1 Basic principles

K. WILSON

1.1 BIOCHEMICAL AND MOLECULAR BIOLOGY STUDIES

1.1.1 Aims of laboratory investigations

Biochemistry involves the study of the chemical processes that occur in living organisms with the ultimate aim of understanding the nature of life in molecular terms. Biochemical studies rely on the availability of appropriate analytical techniques and on the application of these techniques to the advancement of knowledge of the nature of, and relationships between, biological molecules, especially proteins and nucleic acids, and cellular function. In recent years huge advances have been made in our understanding of gene structure and expression and in the application of techniques such as mass spectrometry to the study of protein structure and function. The Human Genome Project in particular has been the stimulus for major developments in our understanding of many human diseases especially cancer and for the identification of strategies that might be used to combat these diseases. The discipline of molecular biology overlaps with that of biochemistry and in many respects the aims of the two disciplines complement each other. Molecular biology is focussed on the molecular understanding of the processes of replication, transcription and translation of genetic material whereas biochemistry exploits the techniques and findings of molecular biology to advance our understanding of such cellular processes as cell signalling and apoptosis. The result is that the two disciplines now have the opportunity to address issues such as:

- the structure and function of the total protein component of the cell (*proteomics*) and of all the small molecules in the cell (*metabolomics*);
- the mechanisms involved in the control of gene expression;

- the identification of genes associated with a wide range of human diseases;
- the development of gene therapy strategies for the treatment of human diseases;
- the characterisation of the large number of 'orphan' receptors, whose physiological role and natural agonist are currently unknown, present in the human genome and their exploitation for the development of new therapeutic agents;
- the identification of novel disease-specific markers for the improvement of clinical diagnosis;
- the engineering of cells, especially stem cells, to treat human diseases;
- the understanding of the functioning of the immune system in order to develop strategies for the protection against invading pathogens;
- the development of our knowledge of the molecular biology of plants in order to engineer crop improvements, pathogen resistance and stress tolerance;
- the application of molecular biology techniques to the nature and treatment of bacterial, fungal and viral diseases.

The remaining chapters in this book address the major experimental strategies and analytical techniques that are routinely used to address issues such as these.

1.1.2 Experimental design

Advances in biochemistry and molecular biology, as in all the sciences, are based on the careful design, execution and data analysis of experiments designed to address specific questions or hypotheses. Such experimental design involves a discrete number of compulsory stages:

- the identification of the subject for experimental investigation;
- the critical evaluation of the current state of knowledge (the 'literature') of the chosen subject area noting the strengths and weaknesses of the methodologies previously applied and the new hypotheses which emerged from the studies;
- the formulation of the question or hypothesis to be addressed by the planned experiment;
- the careful selection of the biological system (species, *in vivo* or *in vitro*) to be used for the study;
- the identification of the variable that is to be studied; the consideration of the other variables that will need to be controlled so that the selected variable is the only factor that will determine the experimental outcome;
- the design of the experiment including the statistical analysis of the results, careful evaluation of the materials and apparatus to be used and the consequential potential safety aspects of the study;
- the execution of the experiment including appropriate calibrations and controls, with a carefully written record of the outcomes;
- the replication of the experiment as necessary for the unambiguous analysis of the outcomes;

- the evaluation of the outcomes including the application of appropriate statistical tests to quantitative data where applicable;
- the formulation of the main conclusions that can be drawn from the results;
- the formulation of new hypotheses and of future experiments that emerge from the study.

The results of well-designed and analysed studies are finally published in the scientific literature after being subject to independent peer review, and one of the major challenges facing professional biochemists and molecular biologists is to keep abreast of current advances in the literature. Fortunately, the advent of the web has made access to the literature easier than it once was.

1.2 UNITS OF MEASUREMENT

1.2.1 SI units

The French **Système International d'Unités** (the SI system) is the accepted convention for all units of measurement. Table 1.1 lists basic and derived SI units. Table 1.2 lists numerical values for some physical constants in SI units. Table 1.3 lists the commonly used prefixes associated with quantitative terms. Table 1.4 gives the interconversion of non-SI units of volume.

1.2.2 Molarity – the expression of concentration

In practical terms one mole of a substance is equal to its **molecular mass** expressed in grams, where the molecular mass is the sum of the atomic masses of the constituent atoms. Note that the term molecular mass is preferred to the older term molecular weight. The SI unit of concentration is expressed in terms of moles per cubic metre ($mol\,m^{-3}$) (see Table 1.1). In practice this is far too large for normal laboratory purposes and a unit based on a cubic decimetre (dm^3, $10^{-3}\,m$) is preferred. However, some textbooks and journals, especially those of North American origin, tend to use the older unit of volume, namely the litre and its subunits (see Table 1.4) rather than cubic decimetres. In this book, volumes will be expressed in cubic decimetres or its smaller counterparts (Table 1.4). The **molarity** of a solution of a substance expresses the number of moles of the substance in one cubic decimetre of solution. It is expressed by the symbol M.

It should be noted that atomic and molecular masses are both expressed in **daltons** (Da) or **kilodaltons** (kDa), where one dalton is an atomic mass unit equal to one-twelfth of the mass of one atom of the ^{12}C isotope. However, biochemists prefer to use the term **relative molecular mass** (M_r). This is defined as the molecular mass of a substance relative to one-twelfth of the atomic mass of the ^{12}C isotope. M_r therefore has no units. Thus the relative molecular mass of sodium chloride is 23 (Na) plus

Table 1.1 **SI units – basic and derived units**

Quantity	SI unit	Symbol (basic SI units)	Definition of SI unit	Equivalent in SI units
Basic units				
Length	metre	m		
Mass	kilogram	kg		
Time	second	s		
Electric current	ampere	A		
Temperature	kelvin	K		
Luminous intensity	candela	cd		
Amount of substance	mole	mol		
Derived units				
Force	newton	N	$kg\,m\,s^{-2}$	$J\,m^{-1}$
Energy, work, heat	joule	J	$kg\,m^2\,s^{-2}$	$N\,m$
Power, radiant flux	watt	W	$kg\,m^2\,s^{-3}$	$J\,s^{-1}$
Electric charge, quantity	coulomb	C	$A\,s$	$J\,V^{-1}$
Electric potential difference	volt	V	$kg\,m^2\,s^{-3}A^{-1}$	$J\,C^{-1}$
Electric resistance	ohm	Ω	$kg\,m^2\,s^{-3}A^{-2}$	$V\,A^{-1}$
Pressure	pascal	Pa	$kg\,m^{-1}\,s^{-2}$	$N\,m^{-2}$
Frequency	hertz	Hz	s^{-1}	
Magnetic flux density	tesla	T	$kg\,s^{-2}\,A^{-1}$	$V\,s\,m^{-2}$
Other units based on SI				
Area	square metre	m^2		
Volume	cubic metre	m^3		
Density	kilogram per cubic metre	$kg\,m^{-3}$		
Concentration	mole per cubic metre	$mol\,m^{-3}$		

Table 1.2 **SI units – conversion factors for non-SI units**

Unit	Symbol	SI equivalent
Avogadro constant	L or N_A	$6.022 \times 10^{23}\ \text{mol}^{-1}$
Faraday constant	F	$9.648 \times 10^{4}\ \text{C mol}^{-1}$
Planck constant	h	$6.626 \times 10^{-34}\ \text{J s}$
Universal or molar gas constant	R	$8.314\ \text{J K}^{-1}\ \text{mol}^{-1}$
Molar volume of an ideal gas at s.t.p.		$22.41\ \text{dm}^3\ \text{mol}^{-1}$
Velocity of light in a vacuum	c	$2.997 \times 10^{8}\ \text{m s}^{-1}$
Energy		
calorie	cal	4.184 J
erg	erg	10^{-7} J
electron volt	eV	1.602×10^{-19} J
Pressure		
atmosphere	atm	101 325 Pa
bar	bar	10^{5} Pa
millimetres of Hg	mm Hg	133.322 Pa
Temperature		
centigrade	°C	(t °C + 273.15) K
Fahrenheit	°F	(t °F – 32)5/9 + 273.15 K
Length		
Ångström	Å	10^{-10} m
inch	in	0.0254 m
Mass		
pound	lb	0.4536 kg

Note: s.t.p., standard temperature and pressure.

35.5 (Cl) i.e. 58.5, so that one mole is 58.5 grams. If this was dissolved in water and adjusted to a total volume of 1 dm^3 the solution would be one molar (1 M).

Biological substances are most frequently found at relatively low concentrations and in *in vitro* model systems the volumes of stock solutions regularly used for experimental purposes are also small. The consequence is that experimental solutions are usually in the mM, μM and nM range rather than molar. Table 1.5 shows the interconversion of these units.

Table 1.3 **Common unit prefixes associated with quantitative terms**

Multiple	Prefix	Symbol	Multiple	Prefix	Symbol
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^{9}	giga	G	10^{-12}	pico	p
10^{6}	mega	M	10^{-15}	femto	f
10^{3}	kilo	k	10^{-18}	atto	a
10^{2}	hecto	h	10^{-21}	zepto	z
10^{1}	deca	da	10^{-24}	yocto	y

Table 1.4 **Interconversion of non-SI and SI units of volume**

Non-SI unit	Non-SI subunit	SI subunit	SI unit
1 litre (l)	10^3 ml	$= 1\ \mathrm{dm}^3$	$= 10^{-3}\ \mathrm{m}^3$
1 millilitre (ml)	1 ml	$= 1\ \mathrm{cm}^3$	$= 10^{-6}\ \mathrm{m}^3$
1 microlitre (μl)	10^{-3} ml	$= 1\ \mathrm{mm}^3$	$= 10^{-9}\ \mathrm{m}^3$
1 nanolitre (nl)	10^{-6} ml	$= 1\ \mathrm{nm}^3$	$= 10^{-12}\ \mathrm{m}^3$

Table 1.5 **Interconversion of mol, mmol and μmol in different volumes to give different concentrations**

Molar (M)	Millimolar (mM)	Micromolar (μM)
$1\ \mathrm{mol\ dm}^{-3}$	$1\ \mathrm{mmol\ dm}^{-3}$	$1\ \mu\mathrm{mol\ dm}^{-3}$
$1\ \mathrm{mmol\ cm}^{-3}$	$1\ \mu\mathrm{mol\ cm}^{-3}$	$1\ \mathrm{nmol\ cm}^{-3}$
$1\ \mu\mathrm{mol\ mm}^{-3}$	$1\ \mathrm{nmol\ mm}^{-3}$	$1\ \mathrm{pmol\ mm}^{-3}$

1.3 WEAK ELECTROLYTES

1.3.1 The biochemical importance of weak electrolytes

Many molecules of biochemical importance are weak electrolytes in that they are acids or bases that are only partially ionised in aqueous solution. Examples include

the amino acids, peptides, proteins, nucleosides, nucleotides and nucleic acids. It also includes the reagents used in the preparation of buffers such as ethanoic (acetic) acid and phosphoric acid. The biochemical function of many of these molecules is dependent upon their precise state of ionisation at the prevailing cellular or extracellular pH. The catalytic sites of enzymes, for example, contain functional carboxyl and amino groups, from the side chains of constituent amino acids in the protein chain, which need to be in a specific ionised state to enable the catalytic function of the enzyme to be realised. Before the ionisation of these compounds is discussed in detail, it is necessary to appreciate the importance of the ionisation of water.

1.3.2 Ionisation of weak acids and bases

One of the most important weak electrolytes is water since it ionises to a small extent to give hydrogen ions and hydroxyl ions. In fact there is no such species as a free hydrogen ion in aqueous solution as it reacts with water to give a **hydronium ion** (H_3O^+):

$$H_2O \rightleftharpoons H^+ + HO^-$$
$$H^+ + H_2O \rightleftharpoons H_3O^+$$

Even though free hydrogen ions do not exist it is conventional to refer to them rather than hydronium ions. The equilibrium constant (K_{eq}) for the ionisation of water has a value of 1.8×10^{16} at 24 °C:

$$K_{eq} = \frac{[H^+][OH^-]}{[H_2O]} = 1.8 \times 10^{16} \quad (1.1)$$

The molarity of pure water is 55.6 M. This can be incorporated into a new constant, K_w:

$$1.8 \times 10^{-16} \times 55.6 = [H^+][HO^-] = 1.0 \times 10^{-14} = K_w \quad (1.2)$$

K_w is known as the **autoprotolysis constant** of water and does not include an expression for the concentration of water. Its numerical value of exactly 10^{-14} relates specifically to 24 °C. At 0 °C K_w has a value of 1.14×10^{-15} and at 100 °C a value of 5.45×10^{-13}. The stoichiometry in equation 1.2 shows that hydrogen ions and hydroxyl ions are produced in a 1 : 1 ratio, hence both of them must be present at a concentration of 1.0×10^{-7} M. Since the Sörensen definition of pH is that it is equal to the negative logarithm of the hydrogen ion concentration, it follows that the pH of pure water is 7.0. This is the definition of neutrality.

Ionisation of carboxylic acids and amines

As previously stressed, many biochemically important compounds contain a carboxyl group (-COOH) or a primary (RNH_2), secondary (R_2NH) or tertiary (R_3N) amine which can donate or accept a hydrogen ion on ionisation. The tendency of a weak acid, generically represented as HA, to ionise is expressed by the equilibrium reaction:

$$\underset{\text{weak acid}}{HA} \rightleftharpoons H^+ + \underset{\text{conjugate base (anion)}}{A^-}$$

This reversible reaction can be represented by an equilibrium constant, K_a, known as the **acid dissociation constant** (equation 1.3). Numerically, it is very small.

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{1.3}$$

Note that the ionisation of a weak acid results in the release of a hydrogen ion and the conjugate base of the acid, both of which are ionic in nature.

Similarly, amino groups (primary, secondary and tertiary) as weak bases can exist in ionised and unionised forms and the concomitant ionisation process is represented by an equilibrium constant, K_b (equation 1.4):

$$\underset{\text{weak base}\ (\text{primary amine})}{RNH_2 + H_2O} \rightleftharpoons \underset{\text{conjugate acid}\ (\text{substituted ammonium ion})}{RNH_3^+ + HO^-}$$

$$K_b = \frac{[RNH_3^+][HO^-]}{[RNH_2][H_2O]} \tag{1.4}$$

In this case, the non-ionised form of the base abstracts a hydrogen ion from water to produce the conjugate acid that is ionised. If this equation is viewed from the reverse direction it is of a similar format to that of equation 1.3. Equally, equation 1.3 viewed in reverse is similar in format to equation 1.4.

A specific and simple example of the ionisation of a weak acid is that of acetic (ethanoic) acid, CH_3COOH:

$$\underset{\text{acetic acid}}{CH_3COOH} \rightleftharpoons \underset{\text{acetate anion}}{CH_3COO^-} + H^+$$

Acetic acid and its conjugate base, the acetate anion, are known as a **conjugate acid–base pair**. The acid dissociation constant can be written in the following way:

$$K_a = \frac{[CH_3COO^-][H^+]}{[CH_3COOH]} = \frac{[\text{conjugate base}][H^+]}{[\text{weak acid}]} \tag{1.5a}$$

K_a has a value of 1.75×10^{-5} M. In practice it is far more common to express the K_a value in terms of its negative logarithm (i.e. $-\log K_a$) referred to as pK_a. Thus in this case pK_a is equal to 4.75. It can be seen from equation 1.3 that pK_a is numerically equal to the pH at which 50% of the acid is protonated (unionised) and 50% is deprotonated (ionised).

It is possible to write an expression for the K_b of the acetate anion as a conjugate base:

$$CH_3COO_3^- + H_2O \rightleftharpoons CH_3COOH + HO^-$$

$$K_b = \frac{[CH_3COOH][HO^-]}{[CH_3COO^-]} = \frac{[\text{weak acid}][OH^-]}{[\text{conjugate base}]} \tag{1.5b}$$

K_b has a value of 1.77×10^{-10} M, hence its pK_b (i.e. $-\log K_b$) = 9.25.

Multiplying these two expressions together results in the important relationship:

$$K_a \times K_b = [H^+][OH^-] = K_w = 1.0 \times 10^{-14} \text{ at } 24\,^\circ C$$

Table 1.6 **pK_a values of some acids and bases that are commonly used as buffer solutions**

Acid or base	pK_a
Acetic acid	4.75
Barbituric acid	3.98
Carbonic acid	6.10, 10.22
Citric acid	3.10, 4.76, 5.40
Glycylglycine	3.06, 8.13
Hepes[a]	7.50
Phosphoric acid	1.96, 6.70, 12.30
Phthalic acid	2.90, 5.51
Pipes[a]	6.80
Succinic acid	4.18, 5.56
Tartaric acid	2.96, 4.16
Tris[a]	8.14

Note: [a]See list of abbreviations at the front of the book.

hence

$$pK_a + pK_b = pK_w = 14 \tag{1.6}$$

This relationship holds for all acid–base pairs and enables one pK_a value to be calculated from knowledge of the other. Biologically important examples of conjugate acid–base pairs are lactic acid/lactate, pyruvic acid/pyruvate, carbonic acid/bicarbonate and ammonium/ammonia.

In the case of the ionisation of weak bases the most common convention is to quote the K_a or the pK_a of the conjugate acid rather than the K_b or pK_b of the weak base itself. Examples of the pK_a values of some weak acids and bases are given in Table 1.6. Remember that the smaller the numerical value of pK_a the stronger the acid (more ionised) and the weaker its conjugate base. Weak acids will be predominantly unionised at low pH values and ionised at high values. In contrast, weak bases will be predominantly ionised at low pH values and unionised at high values. This sensitivity to pH of the state of ionisation of weak electrolytes is important both physiologically and in *in vitro* biochemical studies employing such analytical techniques as electrophoresis and ion-exchange chromatography.

Ionisation of polyprotic weak acids and bases

Polyprotic weak acids and **bases** are capable of donating or accepting more than one hydrogen ion. Each ionisation stage can be represented by a K_a value using the convention that K_a^1 refers to the acid with the most ionisable hydrogen atoms and K_a^n the acid with the least number of ionisable hydrogen atoms. One of the most important

biochemical examples is phosphoric acid, H_3PO_4, as it is widely used as the basis of a buffer in the pH region of 6.70 (see below):

$$H_3PO_4 \rightleftharpoons H^+ + H_2PO_4^- \quad pK_a^1\ 1.96$$
$$H_2PO_4^- \rightleftharpoons H^+ + HPO_4^{2-} \quad pK_a^2\ 6.70$$
$$HPO_4^{2-} \rightleftharpoons H^+ + PO_4^{3-} \quad pK_a^3\ 12.30$$

Example 1 CALCULATION OF pH AND THE EXTENT OF IONISATION OF A WEAK ELECTROLYTE

Question Calculate the pH of a 0.01 M solution of acetic acid and its fractional ionisation given that its K_a is 1.75×10^{-5}.

Answer To calculate the pH we can write:

$$K_a = \frac{[\text{acetate}^-][H^+]}{[\text{acetic acid}]} = 1.75 \times 10^{-5}$$

Since acetate and hydrogen ions are produced in equal quantities, if $x =$ the concentration of each then the concentration of unionised acetic acid remaining will be $0.01 - x$. Hence:

$$1.75 \times 10^{-5} = \frac{(x)(x)}{0.01 - x}$$
$$1.75 \times 10^{-7} - 1.75 \times 10^{-5}x = x^2$$

This can now be solved either by use of the quadratic formula or, more easily, by neglecting the x term since it is so small. Adopting the latter alternative gives:

$$x^2 = 1.75 \times 10^{-7}$$

hence

$$x = 4.18 \times 10^{-4}\ \text{M}$$

hence

$$\text{pH} = 3.38$$

The fractional ionisation (α) of the acetic acid is defined as the fraction of the acetic acid that is in the form of acetate and is therefore given by the equation:

$$\alpha = \frac{[\text{acetate}]}{[\text{acetate}] + [\text{acetic acid}]}$$
$$= \frac{4.18 \times 10^{-4}}{4.18 \times 10^{-4} + 0.01 - 4.18 \times 10^{-4}}$$
$$= \frac{4.18 \times 10^{-4}}{0.01}$$
$$= 4.18 \times 10^{-2} \text{ or } 4.18\%$$

Thus the majority of the acetic acid is present as the unionised form. If the pH is increased above 3.38 the proportion of acetate present will increase in accordance with the Henderson–Hasselbalch equation.

Basic Principles of Wastewater Treatment. Marcos von Sperling. Department of Sanitary and Environmental Engineering Federal University of Minas Gerais, Brazil. Published by IWA Publishing, Alliance House, 12 Caxton Street, London SW1H 0QS, UK. Telephone: +44 (0) 20 7654 5500; Fax: +44 (0) 20 7654 5555; Email: publications@iwap.co.uk Website: www.iwapublishing.com. First published 2007 C 2007 IWA Publishing. Copy-edited and typeset by Aptara Inc., New Delhi, India Printed by Lightning Source.