Introductory Chemistry

for Molecular Biology & DNA Computing

A Study Guide for the Experienced CS Programmer

Covering:

- Chemical notation & reading molecular formulas
- Organic chemistry essentials: functional groups & bonds
- Polymers: what they are and how they form
- Acids, bases, and pH
- The 5′ and 3′ ends of DNA
- Key molecules in molecular biology: ATP, NADH, phosphate groups
- Water, hydrogen bonds, and why biology is wet

Prerequisite:	High school / early college chemistry
Goal:	Read biochemistry literature without notation confusion
Companion to:	Biochemistry & Molecular Biology Study Guide (v1)

All chemical formulas typeset in proper notation. Equations are balanced. CS analogies throughout.

Contents

Module 1 — Reading Chemical Formulas and Notation

Elemental Symbols: The Alphabet of Chemistry

Every element has a 1- or 2-letter symbol. The ones you will encounter constantly in molecular biology:

Symbol	Element	Atomic #	Why it matters in biology
H	Hydrogen	1	Everywhere; H-bonds; protons; pH
C	Carbon	6	Backbone of all organic molecules
N	Nitrogen	7	Bases, amino acids, ATP
O	Oxygen	8	Phosphate groups, hydroxyl groups, water
P	Phosphorus	15	DNA/RNA backbone, ATP, phospholipids
S	Sulfur	16	Disulfide bonds in proteins, cysteine
Na	Sodium	11	Electrolyte; stabilizes DNA (shields phosphate)
K	Potassium	19	Electrolyte; membrane potential
Mg	Magnesium	12	Enzyme cofactor; stabilizes nucleic acids
Ca	Calcium	20	Signaling ion; enzyme cofactor
Cl	Chlorine	17	Counterion; buffer chemistry
Fe	Iron	26	Heme groups; electron transport chain
Zn	Zinc	30	Zinc finger protein domains; enzyme active sites

Molecular Formulas

A **molecular formula** lists the atoms in one molecule with subscript counts.

Formula	Name	Meaning
H_2O	Water	2 H atoms, 1 O atom
CO_2	Carbon dioxide	1 C atom, 2 O atoms
O_2	Molecular oxygen	2 O atoms (diatomic)
NaCl	Sodium chloride	1 Na, 1 Cl (ionic compound)
H_3PO_4	Phosphoric acid	3 H, 1 P, 4 O
NH_3	Ammonia	1 N, 3 H
$C_6H_{12}O_6$	Glucose	6 C, 12 H, 6 O

Structural Formulas and Line Notation

Molecular formulas hide structure. Biochemistry constantly uses **structural** or **skeletal** notation:

Skeletal (line-angle) notation rules:

- Each vertex or endpoint of a line = one carbon atom (C is implied, not written)
- Hydrogen atoms on carbons are *not* shown (also implied)
- Heteroatoms (N, O, S, P) *are* written explicitly
- Each line = one bond

- Double line = double bond; triple line = triple bond

CS Analogy: Skeletal notation is like compressed code. Carbon atoms and their hydrogens are "whitespace" — implied and omitted to reduce visual noise. Only the "non-default" atoms (N, O, S, P) are written explicitly. Reading a structural formula is like reading minified code; once you learn the conventions, you decompress it automatically.

Charges and Ions

Superscripts indicate charge:

$$\mathrm{Na^+} \quad \mathrm{Cl^-} \quad \mathrm{Ca^{2+}} \quad \mathrm{PO_4^{3-}} \quad \mathrm{NH_4^+} \quad \mathrm{OH^-} \quad \mathrm{H^+}$$

Cations = positive ions (lost electrons). **Anions** = negative ions (gained electrons).

At physiological pH (7.4), the DNA backbone phosphate group carries a full negative charge: $-\mathrm{OPO_3^{2-}}$. This is why DNA is negative and migrates toward the positive electrode in gel electrophoresis.

Balancing Equations

Chemical equations must be balanced: same atoms on both sides.

$$2\,\mathrm{H_2} + \mathrm{O_2} \longrightarrow 2\,\mathrm{H_2O}$$

Read: "2 molecules of hydrogen gas react with 1 molecule of oxygen to produce 2 molecules of water." The coefficients (2, 1, 2) balance the equation. You will see balanced equations for enzyme reactions, ATP hydrolysis, and polymerization throughout molecular biology.

Self-Check: Module 1

1. How many atoms total are in one molecule of glucose ($\mathrm{C_6H_{12}O_6}$)?
2. What charge does a phosphate group carry at neutral pH? Write its ionic formula.
3. In the balanced equation $\mathrm{C_6H_{12}O_6 + 6\,O_2 \longrightarrow 6\,CO_2 + 6\,H_2O}$ (cellular respiration), count atoms on each side to verify it is balanced.
4. Why is carbon the backbone of organic molecules rather than, say, silicon? (Hint: think valence electrons.)

Module 2 — Organic Chemistry: Bonds and Functional Groups

Carbon's Bonding Rules

Carbon has 4 valence electrons and forms exactly 4 bonds. This gives it extraordinary versatility:

- **Single bond** (C–C): 2 electrons shared; free rotation
- **Double bond** (C=C): 4 electrons shared; rigid, planar
- **Triple bond** (C≡C): 6 electrons shared; linear; rare in biology
- Carbon bonds to H, N, O, S, P, and other carbons

Carbon chains and rings form the skeleton. Everything attached to the skeleton = functional groups.

The Essential Functional Groups

These are the "side chains" that determine how a molecule behaves chemically. You will see all of these constantly.

Group name	Structure	Symbol	Role in molecular biology
Hydroxyl	–OH	–OH	Polar; H-bond donor/acceptor; sugars, serine, threonine
Carbonyl	C=O	C=O	Aldehydes (–CHO) and ketones; sugars
Carboxyl	–COOH	–COOH or –COO–	Weak acid; deprotonated at pH 7; amino acid C-terminus
Amino	$–NH_2$	$–NH_2$ or $–NH_3^+$	Weak base; protonated at pH 7; amino acid N-terminus; DNA bases
Phosphate	$–OPO_3^{2-}$	$–OPO_3{}^{2-}$	Strongly acidic; DNA/RNA backbone; ATP energy currency
Thiol	–SH	–SH	Disulfide bonds in proteins; cysteine side chain
Methyl	$–CH_3$	$–CH_3$	Nonpolar; epigenetic methylation; lipid tails
Ester	–COO–	–COO–	Links fatty acids to glycerol; formed in polymerization
Amide	–CO–NH–	–CONH–	Peptide bond; links amino acids in proteins
Ether	–O–	–O–	Links sugars in polysaccharides

Critical distinction: A **carboxyl group** (–COOH) is an acid (donates H^+). An **amino group** ($-NH_2$) is a base (accepts H^+). Amino acids have *both* on the same molecule — this is why they are amphoteric (can act as acid or base).

Condensation (Dehydration) Reactions — How Polymers Form

The universal reaction for building biological polymers is the **condensation reaction**: two monomers join, releasing one molecule of water (H_2O).

$$\text{Monomer-OH} + \text{H-Monomer} \longrightarrow \text{Monomer–Monomer} + H_2O$$

The reverse reaction — **hydrolysis** — breaks polymers by adding water:

$$\text{Polymer} + H_2O \longrightarrow \text{Monomer-OH} + \text{H-Monomer}$$

CS Analogy: Condensation polymerization is like linked list construction. Each monomer has a "left pointer" (–OH) and a "right pointer" (H–). Linking them together ejects a water molecule (the pointer bookkeeping overhead). Hydrolysis is like `free()` — reinserts water to disconnect the link.

Types of Bonds in Biology: Strength Hierarchy

Bond type	Strength (kJ/mol)	Broken by	Example
Covalent (single)	200–400	Enzymes / heat	C–C, C–O, C–N, P–O
Covalent (double)	400–700	Enzymes	C=O, C=C
Ionic	20–40	Salt (water)	Na^+ and Cl^-
Hydrogen bond	12–30	Mild heat	A:T (DNA), G:C (DNA), protein folding
Van der Waals	2–4	Thermal motion	Base stacking in DNA
Hydrophobic	Variable	Water	Lipid bilayer interior

Key insight: Individual hydrogen bonds are weak, but *thousands* cooperatively stabilize DNA duplexes and protein structures. This cooperativity is why melting is a sharp transition — it is all-or-nothing at the ensemble level.

Isomers and Chirality

Two molecules with the same molecular formula but different structures are **isomers**. Biologically critical:

- **Structural isomers**: Different connectivity (glucose vs. fructose: both $C_6H_{12}O_6$)
- **Stereoisomers**: Same connectivity, different 3D arrangement
- **Enantiomers**: Mirror images (L- vs. D-amino acids; D-deoxyribose in DNA)

Biology is almost exclusively **L-amino acids** and **D-sugars**. Enzymes are exquisitely stereospecific — a D-amino acid won't fit an active site built for L-amino acids.

Self-Check: Module 2

1. Identify the functional groups in this description of serine: "an amino acid with a hydroxyl group on its side chain, an amino group at its N-terminus, and a carboxyl group at its C-terminus."
2. Write the balanced condensation reaction that joins two amino acids (generic: $H_2N–CHR–COOH$) to form a dipeptide. What bond is formed?
3. Why does double-bond character in a C=O carbonyl group make the carbon atom electrophilic (electron-hungry)?
4. A drug is synthesized as a racemic mixture (50/50 L and D enantiomers). Why might only one enantiomer be therapeutically active?

Module 3 — Polymers: The Four Classes of Biological Macromolecules

What is a Polymer?

A **polymer** is a large molecule built by covalently linking many smaller units called **monomers**. The linkage reaction is almost always condensation (losing H_2O).

The four biological polymer classes:

Polymer	Monomer	Bond name	Example	Function
Nucleic acid	Nucleotide	Phosphodiester	DNA, RNA	Information storage
Protein	Amino acid	Peptide (amide)	Enzymes, antibodies	Catalysis, structure
Polysaccharide	Monosaccharide	Glycosidic	Cellulose, glycogen	Energy, structure
Lipid (partial)	Fatty acid + glycerol	Ester	Phospholipid	Membranes

Nucleotides and Nucleic Acid Polymers

A **nucleotide** = sugar + phosphate + base. Three varieties:

Type	Sugar	Abbreviation	Found in
Deoxyribonucleotide	Deoxyribose ($C_5H_{10}O_4$)	dNMP, dNDP, dNTP	DNA
Ribonucleotide	Ribose ($C_5H_{10}O_5$)	NMP, NDP, NTP	RNA, ATP, NADH

The difference between deoxyribose and ribose is exactly **one oxygen atom**: ribose has an –OH at the 2′ position; deoxyribose has just –H. This single difference has enormous consequences: RNA is far more reactive (and less stable) than DNA.

The Phosphodiester Bond

DNA polymerization joins the 3′–OH of one nucleotide to the 5′–phosphate of the next, releasing pyrophosphate (PP_i):

$$\cdots\text{–sugar–}\underbrace{\text{O–P(O)(OH)–O}}_{\text{phosphodiester}}\text{–sugar–}\cdots + PP_i$$

The resulting linkage is called a **phosphodiester bond** because the phosphate group forms two ester bonds (one to each sugar).

Nucleotide shorthand:

Abbreviation	Full name	Notes
dAMP / dADP / dATP	Deoxyadenosine mono/di/triphosphate	Substrate for DNA pol
dTMP / dTDP / dTTP	Deoxythymidine mono/di/triphosphate	
dGMP / dGDP / dGTP	Deoxyguanosine mono/di/triphosphate	
dCMP / dCDP / dCTP	Deoxycytidine mono/di/triphosphate	
ATP	Adenosine triphosphate	Energy currency; also RNA monomer
GTP	Guanosine triphosphate	Signal transduction; RNA monomer
NAD^+ / NADH	Nicotinamide adenine dinucleotide	Electron carrier in metabolism

Amino Acids and Proteins

Amino acids are the monomers of proteins. The general structure:

$$H_2N{-}CH(R){-}COOH$$

where R = the side chain ("R group") that distinguishes the 20 standard amino acids.

The **peptide bond** (amide bond) forms between the carboxyl of one amino acid and the amino group of the next:

$$-CO-NH-$$

Peptide bond is planar and rigid because of partial double-bond character. This constrains protein backbone geometry and is fundamental to protein structure prediction (AlphaFold models this explicitly).

Amino Acid Side Chain Chemistry Categories

Category	Side chain character	Examples
Nonpolar / hydrophobic	Alkyl / aromatic; no charge	Ala (A), Val (V), Leu (L), Phe (F)
Polar uncharged	–OH, –SH, $–CONH_2$; H-bond capable	Ser (S), Thr (T), Cys (C), Asn (N)
Positively charged (basic)	$–NH_3^+$ or guanidinium at pH 7	Lys (K), Arg (R), His (H)
Negatively charged (acidic)	$–COO^–$ at pH 7	Asp (D), Glu (E)

Proteins are written N-terminus to C-terminus using single-letter or three-letter codes: `M-A-S-K-...` or `Met-Ala-Ser-Lys-...`

Sugars and Polysaccharides

Monosaccharides: simple sugars. Most biologically important are 5-carbon (**pentoses**) and 6-carbon (**hexoses**).

Sugar	Formula	Type	Role
Glucose	$C_6H_{12}O_6$	Hexose	Primary energy source
Fructose	$C_6H_{12}O_6$	Hexose	Isomer of glucose; fruit sugar
Ribose	$C_5H_{10}O_5$	Pentose	RNA backbone
Deoxyribose	$C_5H_{10}O_4$	Pentose	DNA backbone
Galactose	$C_6H_{12}O_6$	Hexose	Cell surface glycoproteins

Sugars join via **glycosidic bonds** (a type of ether linkage: –O–) with loss of water. The naming (α or β, 1→4 or 1→6) specifies which carbons link and the stereochemistry of the bond.

CS Analogy: A polymer is a singly-linked list. Each monomer is a node. The bond (phosphodiester, peptide, glycosidic) is the pointer. The head of the list has a free end (5′–phosphate in DNA, N-terminus in protein). The tail has the other free end (3′–OH in DNA, C-terminus in protein). Directionality is always defined and always matters.

Self-Check: Module 3

1. What single atom distinguishes ribose from deoxyribose, and at which carbon position?
2. A protein of 300 amino acids was formed by 299 condensation reactions. How many water molecules were released during synthesis?
3. What is the chemical linkage in a *peptide bond*? Write its structural formula.
4. ATP is described as the "energy currency" of the cell. What reaction releases energy from ATP, and approximately how much free energy ($\Delta G^{\circ\prime}$) does it release?

Module 4 — Acids, Bases, and pH

The Brønsted–Lowry Definitions

- **Acid**: Proton (H^+) *donor*. Example: $HCl \longrightarrow H^+ + Cl^-$
- **Base**: Proton (H^+) *acceptor*. Example: $NH_3 + H^+ \longrightarrow NH_4^+$
- **Conjugate base**: What remains after an acid donates its proton
- **Conjugate acid**: What forms after a base accepts a proton

Water is **amphoteric** — it can act as both acid and base:

$$H_2O + H_2O \rightleftharpoons H_3O^+ + OH^-$$

The equilibrium constant for water autoionization: $K_w = [H^+][OH^-] = 1.0 \times 10^{-14}$ at 25°C.

The pH Scale

$$\boxed{\text{pH} = -\log_{10}[H^+]}$$

pH	$[H^+]$ (mol/L)	Example
0	$1 \times 10^0 = 1.0$	Stomach acid (HCl)
2	1×10^{-2}	Lemon juice
4	1×10^{-4}	Black coffee
7	1×10^{-7}	Pure water (neutral)
7.4	4×10^{-8}	**Human blood / physiological pH**
8	1×10^{-8}	Seawater
10	1×10^{-10}	Baking soda
14	1×10^{-14}	Concentrated NaOH

Each pH unit = 10-fold change in $[H^+]$. pH 6 is 10× more acidic than pH 7; pH 5 is 100× more acidic.

Strong vs. Weak Acids and Bases

Strong acids/bases dissociate completely:

$$HCl \longrightarrow H^+ + Cl^- \qquad NaOH \longrightarrow Na^+ + OH^-$$

Weak acids/bases partially dissociate. Described by an equilibrium constant K_a:

$$HA \rightleftharpoons H^+ + A^- \quad K_a = \frac{[H^+][A^-]}{[HA]}$$

$$\text{p}K_a = -\log_{10}(K_a)$$

Lower $\text{p}K_a$ = stronger acid (more dissociated). Biologically critical $\text{p}K_a$ values:

Group / molecule	pK_a	State at pH 7.4
Phosphoric acid H_3PO_4 (1st)	2.1	Fully deprotonated
Carboxyl group (–COOH)	4–5	Deprotonated: $-COO^-$
Phosphate ester (DNA backbone)	∼1	Fully deprotonated: $-OPO_3^{2-}$
Imidazole (histidine side chain)	6.0	∼50% protonated
Amino group ($-NH_2$)	9–10	Protonated: $-NH_3^+$
Guanidinium (arginine)	12.5	Fully protonated
Water (H_2O)	15.7	Exists as water

The Henderson-Hasselbalch Equation

For any weak acid equilibrium, the relationship between pH and the ratio of protonated/deprotonated forms:

$$\boxed{\text{pH} = \text{p}K_a + \log_{10}\left(\frac{[\text{A}^-]}{[\text{HA}]}\right)}$$

Key consequence: When pH = pK_a, the acid is exactly 50% dissociated. This is the center of buffering capacity.

Buffers

A **buffer** resists changes in pH when acid or base is added. It consists of a weak acid and its conjugate base in significant concentrations.

The most important biological buffers:

Buffer system	Components	Useful pH range
Bicarbonate (blood)	H_2CO_3/HCO_3^-	6.1–7.1
Phosphate (intracellular)	$H_2PO_4^-/HPO_4^{2-}$	6.2–8.2
Tris (lab buffer)	Tris-H^+/Tris	7.0–9.0
TAE (gel buffer)	Tris-acetate/EDTA	∼8.0
TBE (gel buffer)	Tris-borate/EDTA	∼8.3
HEPES (cell culture)	HEPES-H^+/HEPES	6.8–8.2

EDTA (ethylenediaminetetraacetic acid) appears in nearly every molecular biology buffer. It chelates (tightly binds) divalent cations like Mg^{2+} and Ca^{2+}, thereby inhibiting nucleases (which need these metal ions to function). When you see "10 mM EDTA" in a buffer recipe, it is there to *inactivate* degrading enzymes and protect your DNA.

Why pH Matters for DNA

- At physiological pH (7.4), **all** DNA backbone phosphates are fully deprotonated ($-OPO_3^{2-}$), giving DNA its characteristic negative charge
- Extreme pH (<4 or >10) can **depurinate** DNA (cleave bases from the backbone) or hydrolyze phosphodiester bonds

- The pKa of the nitrogenous bases shifts upon base pairing; protonation/deprotonation of bases at non-physiological pH disrupts Watson-Crick H-bonding and melts duplexes
- Gel electrophoresis is run at pH 8–8.3 (TAE/TBE buffers) to keep DNA fully charged for migration

Self-Check: Module 4

1. Calculate the $[H^+]$ at physiological pH of 7.4. Is blood acidic, neutral, or basic?
2. A weak acid has $pK_a = 6.0$ (like histidine's imidazole). Using Henderson-Hasselbalch, what fraction is in the deprotonated (basic) form at pH 7.4?
3. TAE buffer is made from Tris base, acetic acid, and EDTA. What is the role of each component?
4. If you accidentally omit EDTA from your DNA extraction buffer, what risk do you introduce, and why?

Module 5 — The 5′ and 3′ Ends of DNA and RNA

Carbon Numbering of the Sugar

The 5′ and 3′ notation refers to specific carbon atoms on the deoxyribose sugar. The ring carbons of deoxyribose are numbered 1′ through 5′ (the prime ′ distinguishes them from the base's ring carbons):

Deoxyribose carbon positions and what's attached:

Carbon	Attached to	Role
1′	Nitrogenous base (N-glycosidic bond)	Base attachment site
2′	–H only (deoxyribose) or –OH (ribose)	Distinguishes DNA from RNA
3′	–OH (free end) or phosphodiester bond	**3′ end; chain elongation here**
4′	Part of the ring oxygen	Ring closure
5′	Phosphate group(s) or –OH	**5′ end; free phosphate here**

What the 5′ End Looks Like

The 5′ end of a DNA strand has a **free phosphate group** on the 5′ carbon:

$$^{-}\mathrm{O}-\overset{5'}{P}(=\mathrm{O})(-\mathrm{O}^{-})-\mathrm{O}-\mathrm{CH_2}-[\text{sugar}]-$$

This is written shorthand as **5′-p** or simply described as having a 5′ phosphate. In many lab contexts, the 5′ end is **phosphorylated** (carries a phosphate, enabling ligation) or **dephosphorylated** (–OH only; cannot be directly ligated without phosphorylation by a kinase).

What the 3′ End Looks Like

The 3′ end of a DNA strand has a **free hydroxyl group** (–OH) on the 3′ carbon:

$$-[\text{sugar}]-\overset{3'}{C}\mathrm{H}-\mathrm{OH}$$

This free 3′–OH is the critical site for:

- DNA polymerase extension (adds next nucleotide to 3′–OH)
- Ligation (ligase seals the nick between 3′–OH and 5′-phosphate)
- 3′ exonuclease activity (proofreading digests from 3′ end)

Directionality: Antiparallel Strands

A DNA duplex always has two strands running in opposite directions:

$$5'\text{–}\underbrace{\mathrm{ATCGGATC}}_{\text{top strand}}\text{–}3'$$

$$3'\text{–}\underbrace{\mathrm{TAGCCTAG}}_{\text{bottom strand}}\text{–}5'$$

Reading both strands **left to right**, the top strand goes $5' \to 3'$ and the bottom strand goes $3' \to 5'$. The conventional representation always writes the top strand $5' \to 3'$ left to right.

CS Analogy: Think of a DNA duplex as a double-ended queue (deque) with two separate read-directions. The $5'$ end is the "head" from which sequence is conventionally read. The $3'$ end is the "tail" to which new monomers are appended during synthesis. Every enzyme that acts on DNA cares deeply about which end it is at, exactly as a parser cares about the start of a string.

Why Directionality Controls Every Enzyme

Enzyme	Direction of action	Consequence
DNA polymerase	Synthesizes $5' \to 3'$; reads template $3' \to 5'$	Lagging strand requires Okaza
Exonuclease I	Degrades $3' \to 5'$	Used to remove primers and si
Exonuclease III	Degrades $3' \to 5'$ on dsDNA	Creates single-stranded overha
Lambda exonuclease	Degrades $5' \to 3'$ on phosphorylated end	Generates ssDNA from dsDNA
T4 polynucleotide kinase (PNK)	Phosphorylates $5'$–OH	Prepares DNA for ligation
Alkaline phosphatase (CIP/SAP)	Removes $5'$ phosphate	Prevents self-ligation of vectors

$5'$ and $3'$ Overhangs (Sticky Ends Revisited)

When restriction enzymes cut staggered, they leave single-stranded overhangs:

- **$5'$ overhang** ($5'$ extension): The top strand extends further than the bottom. Most common (*EcoRI*, *BamHI*, *HindIII*).
- **$3'$ overhang** ($3'$ extension): The bottom strand extends further. Less common (*PstI*, *KpnI*).
- **Blunt ends**: No overhang (*SmaI*, *EcoRV*).

For *EcoRI* ($5'$ overhang, 4 nt):

$$5'\text{–G}\quad \underbrace{\text{AATTC}}_{\text{5' overhang}}\ \text{–}3'$$

$$3'\text{–CTTAA}\quad \text{G–}5'$$

RNA vs. DNA Ends: The Critical $2'$–OH Difference

RNA has a $2'$–OH group that DNA lacks. This has two critical consequences:

1. RNA is **chemically less stable**: the $2'$–OH participates in self-cleavage reactions, which is why RNA degrades much faster than DNA under basic conditions
2. The $2'$–OH enables **ribozyme** activity: RNA can catalyze its own cleavage (used in CRISPR RNA processing, spliceosomes)

Self-Check: Module 5

1. You synthesize a PCR product. Which end of each newly synthesized strand was the first nucleotide incorporated? Which end was the last?

2. A DNA strand has no 5′ phosphate (only 5′–OH). Can it be directly ligated to another strand? What enzyme would you use to fix this?
3. Write the antiparallel complement of `5'-GCATTAGC-3'` in the correct orientation.
4. Why does RNA degrade much faster than DNA in alkaline conditions? Name the specific chemical group responsible.

Module 6 — Water, Hydrogen Bonds, and Key Biological Molecules

Water: Why Biology is Aqueous

Water (H_2O) is not a passive solvent — it actively participates in molecular interactions.

Properties arising from water's polarity and H-bonding:

Property	Biological consequence
High specific heat	Buffers temperature changes in cells
High heat of vaporization	Evaporative cooling (sweating)
Cohesion / surface tension	Capillary action in vascular systems
Universal polar solvent	Dissolves ions, sugars, nucleotides, proteins
Hydrogen bond donor/acceptor	Stabilizes DNA duplexes, protein folds
Dielectric constant ($\varepsilon \approx 80$)	Screens electrostatic interactions; weakens ionic bonds

Hydrogen Bond Donors and Acceptors

A **hydrogen bond** forms between:

- A **donor**: an electronegative atom (N or O) covalently bonded to H, which is partially positive
- An **acceptor**: a lone pair on another electronegative atom (N, O, occasionally F)

$$\text{Donor: } \underbrace{\text{N–H}}_{\text{or O–H}} \cdots \underbrace{\text{:O}}_{\text{or :N}} \quad \text{(Acceptor)}$$

Watson-Crick base pairs:

- A:T pair: 2 hydrogen bonds (A donates 1 N–H; T donates 1 N–H)
- G:C pair: 3 hydrogen bonds (G donates 2 N–H; C donates 1 N–H)

Hydrophobic and Hydrophilic

- **Hydrophilic** ("water-loving"): Polar or charged molecules; dissolve readily in water (sugars, DNA, amino acids with polar side chains)
- **Hydrophobic** ("water-fearing"): Nonpolar molecules; excluded from water's H-bond network; aggregate together (lipid tails, aromatic bases in DNA, hydrophobic protein core)
- **Amphipathic**: Has both hydrophilic and hydrophobic regions (phospholipids, detergents — essential for membrane structure)

Base stacking in DNA is partly hydrophobic: the planar aromatic bases prefer to stack face-to-face and exclude water, contributing significantly to duplex stability.

ATP: Structure and Function

Adenosine triphosphate (ATP) is the primary energy currency of cells.

Structure: Adenine base + ribose + **three phosphate groups** (α, β, γ):

$$\text{Adenosine–O–}\underbrace{\text{PO}_3^{2-}}_{\alpha}\text{–O–}\underbrace{\text{PO}_3^{2-}}_{\beta}\text{–O–}\underbrace{\text{PO}_3^{2-}}_{\gamma}$$

Hydrolysis reactions:

$$\mathrm{ATP} + \mathrm{H_2O} \longrightarrow \mathrm{ADP} + \mathrm{Pi} \quad \Delta G^{\circ\prime} \approx -30.5\ \mathrm{kJ/mol}$$

$$\mathrm{ATP} + \mathrm{H_2O} \longrightarrow \mathrm{AMP} + \mathrm{PPi} \quad \Delta G^{\circ\prime} \approx -45.6\ \mathrm{kJ/mol}\ (\mathrm{PPi\ is\ further\ hydrolyzed})$$

The large negative ΔG arises from:

- Electrostatic repulsion between adjacent negatively charged phosphates (relieved on hydrolysis)
- Resonance stabilization of the products (ADP and Pi)
- Increased entropy (one molecule becomes two)

NADH and Electron Carriers

$\mathrm{NAD^+}$ / NADH (Nicotinamide Adenine Dinucleotide) carries electrons in metabolism:

$$\mathrm{NAD^+} + 2\,\mathrm{H^+} + 2\,\mathrm{e^-} \rightleftharpoons \mathrm{NADH} + \mathrm{H^+}$$

- **$\mathrm{NAD^+}$**: oxidized form (electron acceptor)
- **NADH**: reduced form (electron donor; carries "high energy" electrons)
- Used as cofactor by many dehydrogenase enzymes
- Oxidized back to $\mathrm{NAD^+}$ in the electron transport chain, generating ATP

CS Analogy: $\mathrm{NAD^+}$/NADH is a rechargeable battery token. Metabolic reactions charge it ($\mathrm{NAD^+} \rightarrow$ NADH). The electron transport chain drains it (NADH $\rightarrow \mathrm{NAD^+}$) and uses the energy to pump protons and make ATP. The cycle is a producer-consumer queue with ATP as the product.

The Phosphate Group: Biology's Most Important Functional Group

The phosphate group ($-\mathrm{OPO_3^{2-}}$) appears in:

- DNA and RNA backbone (phosphodiester bonds)
- ATP, ADP, GTP (energy and signaling)
- Phospholipids (membranes)
- Phosphorylated proteins (signal transduction; *kinases* add phosphate, *phosphatases* remove it)
- NADH, FAD, CoA (metabolic cofactors)

Why is phosphate so central? At physiological pH, it carries 2 negative charges. This makes phosphorylated molecules membrane-impermeable (charged, can't cross lipid bilayer), soluble (charged, interacts with water), and reactive (high energy when part of anhydride linkages like in ATP).

Common Buffer and Reagent Abbreviations

Abbreviation	Full name	Use
TAE	Tris-Acetate-EDTA	Gel electrophoresis buffer (DNA)
TBE	Tris-Borate-EDTA	Gel electrophoresis buffer (DNA/RNA)
PBS	Phosphate-Buffered Saline	Cell culture, Western blot
TE	Tris-EDTA	DNA storage buffer
DTT	Dithiothreitol	Reduces disulfide bonds; protects thiols
DMSO	Dimethyl sulfoxide	Cryoprotectant; PCR additive
SDS	Sodium Dodecyl Sulfate	Denaturing detergent; SDS-PAGE
EDTA	Ethylenediaminetetraacetic acid	Chelates Mg^{2+}, Ca^{2+}; inhibits nucleases
Tris	Tris(hydroxymethyl)aminomethane	Most common biological buffer base

Self-Check: Module 6

1. ATP hydrolysis ($\Delta G^{\circ\prime} \approx -30.5$ kJ/mol) is coupled to an endergonic reaction with $\Delta G^{\circ\prime} = +22$ kJ/mol. What is the net $\Delta G^{\circ\prime}$ of the coupled reaction? Is it spontaneous?
2. Explain why DNA is soluble in water at physiological pH but RNA degrades faster than DNA in alkaline (high pH) conditions.
3. What does EDTA do to a restriction enzyme reaction, and why would you add it to stop a digest?
4. A kinase adds a phosphate to a protein; a phosphatase removes it. Why would phosphorylation change a protein's activity? (Think: charge, shape, interactions.)

Master Reference: Chemical Symbols and Abbreviations

Inorganic and Small Molecules

Formula	Name	Biological role
H_2O	Water	Universal solvent
O_2	Molecular oxygen	Electron acceptor; aerobic respiration
CO_2	Carbon dioxide	Waste product; pH regulation
H^+	Proton	pH; drives ATP synthase
OH^-	Hydroxide ion	Base; pH
H_2CO_3	Carbonic acid	Blood pH buffer
HCO_3^-	Bicarbonate	Blood buffer; most abundant anion in blood
H_3PO_4	Phosphoric acid	Source of phosphate groups
$H_2PO_4^-$	Dihydrogen phosphate	Phosphate buffer (acidic form)
HPO_4^{2-}	Hydrogen phosphate	Phosphate buffer (basic form)
PO_4^{3-}	Phosphate ion	General phosphate
Pi	Inorganic phosphate	Product of ATP/GTP hydrolysis
PPi	Pyrophosphate	Product of NTP polymerization; hydrolysis drives reactions forward
NH_3	Ammonia	Nitrogen waste (converted to urea)
NH_4^+	Ammonium	Conjugate acid of ammonia; p$K_a \approx 9.2$
Na^+	Sodium ion	Electrolyte; DNA charge shielding
K^+	Potassium ion	Electrolyte; enzyme cofactor
Mg^{2+}	Magnesium ion	Nuclease cofactor; ribosome structure
Ca^{2+}	Calcium ion	Signaling; enzyme cofactor
Cl^-	Chloride ion	Electrolyte; counterion
Fe^{2+}	Ferrous iron	Heme; Fenton reaction
Fe^{3+}	Ferric iron	Oxidized iron; electron carrier
Zn^{2+}	Zinc ion	Zinc-finger domains; enzyme active sites

Nucleotides and Nucleic Acid Components

Symbol	Full name	Notes
dATP, dGTP, dCTP, dTTP	Deoxynucleoside triphosphates	DNA polymerase substrates
ATP, GTP, CTP, UTP	Ribonucleoside triphosphates	RNA polymerase substrates
dNTP	Any deoxynucleoside triphosphate	Generic in protocols
NTP	Any ribonucleoside triphosphate	Generic in protocols
dNMP, dNDP	Mono/diphosphate deoxyribonucleosides	Hydrolysis products
cAMP	Cyclic adenosine monophosphate	Second messenger; signaling
cGMP	Cyclic guanosine monophosphate	Signaling
NAD^+ / NADH	Nicotinamide adenine dinucleotide	Electron carrier
$NADP^+$ / NADPH	Nicotinamide adenine dinucleotide phosphate	Anabolic electron carrier
FAD / $FADH_2$	Flavin adenine dinucleotide	Electron carrier (tightly bound)
CoA / Acetyl-CoA	Coenzyme A	Acyl group carrier

Organic Functional Group Shorthand

Symbol	Group name	Structure
R– or R′–	Alkyl/generic side chain	Any carbon-containing group
–OH	Hydroxyl	Oxygen + hydrogen
–COOH, $–COO^-$	Carboxyl, carboxylate	Acidic; p$K_a \approx 4$
$–NH_2$, $–NH_3^+$	Amino, ammonium	Basic; p$K_a \approx 9$
–SH	Thiol	Sulfur + hydrogen; forms disulfides
–S–S–	Disulfide	Covalent; stabilizes proteins
$–CH_3$	Methyl	Nonpolar; methylation marks
–CHO	Aldehyde	Reactive carbonyl at chain end
C=O (internal)	Ketone	Reactive carbonyl mid-chain
–CO–NH–	Amide (peptide bond)	Links amino acids
–CO–O–	Ester	Links fatty acids; esters of phosphate
–O–	Ether	Links sugars (glycosidic)
$–OPO_3^{2-}$	Phosphate ester	DNA/RNA backbone; signaling

4-Week Reading Roadmap

Week	Focus	Assignments
1	Chemical notation, bonds, functional groups	BTS Ch. 1 (chemistry of life); BTS Ch. 2 (water); Khan Academy "Organic chemistry — functional groups" (videos)
2	Polymers: nucleotides, amino acids, sugars	BTS Ch. 4 intro (nucleic acid structure, nucleotide chemistry); BTS Ch. 3 (amino acids and protein structure intro); MBC Ch. 2
3	Acids, bases, pH, buffers	BTS Ch. 2 (water and pH); Henderson-Hasselbalch derivation; make a personal pK_a cheat sheet for all groups in this guide
4	$5'/3'$ ends, ATP, phosphate, key reagents	MBC Ch. 5 intro; any molecular biology protocols manual (Addgene, NEB) — read 3–5 protocol pages and identify every abbreviation from the master reference table

Practical exercise: Download any published DNA computing paper (e.g., Qian & Winfree 2011) and annotate every chemical abbreviation you encounter using this guide. By week 4 you should recognize all of them.

Self-Assessment Answer Key

Module 1: Chemical Formulas

1. Glucose $C_6H_{12}O_6$: $6 + 12 + 6 = 24$ atoms per molecule.
2. Phosphate at neutral pH: HPO_4^{2-} or $H_2PO_4^-$ (mixture near pK_a 6.8–7.2); the DNA backbone phosphate ester is $-OPO_3^{2-}$ (charge -2).
3. $C_6H_{12}O_6 + 6\,O_2 \longrightarrow 6\,CO_2 + 6\,H_2O$: Left side: 6C, 12H, 18O. Right side: 6C, 12H, 6(2) + 6(1) = 18O. Balanced.
4. Carbon has 4 valence electrons and forms 4 strong, stable covalent bonds, enabling chains, rings, and branching. Silicon also has 4 valence electrons but forms weaker bonds with oxygen and cannot form the same diversity of stable structures in aqueous environments.

Module 2: Organic Chemistry

1. Serine contains: a **hydroxyl group** (–OH) on the side chain; an **amino group** ($-NH_2$) at the N-terminus; a **carboxyl group** (–COOH) at the C-terminus.
2. Condensation joining two generic amino acids (R, R′ = side chains):
`H2N-CH(R)-COOH + H2N-CH(R')-COOH` → `H2N-CH(R)-CO-NH-CH(R')-COOH` + H_2O
The bond formed is a **peptide bond** (amide bond: –CO–NH–).
3. In C=O, oxygen's high electronegativity pulls electron density away from the carbon, leaving it partially positively charged (δ^+). This electrophilic carbon is vulnerable to attack by nucleophiles (electron-rich species). This is the basis of nucleophilic addition reactions.
4. Enzymes (and receptors) have chiral active sites. Only the enantiomer with the correct 3D geometry fits the binding pocket. The wrong enantiomer either doesn't bind or binds incorrectly. Example: thalidomide — one enantiomer was therapeutic, the other teratogenic.

Module 3: Polymers

1. The difference is a single **oxygen atom** at the **2′ position**: ribose has 2′–OH; deoxyribose has 2′–H.
2. $n - 1 = 299$ water molecules released for a polymer of $n = 300$ monomers.
3. Peptide bond: $-CO-NH-$. It is an amide bond formed between the carbonyl carbon of one amino acid's carboxyl group and the nitrogen of the next amino acid's amino group.
4. ATP + $H_2O \longrightarrow$ ADP + Pi; $\Delta G^{\circ\prime} \approx -30.5$ kJ/mol. (Alternatively ATP + $H_2O \longrightarrow$ AMP + PPi, $\Delta G^{\circ\prime} \approx -45.6$ kJ/mol when pyrophosphate is subsequently hydrolyzed.)

Module 4: Acids, Bases, pH

1. $[H^+] = 10^{-7.4} \approx 4.0 \times 10^{-8}$ mol/L. Since pH 7.4 > 7, blood is slightly **basic** (alkaline).
2. Henderson-Hasselbalch: $7.4 = 6.0 + \log([A^-]/[HA])$, so $\log([A^-]/[HA]) = 1.4$, giving $[A^-]/[HA] = 10^{1.4} \approx 25$. Deprotonated fraction $= 25/(25+1) \approx 96\%$ deprotonated (basic form).
3. **Tris base**: the buffering agent (weak base); **acetic acid**: conjugate acid partner with Tris to set buffering pH; **EDTA**: chelates Mg^{2+} and Ca^{2+} to inhibit DNases.

4. Without EDTA, Mg^{2+} and other divalent cations remain active. DNases (nucleases) require these metal ions as cofactors. Active DNases will degrade your DNA sample.

Module 5: 5′ and 3′ Ends

1. The **first nucleotide** incorporated defines the **5′ end** (polymerase starts at the primer's 3′–OH and extends 5′→3′, so the primer end is at 5′ and growth is toward 3′). The **last nucleotide** added is at the **3′ end**.
2. No, it cannot be directly ligated. T4 Polynucleotide Kinase (T4 PNK) adds a phosphate group from ATP to the 5′–OH, generating a 5′-phosphate required for ligation.
3. Complement of `5'-GCATTAGC-3'`: first write base-by-base complement: `CGTAATCG`, then reverse: `3'-CGTAATCG-5'`, or written 5′→3′: `5'-GCTAATGC-3'`.
4. The 2′–OH of ribose is a nucleophile. Under alkaline conditions (excess OH^-), the 2′–OH attacks the adjacent phosphate, forming a 2′,3′-cyclic phosphate intermediate and cleaving the backbone. DNA lacks the 2′–OH entirely, so this attack is impossible and DNA is stable under alkaline conditions.

Module 6: Water, H-bonds, Key Molecules

1. Net $\Delta G^{\circ\prime} = -30.5 + 22.0 = -8.5$ kJ/mol. Negative, so the coupled reaction is **spontaneous**.
2. DNA is soluble because its backbone phosphate groups ($-OPO_3{}^{2-}$) are charged at physiological pH and interact favorably with water. RNA degrades in alkali because the 2′–OH nucleophilically attacks the adjacent phosphate, cleaving the backbone; DNA lacks 2′–OH and is thus alkali-stable.
3. EDTA chelates Mg^{2+}, which restriction enzymes require as a catalytic cofactor. Adding EDTA captures the magnesium, inactivates the enzyme, and stops the digest. This is standard practice to terminate a reaction before loading a gel.
4. Phosphorylation adds 2 negative charges to a protein at the modification site. This can: (a) directly alter protein–protein or protein–DNA interactions by charge repulsion/attraction; (b) cause a conformational change exposing or burying an active site; (c) create a phospho-binding docking site for another protein (e.g., SH2 domains bind phosphotyrosine). Kinase/phosphatase cycling is the most common on/off switch in cell signaling.