cystic fibrosis, cataracts, Tay–Sachs disease, Huntington’s chorea and familial hypercholesterolaemia, but of course such defects can affect many different species.

Carbohydrates

As noted earlier, there are many different ‘types’ of carbohydrate which may be grouped according to the number of structural sugar units, or indeed nutritionally. Typically carbohydrates are classified on the basis of the chemical nature of their carbonyl groups and the number of constituent carbon atoms. Carbohydrates represent major fuel sources for most species. However, in addition to being utilized for storage and transport of energy (e.g. starch, glycogen) they also make up structural components in plants (e.g. cellulose) and animals (e.g. chitin). Given such important roles, it is perhaps not surprising that there are a range of disorders associated with incorrect handling (including storage) and usage of carbohydrates, which include: lactose intolerance, glycogen storage disease, fructose intolerance, galactosaemia, pyruvate carboxylase deficiency (PCD), pyruvate dehydrogenase deficiency (PDHA), and pentosuria. This list is by no means exhaustive but it would be amiss not to mention diabetes mellitus, a metabolic disease that has been described as the ‘epidemic of the twenty-first century’. Insulin is an important regulator of whole body metabolism and in particular glucose control, where insulin depletion and/or impaired insulin sensitivity of body tissues has a major direct impact on blood glucose levels (glycaemia), often resulting in either hypoglycaemia (glucose too low) or hyperglycaemia (glucose too high). Both states are detrimental if left untreated, and both major forms of diabetes (i.e. Type 1

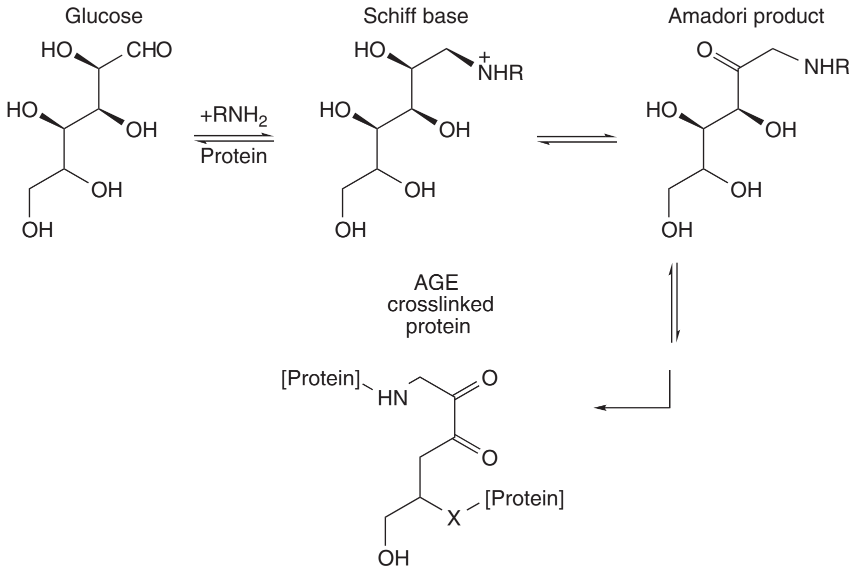


Figure 1.11 **Illustration of advanced glycation end-product (AGE) formation.**

and Type 2) are characterized by hyperglycaemia using various measured parameters including glycated haemoglobin (HbA$_{1c}$). Glycation is a natural process by which endogenous simple sugars (glucose, fructose, galactose) attach to other biomolecules (typically peptides/proteins) in the bloodstream and in tissues, and subsequently alter their biological activity and often also their elimination from the body. Formation of advanced glycation end-products (AGEs) has three principle steps, the first two of which are reversible; namely, Step 1: Schiff base formation; Step 2: Amadori product formation; and Step 3: Formation of AGE (see Figure 1.11). AGE formation can contribute to major pathologies, especially those associated with Type 2 diabetes and microvascular complications (such as retinopathy, where retinal vascular components are altered/damaged). Epidemiological studies in humans have revealed that tighter control of blood glucose levels (glycaemia) reduces the risk of diabetic complications including those associated with glycation, and that such complications are not an inevitable outcome of diseases such as Type 2 diabetes.

Lipids

Lipids are primarily categorized on the basis of the presence or absence of carbon–carbon double bonds in their carbon chain. When lipids contain double bonds they are described as *unsaturated*, and conversely those that have no double bonds are termed *saturated*. As a group, lipids are important when considering nutrition and health, with roles in energy storage, cell membrane structure and cell signalling. However, fats are 'energy dense', readily stored, and there is a large growing body of evidence that over-consumption of dietary fats can directly contribute to major pathologies. This is largely as a consequence of fat storage, which is influenced by both genetic factors (enhancing fat deposition) and environmental factors, particularly food choice and physical activity. Given this, while obesity is primarily associated with over-consumption of high-energy foods, this condition is not a single disorder but rather a heterogeneous group of conditions. The popularization of terms like *good fats* and *bad fats*, while questionable scientifically, has certainly increased public awareness of different dietary fat groups/components, with the aim of public health benefits. Control of dietary habits and increased physical activity can both prevent and alleviate weightiness/obesity and related major metabolic conditions such as Type 2 diabetes. Importantly, diseases such as Type 2 diabetes are as much about impaired fat metabolism, mobilization and handling, where high levels of circulating lipids combined with high glucose levels collectively result in cell and tissue damage (so-called glucolipotoxicity). Much attention has been directed to cholesterol (and hypercholesterolaemia) and trans fatty acids as important risk factors for heart disease, which itself is associated with obesity/being overweight. Given the emerging obesity epidemic (associated with Type 2 diabetes–diabesity), there is particular pressure on the research community and pharmaceutical industry to discover and develop new drugs to combat obesity. Recent efforts have been directed towards agents that enhance metabolism and in particular fat utilization, or alternatively target the suppression of appetite or regulation of feeding behaviour. These efforts are critically important to avoid the necessity for invasive surgical procedures such as gastric bypass to reverse or prevent worsening of obesity.

Key Points

- Biomolecules are naturally occurring chemical compounds found in living organisms that are constructed from various combinations of key chemical elements.
- Biomolecules can be broadly classified into three main categories: small molecules, monomers and polymers.

- Differences in the properties of biomolecules are dictated by their components, design and construction, giving the inherent key features and characteristics of each biomolecule which enables its specific function.
- Nucleotides consist of three key components: a heterocyclic nitrogenous base, a sugar and one or more phosphate groups.
- Nucleic acids are polymers constructed from nucleotides (monomers), where ribonucleic acid (RNA) comprises ribonucleotides, and deoxyribonucleic acid (DNA) contains deoxyribonucleotides.
- Amino acids are biomolecules that contain a central carbon atom (alpha-carbon) attached to a carboxyl group (COOH), an amine group (NH_2), hydrogen atom (H) and a side chain (R group).
- A peptide is a short molecule formed by amino acids linked through amide (peptide) bonds, while proteins comprise one or more polypeptides.
- Carbohydrates are simple neutral biomolecules composed of C, H, and O, and are classified on the basis of their number of structural sugar units and functional group (aldehyde or ketone).
- Lipids represent diverse classes of hydrocarbon-containing biomolecules, containing polar heads, with important roles in energy storage, cell membrane structure and cell signalling.
- The structure of a biomolecule will confer certain functional attributes which are a key defining feature of that biomolecule.