1.1 BONDS

Living organisms are made of up of organic molecules consisting mainly, though not exclusively, of carbon (C), hydrogen (H), oxygen (O), nitrogen (N), and sulphur (S). These are held together by bonds which vary in strength and length.

**Covalent bonds**

- **Covalent bonds** form when electrons are shared between atoms within a molecule.
- These are usually the most stable type of bond, although some will break spontaneously.
- e.g. C–C or C–H
- Can be single, double, or triple bonds, e.g.

  \[
  \begin{align*}
  &\text{H}_3\text{C} &\text{H}_2\text{C} &\text{HC} \\
  &\text{CH}_3 &\text{CH}_2 &\text{CH} \\
  &\text{ethane} &\text{ethene} &\text{ethyne} \\
  \\
  &\text{(ethylene)} &\text{(ethene)} &\text{(acetylene)}
  \end{align*}
  \]

1. **sigma (σ):** single bond, strongest, formed by head-on orbital overlap, symmetrical with respect to rotation around bond axis.
2. **pi (π):** double bond where two lobes of one atomic orbital of one atom overlap two lobes of the other, weaker than σ, on-rotational.
Bonds

iii. triple bonds formed by one $\sigma$- and two $\pi$-bonds—stronger than each individual bond

- More energy is required to break double and triple bonds than single bonds
- Electrons can be shared equally (e.g. C–C) or unequally (e.g. O–H) in which case the bond is polarized with the electrons attracted to the electronegative O atom (forming a dipole)
- Many require enzymes to make or break them

Ionic bonds

- Ionic or electrostatic bonds form between positively and negatively charged ions
- Have some degree of covalent bond nature as atoms close together share electron density
- Intermediate in strength between covalent and hydrogen bonds
- e.g. Na$^+$Cl$^-$

Dipole–dipole interactions

- Dipole = separation of positive and negative charge
- Interaction between permanent dipoles increases attraction between molecules
- e.g. HCl

\[
\begin{align*}
\delta^+ & \quad \delta^- \\
H & \quad Cl \quad … \quad H & \quad Cl
\end{align*}
\]

Hydrogen bonds (H-bonds)

- Hydrogen or H-bonds form between polar molecules
- When H is attached to an electronegative atom such as O or N, the bond becomes polarized
- Bond formed by sharing of non-bonding (lone pair) electrons from one atom with an H covalently attached to an electronegative atom and therefore starved of electrons
- Weaker than ionic bonds but stronger than hydrophobic or van der Waals forces—essentially a strong dipole–dipole interaction
- e.g. stabilize secondary structures of proteins (alpha helices and beta sheets), base pairs in DNA double helix and between water molecules (as shown below)

\[
\begin{align*}
\delta^+ & \quad \delta^- \\
\delta^- & \quad \delta^+ \\
\delta^+ & \quad \delta^- \\
H & \quad O & \quad H & \quad O
\end{align*}
\]
van der Waals forces

- van der Waals forces occur between any atoms
- Transient dipoles formed by electron movement lead to electrostatic attraction between atoms/molecules a short distance apart
- Can form between polar or hydrophobic molecules
- Individually weak but additive in large molecules
- At very short distances, van der Waals interactions become strongly repulsive (≤atomic radius)

Hydrophobic effect

- The hydrophobic effect will drive the association of hydrophobic molecules in a polar, aqueous environment to exclude water and maximize entropy
- e.g. fatty acids in centre of lipid bilayer or fat droplet
- Individually weak but additive in large molecules

Comparison of bond types

For comparison of properties of important bond types see Table 1.1.

<table>
<thead>
<tr>
<th>Bond</th>
<th>Approx. strength (kJ/mol at 25°C)</th>
<th>Length (internuclear separation distance) (Å)</th>
<th>Where found</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covalent</td>
<td>200–1000</td>
<td>0.7–2.7, e.g. 1.4 for C–C in graphite</td>
<td>intermolecular</td>
<td>peptide bond</td>
</tr>
<tr>
<td>Ionic or electrostatic</td>
<td>-40</td>
<td>2.8</td>
<td>inter- and intramolecular</td>
<td>salt bridges</td>
</tr>
<tr>
<td>Dipole–dipole</td>
<td>4–20</td>
<td></td>
<td>inter- and intramolecular</td>
<td></td>
</tr>
<tr>
<td>Hydrogen</td>
<td>10–30</td>
<td>2.7–3.1</td>
<td>inter- and intramolecular</td>
<td>between polar side chains of amino acids</td>
</tr>
<tr>
<td>van der Waals</td>
<td>4</td>
<td>3–4</td>
<td>inter- and intramolecular</td>
<td>between tightly packed atoms in centre of protein molecule</td>
</tr>
<tr>
<td>Hydrophobic effect</td>
<td>&lt;1 per CH₃</td>
<td>3–5</td>
<td>inter- and intramolecular</td>
<td>fatty acids in lipid droplet or membrane</td>
</tr>
</tbody>
</table>

Table 1.1 Comparison of bond types
Note: strength is approximate and varies according to molecules involved.

Check your understanding

Describe the different bond types occuring in organic molecules. *(Hint: can you give an example of each type?)*
Proteins

1.2 PROTEINS

Proteins are essential to all living organisms as they catalyse the majority of enzymatic reactions in the cell and also play an essential structural role. Proteins are polymers of amino acids also referred to as polypeptide chains. The function of a protein is intimately linked to its structure.

Key functions of proteins

- Structural, e.g.
  i. collagen, which forms long fibres
  ii. actin, which forms long filaments made up of short monomers
- Enzymes, e.g. lysozyme
- Carriers, e.g. haemoglobin
- Transmembrane, e.g. ion channels
- Signalling, e.g. insulin

Amino acids

- Fundamental building block of proteins
- General structure of an amino acid (shown in non-ionized form below and as zwitterion)

\[
\begin{align*}
\text{H} & \quad \text{N} & \quad \text{C} & \quad \text{C} & \quad \text{O} \\
\text{H} & \quad \text{R} & \quad \text{H} & \quad \text{O} & \quad \text{H} \\
\text{N}^+ & \quad \text{C} & \quad \text{C} & \quad \text{O} & \quad \text{O}^-
\end{align*}
\]

- Amino acids are chiral about the central carbon atom next to the COOH (α-carbon)
- Natural amino acids are the l form (laevorotatory)
- Twenty naturally occurring in proteins that differ in the R side chain (Table 1.2)
- Side chains confer one or more specific characteristics on the amino acid, e.g. tyrosine has an aromatic side chain that possesses an –OH group which can be phosphorylated
- Amino acids form a zwitterion at neutral pH (i.e. in the cytosol) with a positive charge on the amine group and a negative charge on the carboxyl group
- They act as buffers (see Figure 1.1 for titration curve)
- Excess dietary amino acids are broken down:
  i. amino portion is used to form urea in mammals
  ii. carbon skeletons are recycled for the formation of glucose via \textit{gluconeogenesis} (glucogenic), or of \textit{ketone bodies} (ketogenic)

See 5.8 Amino acid breakdown and synthesis (p. 156) for metabolism of amino acids
## Proteins

<table>
<thead>
<tr>
<th>Property</th>
<th>Amino acid</th>
<th>Abbreviation</th>
<th>Side chain</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aliphatic non-polar</strong></td>
<td>alanine</td>
<td>ala</td>
<td>CH₃</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cysteine</td>
<td>cys</td>
<td>SHCH₂</td>
<td>sulphur containing—can form disulphide bonds</td>
</tr>
<tr>
<td></td>
<td>isoleucine</td>
<td>ile</td>
<td>C₂H₅(CH₃)CH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>leucine</td>
<td>leu</td>
<td>(CH₂)₂CHCH₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td>glycine</td>
<td>gly</td>
<td>H</td>
<td>smallest amino acid—makes up ~1/3 collagen</td>
</tr>
<tr>
<td></td>
<td>methionine</td>
<td>met</td>
<td>CH₃SCH₂CH₂</td>
<td>initiator amino acid for all proteins; sulphur containing</td>
</tr>
<tr>
<td></td>
<td>valine</td>
<td>val</td>
<td>(CH₂)₂CH</td>
<td></td>
</tr>
<tr>
<td><strong>Aliphatic polar—hydroxyl</strong></td>
<td>serine</td>
<td>ser</td>
<td>CH₂OH</td>
<td>can be phosphorylated or glycosylated (O-linked)</td>
</tr>
<tr>
<td></td>
<td>threonine</td>
<td>thr</td>
<td>CH₂CHOH</td>
<td>can be phosphorylated or glycosylated (O-linked)</td>
</tr>
<tr>
<td><strong>Aliphatic polar—amide</strong></td>
<td>asparagine</td>
<td>asn</td>
<td>CONH₂CH₂</td>
<td>can be glycosylated—N-linked</td>
</tr>
<tr>
<td></td>
<td>glutamine</td>
<td>gln</td>
<td>CONH₂CH₂CH₂</td>
<td>can be glycosylated—N-linked</td>
</tr>
<tr>
<td><strong>Aromatic</strong></td>
<td>phenylalanine</td>
<td>phe</td>
<td>C₆H₅CH₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tryptophan</td>
<td>trp</td>
<td>C₆H₄NHCH₂</td>
<td>disrupts alpha helices</td>
</tr>
<tr>
<td></td>
<td>tyrosine</td>
<td>tyr</td>
<td>C₆H₄OHCH₂</td>
<td>polar—can be phosphorylated</td>
</tr>
<tr>
<td><strong>Basic</strong></td>
<td>arginine</td>
<td>arg</td>
<td>(CN₃H₅)⁺(CH₂)₃</td>
<td>can be acetylated or methylated</td>
</tr>
<tr>
<td></td>
<td>histidine</td>
<td>his</td>
<td>C₆N₂H₃CH₂</td>
<td>pKₐ ~7</td>
</tr>
<tr>
<td></td>
<td>lysine</td>
<td>lys</td>
<td>(NH₃)⁺(CH₂)₄</td>
<td>can be acetylated or methylated</td>
</tr>
<tr>
<td><strong>Acidic</strong></td>
<td>aspartic acid</td>
<td>asp</td>
<td>COO⁻HCH₂</td>
<td>normally negative</td>
</tr>
<tr>
<td></td>
<td>glutamic acid</td>
<td>glu</td>
<td>COO⁻H₂CH₂</td>
<td>normally negative</td>
</tr>
<tr>
<td><strong>Cyclic</strong></td>
<td>proline</td>
<td>pro</td>
<td>C₆H₄NH⁺⁺</td>
<td>disrupts alpha-helices; can be hydroxylated, e.g. in collagen</td>
</tr>
</tbody>
</table>

*Table 1.2 Amino acid side chains R in R–CH(NH₂)COOH*
Proteins

![Graphs of amino acid titration]

Figure 1.1 Titration of amino acids. Titration curves of amino acids with (A) uncharged, (B) acidic and (C) basic side chains. Grey line = midpoint of titration.

Key features of protein structure

- Four main levels of structure—primary, secondary, tertiary, and quaternary (see below)
- Three-dimensional shape is dependent on primary amino acid sequence
- Proteins fold to adopt the most thermodynamically stable conformation
- Chaperone molecules are required in the cell to stabilize partially folded intermediates and promote correct protein folding
- Proteins can show some flexibility in structure allowing conformational changes to regulate activity, e.g. allosteric regulator allolactose binds to lac repressor protein and leads to its dissociation from DNA
  - see Chapter 4: Regulation of transcription in prokaryotes (p. 94) for regulation of lac operon
- Interior of the protein will contain hydrophobic interactions between side chains and ionic and H-bonds between polar/charged side chains

continued
Primary structure of proteins

- Linear arrangement of amino acids
- Determined by order of nucleotides in mRNA (and ultimately DNA)
- Amino acids are linked together by peptide bonds

![Peptide Bond Diagram]

- Peptide bond (dotted box) has partial double bond character due to delocalization of electrons, so is rigid and planar
- R side chains are in the trans configuration, i.e. on opposite sides of the peptide bond
- All polypeptide chains have an amino [N] and carboxy [C] terminus
- Exist as zwitterions at physiological pH, i.e. at least one positive (–NH₃⁺) and one negative (–COO⁻) charge per molecule, plus further charges associated with the R side chains
- Depending on the number of acidic and basic side chains, the pH at which each protein is neutral will differ. This pH is known as the isoelectric point (pI). pH gradients can be used to separate proteins of different charge.

> see 9.9 Protein purification (p. 249) and 9.10 Protein analysis (p. 253) for protein separation
Proteins

Secondary structure of proteins

- Local folding of amino acids in three dimensions, defined mainly by H-bonds between residues of the peptide backbone (Figure 1.2)
- Two major forms of secondary structure:
  i.  alpha-helix
  ii.  beta-pleated sheet
- $\alpha$-helix—hydrogen bonds every four residues within a strand. R groups face outwards
- $\beta$-pleated sheet—held together by hydrogen bonds between strands, either parallel or antiparallel

Figure 1.2 Protein secondary structure motifs. (A) The $\alpha$-helix is stabilized by H-bonds within the chain. There are around 3.6 residues for every turn and the R groups face outwards. (B) The $\beta$-sheet is stabilized by H-bonds between strands. (C) In cartoons of protein structures, $\beta$-sheets are usually shown as thick arrows, which can lie (i) parallel or (ii) antiparallel to each other. The R groups alternately lie above and below the plane of the sheet. (D) Cartoon of a protein structure (monomer of PCNA) to show $\beta$-sheets and $\alpha$-helices. Thin lines represent regions that are less ordered.
• Secondary structure of proteins is vital to their function, e.g. keratin (skin) is largely \( \alpha \)-helix
• Collagen is made up of a triple helix (coiled coil) with glycine every third residue—small side chain (hydrogen) allows strands to come into close proximity for tight packing, important for structural strength. NB The collagen triple helix is unique and distinct from the \( \alpha \)-helix

**Tertiary structure of proteins**

• Folding of the protein in three dimensions
• May involve interactions between residues that are far apart in the primary sequence
• Defined via interactions between the R side chains
• Tertiary structure stabilized by large number of mainly weak bonds—allows proteins to be flexible and change shape (important, for example, in allosteric regulation)
• Forces/bonding can be hydrophobic, ionic, or covalent (disulphide S–S bridges)
• S–S bonds form between cysteine residues which are brought together by tertiary structure and stabilize it—prevalent in proteins on cell surface or secreted from cell
• Tertiary structure commonly subdivided into **domains**, which are independently folding regions often with particular functions
• Most enzymes are globular proteins
• Globular proteins, e.g. myoglobin—ability to bind oxygen is defined by the precise tertiary structure (hyperbolic \( O_2 \) dissociation curve)
• Fibrous proteins serve structural roles, e.g. \( \alpha \)-helical coiled coil of keratins in skin aids protein stability and confers strength
• Some proteins have both fibrous and globular regions, e.g. myosin in muscle—the structural region is fibrous (twisted \( \alpha \)-helices) and the enzymatic ATPase activity is located in a globular head domain

**Quaternary structure of proteins**

• Arrangement of polypeptide chains in a multi-subunit protein, or inclusion of **prosthetic groups**
• e.g. haemoglobin: interaction of four subunits leads to co-operative binding of oxygen (**sigmoidal** dissociation curve for \( O_2 \) compared with hyperbolic for myoglobin, which has very similar unit structure)

**Key evidence for protein structure**

• Denatured **RNaseA** (no enzyme activity) can be refolded *in vitro* to restore activity—active structure is thermodynamically the most stable
• Single amino acid changes can alter protein function, e.g. mutant haemoglobin in sickle cell anaemia and thalassaemias
• Structures of many proteins and domains are known, and side chain interactions follow expected rules
Lipids

• Structural evidence from:
  i. nuclear magnetic resonance (NMR) can be used to deduce the structure of polypeptides up to 200 amino acids long in solution
  ii. X-ray crystallography can be used to deduce the structure of larger proteins as long as high quality crystals are available
  iii. single particle electron microscopy (EM) and cryo EM, atomic force microscopy

> see also 9.12 Biophysical techniques (p. 262) for methods used to determine protein structure

1.3 LIPIDS

Lipids are water-insoluble, structurally diverse organic compounds. Functions include structural components of cell membranes, insulators (thermal, chemical, and electrical), energy stores, and biological signalling molecules.

**Key features of lipids**

• Most common lipids comprise fatty acids covalently bound to alcohol, e.g. glycerol or sphingosine; many variations in structure

• Two major types of lipids:
  i. hydrophobic (non-polar), e.g. triglycerides
  ii. amphipathic (polar head and non-polar tail), e.g. phospholipids

• Phospholipid polar head oriented toward the solvent and non-polar tails away from the solvent in aqueous solution—important in biological membranes

• Non-polar lipids are insoluble in aqueous solvents

• Amphipathic lipids form micelles (unstable) or self-healing bilayers (stable)

• Complex lipids form the structural basis of biological membranes and include:
  i. phospholipids (lipid + phosphate)
  ii. sphingolipids (lipid + sphingosine)
  iii. glycosphingolipids (sphingolipid + carbohydrate)

• Triglycerides (fats) are major molecules of fuel storage:
  i. fats are stored predominantly in adipocytes in mammals
  ii. stored fats metabolized to acetyl CoA by β oxidation in mitochondria, or formed into ketone bodies for utilization as preferred fuel of mammalian heart muscle, and of brain on starvation

> see 5.6 Lipid breakdown and synthesis (p. 147) and 5.7 Ketone body breakdown and synthesis (p. 154) for lipid and ketone body metabolism
  iii. carried as lipoproteins in mammalian blood
Fatty acids (FA)

- 12–20 carbons in aliphatic, unbranched hydrocarbon chain (usually even number) terminating in a carboxyl group
- Two major types
  - saturated, i.e. no double bonds
  - unsaturated, i.e. contains C=C double bonds
- Unsaturated lipids can be:
  - monounsaturated has one double bond
  - polyunsaturated has ≥2 double bonds
- Each fatty acid has specific $T_m$ (melting temperature) dependent on how closely the individual molecules pack together
- Fatty acids are bound to serum albumin in blood for transport to heart, skeletal muscle, liver, etc. in mammals

Key aspects of fatty acid structure

- Fatty acids are often represented numerically as $x:y\Delta^z$, where:
  - $x$ = number of C in hydrocarbon chain
  - $y$ = number of double bonds
  - $z$ = position of double bonds (counted from carboxylate end)
- e.g. palmitate 16:0, oleate 18:1$\Delta^9$, arachidonate 20:4$\Delta^5,8,11,14$
- Carbons of fatty acid are numbered from the carboxyl carbon (C1); alternatively, first carbon of the hydrocarbon chain can be denoted the \( \alpha \) carbon, and the last carbon in chain as \( \omega \) (irrespective of chain length)
- Position of double bonds may be indicated with respect to \( \omega \) carbon, e.g. oleate 18:1$\Delta^9 = \omega$-9
- Double bonds in fatty acids either cis or trans
  
  \[
  \text{cis} \quad / \quad \text{trans} \\
  \text{HC} \quad \text{CH} \quad \text{HC} \quad \text{CH}
  \]

- Cis are more common in fatty acids found in nature
- Cis double bonds introduce bends or kinks into the hydrocarbon chain and lead to a decrease in $T_m$ because the fatty acid chains cannot pack as closely
- Trans double bonds do not appreciably affect the direction of fatty acid chains—can pack tightly

Key physical properties of fatty acids

- Oxidation yields ~37 kJ/g
- $T_m$ increases with length hydrocarbon chain; also degree of saturation
- Saturated fats are waxy at room temperature (e.g. butter)
- Unsaturated fats are generally liquid at room temperature (e.g. olive oil)
Lipids

- Forces: covalent, hydrophobic, and van der Waals

\[ \begin{align*}
\text{glycerol} & \quad \text{triglyceride} \\
& \quad \begin{array}{c}
H_2C-OH \\
HC-OH \\
H_2C-OH
\end{array}
\end{align*} \]

Triacylglycerols (triglycerides, TAG)

- Three fatty acid chains linked to glycerol (three-carbon alcohol = triol) via ester bond (condensation reaction)
- NB glycerol forms the backbone of many naturally occurring lipids
- Most abundant lipids in mammals
- Act as compact, neutral, anhydrous storage form of fatty acids
- Found especially as fat droplets in, for example, adipocytes of mammals
- Not present in biological membranes
- Major dietary source of lipids for mammals and major source of stored energy

\[ \begin{align*}
\text{H}_2\text{C}-&\text{OH} \\
\text{HC}-&\text{OH} \\
\text{H}_2\text{C}-&\text{OH}
\end{align*} \]

Glycerophospholipids (phosphoglycerides)

- Most abundant lipids in biological membranes
- Amphipathic (polar head, non-polar tail)
- Highly diverse class according to head group and/or two possible different fatty acid chains
- e.g. human red blood cells contain \( \geq 21 \) different types of phosphatidylcholine which differ not in head group but in combination of fatty acid chains
- Generalized phosphoglyceride:

\[ \begin{align*}
X-O-P-O- & \quad \text{CH}_2 \\
& \quad \begin{array}{c}
\text{HC}-O-C-R_1 \\
\text{H}_2\text{C}-O-C-R_2
\end{array}
\end{align*} \]

12 **Thrive in** Biochemistry & Molecular Biology
Key types of glycerophospholipid

<table>
<thead>
<tr>
<th>name</th>
<th>–O–X of general formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>phosphatidate</td>
<td>–O–H</td>
</tr>
<tr>
<td>phosphatidylcholine (PC)</td>
<td>–O–CH₂CH₃N(CH₃)₃</td>
</tr>
<tr>
<td>phosphatidylethanolamine (PE)</td>
<td>–O–CH₂CH₂NH₃⁺</td>
</tr>
<tr>
<td>phosphatidylerserine (PS)</td>
<td>–O–CH₂CH(NH₃⁺)COO⁻</td>
</tr>
</tbody>
</table>

Ether phospholipids

- Formed by reduction of carbonyl groups of phospholipids
- Can serve as important signalling molecules, acting at low concentrations
- e.g. platelet activating factor (1-alkyl, 2-acetyl ether analogue of phosphatidyl choline)
  i. promotes platelet aggregation
  ii. promotes dilation of blood vessels

Sphingolipids

- Found in both plant and animal cell membranes, especially in mammalian central nervous system
- Sphingosine forms the core backbone with different fatty acids attached
- Sphingosine = C18 unbranched alcohol
- Ceramide is the metabolic precursor in the formation of all sphingolipids, with a fatty acyl group amide bonded to the C2 of sphingosine
- Three major classes of sphingolipids
  i. sphingomyelins have phosphate group therefore = phosphosphingolipid
  ii. cerebrosides have carbohydrate therefore = glycosphingolipids
  iii. gangliosides have carbohydrate therefore = glycosphingolipids

Isoprenoids

- Include steroids
- Key example: cholesterol (found in animal plasma membranes, rarely in plants and never in prokaryotes)
- Other examples:
  i. stigmasterol (plants)
  ii. mammalian bile salts
  iii. mammalian hormones such as the estrogens and androgens (e.g. testosterone)
  iv. lipid vitamins, e.g. A, D, E, and K
  v. sterols of plants, fungi, and yeast
Carbohydrates

Waxes

- Non-polar esters of long chain fatty acids and long chain monohydric alcohols
- Protective role, forming a waterproof coating on:
  - leaves and fruits of some plants
  - skin or fur (mammals)
  - feathers (birds)
  - exoskeleton (insects)
- e.g. beeswax = ester of palmitate (16:0) and 30C myricyl alcohol

Eicosanoids

- Oxygenated derivatives of C20 polyunsaturated fatty acids, e.g. arachidonic acid
- Both physiological and pathological roles, e.g. vasoconstriction, inflammation, etc.
- Prostaglandins are eicosanoids with a cyclopentane ring

Enzymes that cleave phospholipids

- Phospholipases cleave phospholipids at specific positions
- Different classes according to position cleaved in phospholipid
- Digestive enzymes, e.g.
  - mammalian pancreatic PLA$_2$
  - snake venom PLA$_2$
  - secreted by bacteria, e.g. Clostridium perfringens $\alpha$-toxin = PLC
- Important in signalling cascades, e.g. inflammatory cascade: PLA$_2$ generates arachidonic acid, a precursor of prostaglandins

1.4 CARBOHYDRATES

Carbohydrates make up most of the organic matter on earth. Carbohydrates are aldehydes or ketones with multiple hydroxyl groups or their polymers. They act as energy stores and fuels, as well as having a structural role.

Key features of sugars

- Hydroxylated ketones or aldehydes that can be linked by glycosidic bonds to form oligo- and polysaccharides
- Monomeric sugars act as intermediates in metabolism, e.g. glucose and other molecules in the glycolytic pathway

continued
Monosaccharides

- Aldehydes and ketones with two or more hydroxyl groups \((\text{CH}_2\text{O})_n\) where \(n \geq 2\) (see Table 1.3)
- Contain multiple asymmetric carbon atoms
- Naturally occurring hexoses tend to be \(\alpha\) isomers e.g. \(\alpha\)-glucose and \(\alpha\)-fructose
- Pentoses and hexoses can cyclize to form ring structures
- Rings can be six-membered (pyranose) or five-membered (furanose)
- Can be linked together to form oligo- and polysaccharides
- Each ring structure can exist as either \(\alpha\) or \(\beta\) forms:
  - \(\alpha\) (alpha) sugars have the OH group of C1 below the plane of the ring
  - \(\beta\) (beta) sugars have the OH group of C1 above the plane of the ring
- The C1 carbon is the one with the aldehyde group on for aldoses, or the one adjacent to the ketone group of ketoses
- Glucose conformations:
Carbohydrates

<table>
<thead>
<tr>
<th>General name</th>
<th>Number of carbons</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triose</td>
<td>3</td>
<td>aldose</td>
<td>glyceraldehyde</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>ketose</td>
<td>dihydroxyacetone</td>
</tr>
<tr>
<td>Hexose</td>
<td>6</td>
<td>aldose</td>
<td>glucose</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>ketose</td>
<td>fructose</td>
</tr>
</tbody>
</table>

Table 1.3 Common sugars

Key monosaccharides
- Glucose (C₆ aldose)
- Galactose (C₆ aldose)
- Mannose (C₆ aldose)
- Fructose (C₆ ketose)
- N-Acetylglucosamine (C₆ aldose)
- N-Acetylglactosamine (C₆ aldose)

Disaccharides
- Two monosaccharides joined together by glycosidic bonds
- These bonds can be either α or β configuration
- Bond formation requires activation of the sugar monomer in order to make it reactive, usually by transient association with a nucleotide triphosphate, e.g. UTP

Key disaccharides
- Sucrose = glucose + fructose (α-D-glucopyranosyl-(1-2)-β-D-fructofuranoside)
- Maltose = glucose + glucose (α-D-glucopyranosyl-(1-4)-α-D-glucopyranose)
- Lactose = galactose + glucose (β-D-galactopyranosyl-(1-4)-α-D-glucopyranose)

Polysaccharides
- Multiple monosaccharide units linked together by glycosidic bonds to form a polymer
- Can serve as fuel stores, e.g. glycogen in animals and starch in plants
- Can have structural role, e.g. in plant cell wall (cellulose) and insect exoskeleton (chitin)

Glycogen
- Large branched polymer of glucose
- Most glucose units are linked by α-1,4 linkages; branches are formed by α-1,6 glycosidic bonds
- Branches form at approximately every 10th residue
- Fuel store in animal cells—predominantly liver and muscle of mammals

16  Thrive in Biochemistry & Molecular Biology
Starch

- Major fuel store in plant cells
- Very long polymers of glucose synthesized in chloroplast
- Can be stored in cytoplasm or chloroplast
- Exists with two components:
  i. amylose—unbranched chains of glucose with $\alpha$-1,4 linkages
  ii. amylopectin—like amylose but with a branch with an $\alpha$-1,6 linkage approximately every 30 residues

Dextran

- Fuel storage polysaccharide in yeast and bacteria
- Polymer of glucose
- Almost exclusively linked $\alpha$-1,6, though a few branches with $\alpha$-1,2, $\alpha$-1,3, or $\alpha$-1,4 linkages
Cellulose
- Structural molecule of plant cell walls, giving strength and stability
- Unbranched polymer of glucose with β-1,4 linkages
- Stabilizing H-bonds between adjacent glucose units give cellulose its tensile strength—not found in α linkage form (glycogen, starch)

Chitin
- Structural molecule of insects, forming exoskeleton
- Polymer of N-acetylglucosamine residues in β-1,4 linkage (like cellulose except for acetylamino group on C2)
Glycoconjugates

- Oligosaccharides can be attached to protein and lipids by glycosidic bonds
- Glycoprotein normally on extracellular domain:
  - $O$-glycosidic bonds to serine or threonine
  - $N$-glycosidic bond to asparagine
- e.g. cell adhesion molecules (CAMs) are glycoproteins
- Sugars provide essential ‘address label’ on many proteins, e.g. lysosomal enzymes carry characteristic mannose 6-phosphate tag
- Sugar groups may also be important in the mechanics of protein trafficking within the cell, e.g. nuclear pore complex proteins are glycosylated, and nuclear import is blocked by lectins that bind to the sugars
- Glycolipids are found predominantly in plasma membranes of cells in the vertebrate nervous system—may be important both in cell recognition and in electrical insulation

1.5 DNA

DNA (deoxyribonucleic acid) is the genetic material of most living organisms. It exists as a very long filamentous double helix (i.e. two strands of DNA).

Key features of DNA (deoxyribonucleic acid)

- Polymer of deoxyribonucleotides linked by phosphodiester bonds
- Chemically, DNA is made up of four nitrogenous bases (A, T, G, C) linked to a deoxyribose sugar bearing a phosphate group
- Phosphodiester bonds form between adjacent sugars
- DNA double helix is formed by complementary base pairing through hydrogen bond formation between bases on opposite strands of the double helix
- A always pairs with T, and G always pairs with C: the two strands are complementary
- Sugar phosphate backbone, which is negatively charged, interacts with aqueous environment
- Genetic code is determined by the order of the four nitrogenous bases
DNA

Chemical composition of DNA

- There are four nitrogenous bases in DNA: cytosine (C), guanine (G), thymine (T), and adenine (A)
- Pyrimidines (C and T) have a single nitrogenous ring:

\[
\text{cytosine} \quad \text{thymine}
\]

- Purines (G and A) have double ring structure:

\[
\text{adenine} \quad \text{guanine}
\]

- Deoxyribonucleosides = base + sugar
  i. nitrogenous bases linked to the C1 carbon of 2’ deoxyribose by a β-N-glycosidic bond
- Deoxyribonucleotides = base + sugar + phosphate
  i. nucleoside with the 5’ carbon of deoxyribose bound by an ester linkage to a mono-, di-, or tri-phosphate group (dNMP, dNDP, or dNTP)
  ii. dNTPs are substrates for DNA synthesis during DNA replication

➤ see 4.1 DNA replication (p. 78)
  iii. general structure of a deoxyribonucleotide (dNTP):

\[
\text{base} \quad \text{OH} \quad \text{H}
\]

Key structural features of DNA

- DNA is a polymer of four different deoxyribonucleotides (dAMP, dTMP, dCMP, dGMP) linked by phosphodiester bonds

➤ see Chapter 4: DNA polymerases (p. 80) for bond formation

20 Thrive in Biochemistry & Molecular Biology
• Polymer has polarity: the 5′ terminal nucleotide is a deoxyribonucleotide triphosphate; the 3′ terminal nucleotide has a free 3′ hydroxyl group—always drawn 5′→3′ by convention.

• Acidic due to the phosphate groups.

• Bases are hydrophobic and reside on inside of helix: sugar–phosphate backbone is hydrophilic and faces the aqueous environment.

• Order of nitrogenous bases along the DNA determines the genetic code.

• Two single strands of DNA coil around each other in anti-parallel fashion forming double stranded helix, or duplex DNA (see Figure 1.3), held together by complementary base pairing: A always base pairs with T (two hydrogen bonds); C always base pairs with G (three hydrogen bonds).

• Width of AT pair = width of GC pair (precise, so no helical disruption).

• DNA is a stable molecule but subject to spontaneous and environmental damage—see 4.4 DNA repair (p. 112) for details.

• Three major conformations of duplex DNA: A, B, and Z:
  i. B form DNA is hydrated right-handed helix, probably the form found in nature (see Figure 1.3)
    a. double helix with major and minor groove; helix diameter is 23.7 Å (2.37 nm)
    b. ten bases per turn of helix—3.4 Å (0.34 nm) per base giving a periodicity (one complete turn of the helix) of 34 Å or 3.4 nm
    c. hydrogen bonds between complementary base pairs are perpendicular to the axis of the helix
  ii. A is dehydrated right-handed helix, more tightly wound than B form
  iii. Z is left-handed helix, with no grooves, and occurs on repetitive alternating stretches of Pu/Py

• Bonding in double helix:
  i. hydrogen bonds between complementary base pairs: A forms two hydrogen bonds with T, and C forms three hydrogen bonds with G
  ii. hydrophobic effects (bases inside backbone)
  iii. base stacking interactions (van der Waals contacts—individually weak but additive, so significant for long DNA polymer)
  iv. charge–charge interactions: the large negative charge of the backbone that would otherwise destabilize DNA is generally neutralized in vivo by presence of cations (e.g. Mg$^{2+}$) and cationic proteins (e.g. histones in eukaryotes)

Key evidence for DNA structure

• Base pairing deduced from base ratios: A + G = T + C; A = T, G = C ( Chargaff rules—from TLC analysis)

• X-ray crystallography (Franklin and Wilkins): diffraction pattern shows double helix.
DNA

Figure 1.3 (A) DNA double helix. Two antiparallel strands of DNA, held together by H-bonds between the bases, twist to form a double helix with the hydrophilic sugar-phosphate backbone on the outside. There are ten bases per turn, aligned perpendicular to the helical axis. Note major and minor grooves. (B) Nucleosome. An octamer of core histones (two each of H2A, H2B, H3, and H4, pale grey) has two turns of DNA wrapped around it (approx. 146 bp). The linker histone H1 (dark grey) may be present at the entry/exit sites.

- Model building based on X-ray crystallography (Watson and Crick): bases must be on inside and purine must base pair with pyrimidine
- Thermal denaturation with $A_{260}$ nm measurements give $T_m$ (melting temperature) for DNA—absorbance increases sharply on disruption of cooperative interactions between stacked H-bonded base pairs

Higher order structure of DNA

Due to the long length of DNA compared with the size of cell, DNA must be compacted but remain accessible for replication and gene expression. A chromosome is a single molecule of double-stranded DNA. Prokaryotes have circular chromosomes and eukaryotes have linear chromosomes.

Prokaryotes
- e.g. *Escherichia coli* 4.2 × 10⁶ bp of genomic DNA in one circular chromosome = 1.3 mm long; cell diameter <1 μm

22 **Thrive in** Biochemistry & Molecular Biology
• Circular DNA can be supercoiled (linking number = number of supercoils)
• May be bound by HU proteins which are basic
• Cell may also contain extra-chromosomal plasmids—usually large supercoiled circular DNA molecules (e.g. up to 100 kb) that may encode, for example, fertility factors or antibiotic resistance. NB Naturally occurring plasmids are much bigger than plasmids used for cloning in molecular biology (which may be \( \sim 3–10 \) kb)
• Genomic DNA of prokaryotes is located in the cytoplasm as nucleoid
• In prokaryotes, much of the DNA encodes proteins or structural RNA molecules

**Eukaryotes**

• e.g. *Homo sapiens* \( \sim 2.9 \times 10^9 \) bp of DNA in \( 2 \times 23 \) chromosomes = 1.8 m long.
  2 nm diameter; nuclear diameter \(< 10 \) \( \mu \)m
• Compartmentalized into the nucleus of eukaryotes
• Complex of DNA with associated proteins is known as chromatin
• DNA packaged into hierarchical structure giving overall \( 10^4 \)-fold compaction
  i. duplex DNA 2 nm diameter
  ii. wrapped around octamer of basic histone proteins (core nucleosome), giving ‘beads on a string’ appearance in EM (11 nm fibre)
  iii. 30 nm diameter fibre with linker histone H1
  iv. 300 nm diameter solenoids
  v. 1,400 nm diameter metaphase bivalent chromosomes (each sister chromatid 700 nm diameter)
• Nucleosome consists of a histone octamer (two each of four core histones H2A, H2B, H3, and H4) with two turns of DNA wrapped around (146 base pairs) (see Figure 1.3B)
• Linker distance between nucleosomes can vary
• Actively transcribed genes are found in more loosely packed chromatin known as euchromatin
• Regions of DNA not containing actively transcribed genes are often more tightly packed into heterochromatin
• Euchromatin and heterochromatin are associated with different histone modifications
  ➔ see Chapter 4: Regulation of initiation of transcription in eukaryotes (p. 99) for histone modifications associated with gene expression

**Key evidence for higher order structure**

• DNA spreads from bacterial cells or eukaryotic chromosomes treated with detergent and/or EDTA; examination by electron microscopy (EM) (size of genomic DNA relative to cell) shows large amount of DNA spilling out from lysed cell
• X-ray crystallography (nucleosome)
DNA

- EM shows various levels of packaging
- Light microscopy (chromosomes)
- Partial micrococcal nuclease digestion of eukaryotic chromatin—cuts DNA to produce a ladder of fragments of approx. 140 bp as the DNA is cleaved in linker DNA between regularly spaced nucleosomes
- Agarose gels ± treatment with topoisomerase to separate differently supercoiled forms demonstrates supercoiling in prokaryotes

Key functions of DNA

- DNA is the genetic material of the cell
- Coding information is arranged as genes (~1–10% of higher eukaryotic DNA is made up of protein coding genes) and intergenic regions, together with specific structural and regulatory regions
- Majority of genomic DNA in prokaryotes and lower eukaryotes codes for proteins or structural RNA molecules
- In higher eukaryotes such as man, ~1% of the genome codes for proteins

Genes

- Encode structural proteins (via the intermediate mRNA) and functional RNA molecules (rRNA, tRNA, snRNA, microRNA, and other regulatory RNAs)
- Order of the bases in the genes determines the order of amino acids in proteins
  - see 4.3 Protein synthesis (translation) (p. 101)
- Eukaryotic coding regions of genes (exons) can be interrupted by non-coding introns (intervening sequences)—rare in yeast, also found in some bacteria and bacteriophage

Intergenic regions (between genes) in eukaryotes include

- Promoters and gene regulatory sequences
  - see 4.2 RNA synthesis (transcription) (p. 88)
- Origins of replication
  - see 4.1 DNA replication (p. 78)
- Matrix attachment sites to anchor DNA to proteinaceous structure in nucleus
- Telomeres (ends of chromosomes)—important for chromosomal stability; measure of cell age as they shorten in each round of DNA replication in somatic cells
- Centromeres (middles of chromosomes)—essential for attachment to microtubules of spindle during cell division to ensure accurate segregation of sister chromatids or homologous chromosomes to daughter cells in mitosis or meiosis, respectively
- Repetitive DNA elements, e.g.
  i. microsatellite and minisatellite, e.g. Alu in *H. sapiens*
  ii. transposable elements (e.g. copia) and interspersed elements (LINES, SINES)
Uncharacterized regions—originally called ‘junk DNA’? but much of this DNA may encode critical regulatory small RNAs

**Looking for extra marks?**

**Pseudogenes** are potential genes that are not expressed, or genes without promoters—often also lack introns and may have arisen from mRNA by reverse transcription.

### Key evidence that DNA is the genetic material

- White blood cells from pus on surgical bandages treated with HCl gave a precipitate containing carbon, hydrogen, oxygen, nitrogen, and phosphorus—called ‘nuclein’ (Miescher 1869)
- Factor could be transferred from heat-killed pathogenic (smooth) pneumococci to non-pathogenic (rough) strain—the ‘transforming principle’ (Griffith 1928) (Figure 1.4)
- Transforming principle was resistant to lipases, proteases, and heat, but was destroyed by treatment with nucleases. Purification showed that the principle was DNA (Avery, MacLeod, and McCarty 1944)
- Bacteriophage radiolabelled with $^{35}\text{S}$ (proteins) or $^{32}\text{P}$ (nucleic acid) infected into bacteria. Only the $^{32}\text{P}$ transferred into the bacterial cell, therefore genetic material of phage is nucleic acid not protein (Hershey and Chase 1952) (Figure 1.5)
- DNA and protein sequencing combined with mutational analysis to alter DNA bases in bacteria demonstrates that order of bases in DNA defines order of amino acids in protein
- Classical genetics—mutations in DNA leading to changes in phenotype
- Genome sequencing projects including the EST (expressed sequence tag) database and specific projects to sequence the DNA from individual species (e.g. Human Genome Project) relate DNA sequence to product
- Fluorescence *in situ* hybridization (FISH) demonstrates presence and location of specific stretches of DNA on chromosomes, e.g. telomere-specific probes
- Density gradient ultracentrifugation can be used to purify repetitive elements of particular density, demonstrating that they exist as repeats rather than as single copies

**Exam tip**

In questions about DNA structure, make sure you draw a clear, correctly labelled diagram of DNA—see Figure 1.3.
RNA

'rough' avirulent pneumococcus

R

Mouse lives

'Smooth' virulent pneumococcus isolated from dead mouse

S

'smooth' virulent pneumococcus

S

Mouse dies

rough + heat-killed smooth pneumococcus

R +

Mouse dies

'Smooth' virulent pneumococcus isolated from dead mouse

S

Figure 1.4 Transforming principle (Avery). Mice were infected with one of two strains of pneumococcus bacteria: The rough 'R' form is non-lethal but the smooth 'S' strain kills the mice. It is possible to heat-kill the 'S' bacteria (proteins are denatured, DNA remains intact); these dead bacteria do not cause illness in mice. However, if a mouse is infected with both the heat-treated S form and the non-pathogenic R form, the R strain of the bacteria can take up DNA from the dead S strain and acquire their pathogenic properties, and so the mice die. The bacteria isolated from the dead mouse are now the virulent 'S' strain.

1.6 RNA

RNA (ribonucleic acid) is a single-stranded polymer of ribonucleotides linked by phosphodiester bonds. Some viruses have an RNA genome (mammalian retroviruses, e.g. human immunodeficiency virus). RNA can have structural, regulatory, and catalytic roles as well as acting as the intermediate between DNA and proteins. RNA is made during the process of transcription, and is complementary to the sequence of the DNA from which it is transcribed.
Chemical composition of RNA

- **Ribonucleosides** = base + sugar, i.e. nitrogenous bases linked to the C1 carbon of ribose by a β–N-glycosidic bond (note deoxyribose in DNA)
- **Ribonucleotides** = base + sugar + phosphate, i.e. nucleoside with the 5′ carbon of ribose bound by an ester linkage to one or more phosphate groups. rNTPs act as substrates for RNA synthesis during transcription
  
  ![Diagram of RNA structure](image)

  **Figure 1.5** Genetic material is nucleic acid (Hershey and Chase). Bacteriophages were radioactively labelled with \(^{35}\text{S}\) in the protein and \(^{32}\text{P}\) in the nucleic acid. Only the \(^{32}\text{P}\) entered the cell and was incorporated into new phage particles. Therefore the genetic material is nucleic acid, not protein.

- General structure of a ribonucleotide (rNTP, aka NTP):

  ![Ribonucleotide structure]

  ![Diagram of RNA structure](image)
RNA

- Bases in RNA:
  i. pyrimidines (C and U) have a single nitrogenous ring
  ii. purines (G and A) have double nitrogenous ring structure
      (see 1.5 DNA (p. 20))
  iii. structure of uracil (U):

```
O
\H\NH
\NH\\N\\O
```

Key features of ribonucleic acid (RNA)

- Polymer of ribonucleotides linked by phosphodiester bonds
- RNA is composed of four different types of nitrogenous base linked by a glycosidic bond to the C1 of ribose (compared to deoxyribose in DNA)
  i. sugar in RNA has OH at 2’ position as well as 3’ position seen in DNA—makes RNA more chemically reactive than DNA
  ii. nitrogenous bases of RNA: adenine (A), cytosine (C), guanine (G), and uracil (U) (instead of thymine in DNA)
- Polymer has polarity, with a triphosphate on the 5’ carbon of the ribose at one end and a free OH group on the 3’ carbon of the other terminal nucleotide—always drawn 5’→3’ by convention
- RNAs are single-stranded but in physiological solution can adopt various secondary structures according to sequence
- Palindromic regions of complementary base pairs can form hairpin loops, e.g. stem–loop structure of tRNA molecules and termination loop during transcription.
- Double-stranded RNA (e.g. some viruses) resembles ‘A form’ double-stranded DNA

Key classes of RNA

- rRNA
  ○ ribosomal RNA
  ○ functions as structural and enzymatic component of ribosomes
  ○ 80% of total cellular RNA
- mRNA
  ○ messenger RNA
  ○ acts as intermediate between genetic code of DNA and amino acid sequence of proteins
3% of total cellular RNA

• tRNA
  • transfer RNA
  • adapter molecule that carries amino acids to specific codon on mRNA in ribosome during protein synthesis
  • 73–95 nucleotides long
  • 5% total cellular RNA

• snRNA
  • small nuclear RNA, in eukaryotes
  • involved in removal or splicing of introns from eukaryotic mRNA
  • catalytic or associated with catalytic proteins

• miRNA
  • small RNA (∼22 nucleotides), often complementary to untranslated region of mRNA
  • involved in regulating gene expression
  • ∼1,000 in human genome, regulating ∼60% of genes

• piRNA
  • piwi-interacting RNA
  • 26–31 nucleotides long
  • largest class of mammalian small RNAs
  • role in transposon and gene silencing

• siRNA
  • short interfering double-stranded RNA molecules
  • usually 21 bases long with an overhang of two at each 3’ terminus
  • recruits RNA-induced silencing complex (RISC) for mRNA cleavage and gene silencing
  • used experimentally to knock down specific gene expression

Check your understanding

Compare the chemical nature and structure of DNA and RNA. (Hint: remember RNA uses uracil instead of thymidine; the extra OH group at the 2’ position in RNA increases its reactivity; higher order structures due to base pairing.)

How does the primary sequence of a protein define its final structure? (Hint: remember all four levels of protein structure and the different bonding types involved; give examples of globular and structural proteins as well as membrane proteins.)

Describe the structure of three polymers found in cells. (Hint: make sure you chose polymers that you can draw and describe in adequate detail.)

What is the structure of glucose and its naturally occurring polymers? (Hint: remember to include starch and glycogen.)