1

Units

THE metric system has been a part of pharmaceutical practice for many years; indeed, as long ago as 1867, the first edition of the *British Pharmacopoeia* used metric quantities for volumetric analysis. This was extended in the 1898 edition to gravimetric analysis, and quantities for preparations were stated in parallel, using the metric and the Imperial systems. In the 1914 *Pharmacopoeia*, the quantities for preparations were expressed in metric only, although doses still used Imperial weights and measures, with an approximate metric equivalent. This continued until 1963, when doses were stated in the metric system only, so that all official standards became entirely metric. The metrication of pharmacy was completed on 1st March 1969, when all dispensing under the National Health Service was transferred to this system.

The use of metric weights and measures is, however, only part of a much greater change in which a rationalised system of metric units is coming into international use. Known as the *Système International d'Unités*, with *SI* as the accepted abbreviation, the objective is to derive almost all the quantities needed from a few base units, and to provide rules for the use of SI units to reduce errors and avoid ambiguities.

This book uses SI units, with recommended multiples and sub-multiples, and follows the recommendations of the International Organisation for Standardisation (ISO) and the British Standards Institution (BSI) with a few exceptions that are finding general acceptance where other permitted units are more convenient in practice than the recommended SI units.

In addition, it will be found that non-recommended units are used in certain cases where quotation is made from sources such as the *British Pharmacopoeia* which, at the time of writing, uses metric units, but does not yet follow SI recommendations.

The system of SI units has two important features:

1. All relationships are decimal.
2. It is a coherent system of units and in this respect it differs from metric systems previously in use. A coherent system is one in which the product or quotient of any two unit quantities is the unit of the resultant quantity. Thus, unit velocity results when unit length is divided by unit time, unit acceleration when unit velocity is divided by unit time, and unit force when unit mass is multiplied by unit acceleration. As well as providing straightforward relationships, this eliminates a number of special units and avoids former problems, such as the confusion between mass, force, and weight.

SI BASE UNITS

Six base units are used in the SI system (Table 1.1).

Table 1.1

Quantity	*Unit*	*Symbol*
length	metre	m
mass	kilogramme	kg
time	second	s
electric current	ampere	A
thermodynamic temperature	kelvin	K
luminous intensity	candela	cd

SUPPLEMENTARY SI UNITS

Certain additional supplementary units are used, including the radian (rad) for plane angles.

DERIVED SI UNITS

Further SI units are derived from the base units and are stated in terms of these units. Some of the derived units, which are used in this book and which have special names, are listed in Table 1.2. The remainder are included in the list of all symbols and units given in Table 1.4.

MULTIPLES AND SUB-MULTIPLES

As far as possible, numerical values should be kept between 0·1 and 1000, so that multiples and

Table 1.2
Derived SI Units with Special Names

Quantity	*Name of SI unit*	*Symbol*	*Expressed in terms of SI base units or derived units*
frequency	hertz	Hz	1 Hz = 1/s
force	newton	N	1 N = 1 kg m/s^2
work, energy, quantity of heat	joule	J	1 J = 1 N m
power	watt	W	1 W = 1 J/s

submultiples are recommended for use, based on the decimal range given in Table 1.3.

Thus, kilogramme represents 10^3 g and microgramme represents 10^{-6} g.

Table 1.3
Decimal Multiples and Sub-multiples

Factor by which unit is multiplied	*Prefix*	*Symbol*
10^{12}	tera	T
10^9	giga	G
* 10^6	mega	M
* 10^3	kilo	k
10^2	hecto	h
10	deca	da
10^{-1}	deci	d
10^{-2}	centi	c
* 10^{-3}	milli	m
* 10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a

To minimise the number of multiples and submultiples and to reduce the risk of error, only prefixes that represent 10 to a power that is a multiple of 3 are recommended. Those in use in this book are marked in Table 1.3 with an asterisk (*).

To avoid confusion, all calculations are performed in SI units themselves and not in decimal multiples or sub-multiples. In addition, only one prefix is applied to a unit at one time; for example, the term micrometre is used for one millionth of a metre, and not the term milli-millimetre.

EXCEPTIONS TO THE USE OF PREFERRED SI UNITS

Volume

The SI unit for volume is the cubic metre, with the cubic millimetre as the recommended sub-multiple. In practice, it is useful to have a unit of intermediate size and the litre is in common use. This name has been adopted as a synonym for the cubic decimetre and is used for general statement of volumes, with the millilitre as sub-multiple.

Pressure

The SI unit of force is the *newton*, defined as that force which when applied to a mass of 1 kilogramme gives it an acceleration of 1 metre per second per second. The newton per square metre is the SI unit for pressure, but for many purposes the unit is too small. For convenience, a unit known as the *bar* has been suggested, where 1 bar = 10^5 N/m^2. The advantage of the bar is that, for all practical purposes, 1 bar = 1 atmosphere pressure, so that pressures from 0 to 1 bar refer to pressures below atmospheric, while pressures from 1 bar upwards are above atmospheric. For high vacuum work, the unit N/m^2 is retained.

Temperature

The SI unit for temperature is the base unit of thermodynamic temperature, namely the kelvin (K), which is used in all calculations, and generally for physicochemical purposes.

As a 'customary' or practical unit, the degree Celsius (°C) is retained. This is identical to the former degree Centigrade, the change in name being due to the fact that the 'grade', and hence the 'centigrade', has another meaning in some countries.

The degree Celsius and the kelvin are identical units of temperature interval, but since 0°C = 273·15 K, then T°C = T + 273·15 K. It follows, therefore, that temperature intervals or temperature differences are the same in both cases, and these are stated in calculations as K. Hence, if water at 20°C is to be heated by a steam jacket at 100°C, the temperature difference is 80 K. To illustrate these various points:

The unit K is used in the Arrhenius equation, Eqn (6.35), which indicates the effect of temperature on the rate constant of a chemical reaction.

The coefficient of thermal conductivity is used in calculations in Chapter 12 and refers to a temperature interval, so that the units are W/mK.

The reference to the saturation temperature of steam at a pressure of two bars, made in Chapter 12,

Table 1.4

List of SI Units and Decimal Multiples and Sub-multiples

Quantity	*SI unit*	*Multiple or sub-multiple units*	*Other permitted units*
Part I: Space and time			
plane angle	rad (radian)		
length	m (metre)	mm μm	
area	m^2	mm^2	
volume	m^3	mm^3	litre and ml (1 litre = 1 dm^3)
time	s (second)	ks ms	minute, hour, day
velocity	m/s		
Part II: Periodic and related phenomena			
frequency	Hz (hertz)		
rotational frequency	1/s		
Part III: Mechanics			
mass	kg (kilogramme)	Mg g mg μg	
density (mass density)	kg/m^3		
force	N (newton)		
pressure	N/m^2	kN/m^2 MN/m^2	bar (1 bar = 10^5 N/m^2)
viscosity (dynamic)	N s/m^2		
kinematic viscosity	m^2/s		
surface tension	N/m		
energy, work	J (joule)	kJ	
power	W (watt)		
Part IV: Heat			
thermodynamic temperature	K (kelvin)		
Celsius temperature			[°C] (degree Celsius)
temperature interval	K		
heat, quantity of heat	J	kJ	
heat flow rate	W	kW	
density of heat flow rate	W/m^2		
thermal conductivity	W/mK		
coefficient of heat transfer	W/m^2K		
heat capacity	J/K	kJ/K	
specific energy	J/kg	kJ/kg	
specific latent heat	J/kg	kJ/kg	

is a customary temperature, so that it is given as 120·2°C.

Similarly, Chapter 24 gives data on equilibrium moisture contents for various materials at 20°C.

Amount of Substance (*Symbol: n*)

An additional unit, the mole, corresponding to the quantity 'amount of substance', has been recommended but it is not as yet approved by the international body concerned with the metric system.

The mole is the amount of substance that contains as many elementary units as there are atoms in 0·012 kg of carbon-12. The elementary unit must be specified and may be an atom, an ion, a molecule, etc. or a specified group of such species. For example—

1 mole of K_2SO_4 has a mass equal to 0·174 16 kg
1 mole of K^+ has a mass equal to 0·039 10 kg
1 mole of K_2^{++} has a mass equal to 0·078 20 kg
1 mole of SO_4^{--} has a mass equal to 0·095 96 kg
1 mole of $\frac{1}{2}SO_4^-$ has a mass equal to 0·047 88 kg

Units such as the gramme-molecule, gramme-equivalent, equivalent or gramme-ion will, therefore, become obsolete.

PRESENTATION OF NUMERICAL VALUES

In many countries, the comma is used as the decimal point, whereas it has been practice in the UK to use the comma to separate the digits of large numbers into groups of three.

To avoid confusion, this book follows present British practice whereby the decimal point is a full point placed above the line (·) and digits are separated by a small space between groups of three when five or more digits are used, except when data are expressed in tabular form; thus 1000 and 10 000.

BIBLIOGRAPHY

ANDERTON, P. and BIGG, P. M. (1967) *Changing to the Metric System*. London, H.M. Stationery Office.

British Standards Institution (1969) PD 5686, 1969 *The Use of SI Units*. London, The British Standards Institution.

MCGLASHAN, M. L. (1968) *Physico-chemical Quantities and Units*. London, Royal Institute of Chemistry.

2

Solutions

AN understanding of the properties of solutions, and the factors that affect solubility, is essential to pharmacists because of the importance of solutions in so many areas of pharmaceutical interest.

A true solution may be defined as a mixture of two or more components that form a homogeneous molecular dispersion; i.e. a system in which one component is dispersed as small molecules or ions throughout the other component. This definition differentiates true solutions from homogeneous colloidal dispersions, in which the particles of the dispersed component are larger than those in true solutions. (*See* Chapter 5 for a classification of disperse systems.)

In a solution composed of only two components (a binary system), that component which is dispersed throughout the other is termed the solute, and the component in which the dispersion occurs is termed the solvent. In general, the solvent is present in the greater amount, but several exceptions occur. For example, syrup contains about 65 per cent of sucrose in aqueous solution. In addition, it is often difficult to classify the components of mixtures of miscible liquids, such as alcohol and water, as solutes or solvents because either component may be regarded as the solvent depending on the composition of the mixture.

The Process of Dissolution

The transference of a molecule of a solute into solution in a solvent involves a change in the environments of both solute and solvent. The solute is separated from other similar molecules and becomes surrounded by solvent molecules. In addition, the solvent molecules are separated sufficiently from other similar molecules to create space for the accommodation of the solute molecule. Thus, dissolution will occur only if the solute and the solvent are mutually attracted to a degree that is sufficient to overcome the solute-solute and solvent-solvent intermolecular attractive forces.

Similar types of intermolecular force may contribute to solute-solvent, solute-solute, and solvent-solvent interactions. However, the strengths of these various contributing forces differ considerably. The Appendix shows that the attractive forces exerted between polar molecules are much stronger than those that exist between polar and non-polar molecules or between non-polar molecules themselves.

Thus, in a polar solute, where the intermolecular interaction is appreciable, transference of solute molecules into solution will occur only if the solute-solvent association is even stronger. Such strong solute-solvent interaction will result only if the solvent is also a polar substance (e.g. water), since a non-polar solvent (e.g. benzene) will be unable to exert sufficient attraction on a molecule to cause it to separate from other solute molecules.

Conversely, the dissolution of a non-polar substance such as paraffin wax, the intermolecular attractions of which are relatively weak, will occur only if the solute-solvent interaction is stronger than the solvent-solvent interaction. A marked intermolecular association between solvent molecules, such as that which exists between the molecules of a polar solvent (e.g. water) will therefore tend to prevent dissolution of a non-polar solute, so that solvents for this type of solute tend to be restricted to non-polar liquids (e.g. benzene).

The above considerations are often expressed in the very general manner that 'like dissolves like'; i.e. a polar substance will dissolve in a polar solvent and a non-polar substance will dissolve in a non-polar solvent. However, such a generalisation should be treated with caution, since the intermolecular attractions involved in the process of dissolution are influenced by factors that are not obvious from a consideration of the overall polarity of a molecule. For example, the possibility of hydrogen-bond formation (*see* Appendix) between solute and solvent may be more significant than polarity.

Methods of Expressing the Concentration of Solutions

It is assumed that the student is familiar with concentration terms such as normality and molarity.

Although such expressions are included in many existing reference books for pharmacists, the introduction of the SI system of units has either made them obsolete or will eventually lead to their abandonment.

The methods recommended by McGlashan (1968) for expressing the amount of substance in a given solution include—

(*a*) the molality of solute, which is defined as the number of moles of solute divided by the mass of the solvent, and its SI units are mole kg^{-1}. (The term molality was used in the cgs scale and, although the name is likely to be retained as the name of an SI quantity, the use of the abbreviation 'molal' is not recognised as a unit symbol by the General Conference on Weights and Measures, and the abbreviation 'm' for molality is to be discouraged because this letter is used as a symbol for metre.)

(*b*) the concentration of solute, which is defined as the number of moles of solute divided by the volume of the solution, and its SI units are mole m^{-3}.

In addition, it is also possible to express concentration of a solution in terms of the mass or volume of a solute contained in a given mass or volume of solution. It should be remembered that the basic SI units for mass and volume are the kilogramme and cubic metre, respectively. These units may be unwieldy in certain instances. and appropriate prefixes to indicate decimal fractions or multiples should therefore be used. The previous use of the litre as a common unit of volume will probably lead to the use of the cubic decimetre (dm^3) as a common unit of volume. It is unfortunate that the term kilogramme and its symbol kg suggest a multiple of a basic unit, and make difficult construction of decimal fractions of this unit by the addition of prefixes. A new name and symbol will probably be introduced in the future.

PERCENTAGE EXPRESSIONS

The concentration of a solution may be expressed as a percentage; i.e.

$$\text{concentration} = \frac{\text{mass or volume of solute}}{\text{mass or volume of solution}} \times 100$$

Percentage weight in volume (% w/v), percentage weight in weight (% w/w), percentage volume in weight (% v/w) and percentage volume in volume (% v/v) are often used in pharmaceutical practice. For example, the strengths of Pharmacopoeial preparations may be defined as percentages; e.g. Belladonna Tincture (BP 1968) contains 0·03% w/v of the alkaloids of Belladonna Herb.

Solubility

The solubility of a substance in a solvent at a given temperature and pressure, is the amount of substance that has passed into solution when equilibrium is attained between the solution and excess, i.e. undissolved, substance. The solution that is obtained under these conditions is termed a saturated solution. It is possible to obtain solutions that are supersaturated. However, they are unstable, and scratching the side of the container, the presence of dust, or the addition of undissolved solute will provide nuclei that readily lead to precipitation of the excess solute.

METHODS OF EXPRESSING SOLUBILITY

Solubilities may be expressed quantitatively by the same methods as are used for stating concentration. The *British Pharmacopoeia* (1968) expresses solubilities as the number of parts by volume of solvent required to dissolve one part by weight of a solid or one part by volume of a liquid. Unless otherwise specified, these solubilities apply at room temperature. Figures given under the side heading 'Solubility in the Pharmacopoeial monographs, are only approximate and are not intended to be official requirements. However, statements under side headings such as 'Solubility in Alcohol' are exact and are intended as part of the official requirements for that substance.

TYPES OF SOLUTION

Solutions may be classified on the basis of the physical states of the components. Since there are three states of matter, i.e. solid, liquid, and gas, nine different types of solution with two components are possible, as shown in Table 2.1.

Solutions of solids in liquids are the most important in pharmacy, and only this type of system is

Table 2.1
Types of Solution

Solute	*Solvent*
Gas	Gas
Liquid	Gas
Solid	Gas
Gas	Liquid
Liquid	Liquid
Solid	Liquid
Gas	Solid
Liquid	Solid
Solid	Solid

discussed in the present chapter. Further information on most of the systems shown in Table 2.1 that are involved in pharmaceutical processes or products is given in the following chapter.

DETERMINATION OF THE SOLUBILITY OF SOLIDS IN LIQUIDS

The following points should be observed in all solubility determinations.

(*a*) The solvent and solute must be pure.
(*b*) A saturated solution must be obtained before any solution is removed for analysis.
(*c*) The method of separating a sample of saturated solution from undissolved solute must be satisfactory.
(*d*) The method of analysing the solution must be reliable.
(*e*) Temperature must be adequately controlled.

A saturated solution is obtained either by stirring excess powdered solute with solvent for several hours at the required temperature until equilibrium has been attained, or by warming the solvent with an excess of the solute and allowing the mixture to cool to the required temperature. It is essential that some undissolved solid should be present at the completion of this stage in order to ensure that the solution is saturated.

A sample of the saturated solution is obtained for analysis by separating the solution from the undissolved solid. Filtration is usually used, but precautions should be taken to ensure that: (*a*) it is carried out at the temperature of the solubility determination, in order to prevent any change in the equilibrium between dissolved and undissolved solute; and (*b*) loss of a volatile component does not occur.

Different methods of analysis may be applied to the saturated solution depending on the type of system involved. For example, if one of the components is volatile and one is non-volatile, the amount of the latter can be determined by heating to constant weight. Alternatively, the solute may be converted to an insoluble compound by chemical reaction, and the weight of this may be obtained after filtration and drying. Volumetric analysis may be used, especially for those compounds that exhibit the reactions of acids, alkalis, chlorides, etc. which are readily determined by this means. Physical measurements offer a further means of analysis. For example, electrical conductivity measurements are suitable for sparingly soluble electrolytes, optical rotation may be used for optically active compounds, or a radioactive indicator method may be employed. This latter method involves the preparation of the test material in such a way that it contains a known proportion of a radioactive indicator. A saturated solution is made and its level of radioactivity may be used to determine the concentration of solute.

FACTORS AFFECTING THE SOLUBILITY OF SOLIDS IN LIQUIDS

1. *Temperature*

In most cases the dissolution of a solid in a liquid involves the absorption of heat; i.e. it is an endothermic process with a positive heat of solution. If this type of system is heated, it will tend to react in a way that will nullify the constraint imposed upon it; i.e. the rise in temperature. This tendency is an example of Le Chatelier's principle. Thus, a rise in temperature will lead to an increase in the solubility of a solid with a positive heat of solution. Conversely, if the dissolution of a solid involves the liberation of heat; i.e. it is an exothermic process with a negative heat of solution, then an increase in temperature will lead to a decrease in the solubility.

Solubility curves are often used to indicate the effect of temperature on the solubility of a given substance. Some of these are shown in Fig. 2.1. It can be seen that potassium nitrate shows a marked increase in solubility with rise in temperature, while calcium acetate shows a small decrease. These

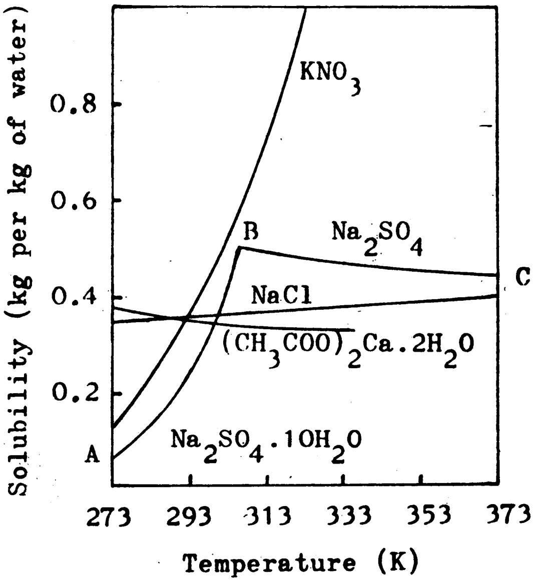


Fig. 2.1 Solubility curves for various substances in water compounds are therefore examples of substances with positive and negative heats of solution, respectively. The dissolution of sodium chloride involves little absorption of heat, as indicated by the approximately horizontal solubility curve for this compound.

The majority of solubility curves are continuous curves, but abrupt changes in slope may sometimes be observed if the nature of the solid phase in contact with the solution alters. For example, sodium sulphate exists as the decahydrate $Na_2SO_4.10\ H_2O$ up to a temperature of 305·55 K, and its dissolution in water is an endothermic process. Above this temperature the solid is converted into the anhydrous form, Na_2SO_4, and the dissolution of this compound is an exothermic process. The solubility curve therefore exhibits a break at 305·55 K, which is known as a transition point, and *AB* and *BC* in Fig. 2.1 represent the solubility curves for the decahydrate and the anhydrous form of sodium sulphate, respectively.

2. *Particle Size of the Solid*

The changes in interfacial free energy (*see* p. 33) that accompany the dissolution of particles of varying sizes cause the solubility of a substance to increase with decreasing particle size, as indicated by Eqn (2.1),

$$\log \frac{s}{s_0} = \frac{2\gamma M}{2{\cdot}303 RT\rho r} \tag{2.1}$$

where s is the solubility of small particles of radius r,
s_0 is the normal solubility (i.e. of a solid consisting of fairly large particles),
γ is the interfacial energy (*see* Chapter 4),
M is the molecular weight of the solid,
ρ is the density of the bulk solid,
R is the gas constant, and
T is the thermodynamic temperature.

This effect may be significant in the storage of pharmaceutical suspensions, since the smaller particles in such a suspension will be more soluble than the larger ones. As the small particles disappear, the overall solubility of the suspended drug will decrease, and growth of the larger particles will occur. The occurrence of crystal growth by this mechanism is of particular importance in the storage of suspensions intended for injection (Gunn and Carter, 1965).

The increase in solubility with decrease in particle size ceases when the particles have a very small radius, and any further decrease in size causes a decrease in solubility. It has been postulated that this change arises from the presence of an electrical charge on the particles and that the effect of this charge becomes more important as the size of the particles decreases (Buckley, 1951).

3. *Solvent*

It will be appreciated from the discussion of the mechanism of solution that the solubility of a solid depends on the nature of the solvent. The solubilities of a particular solid in a variety of liquids will, therefore, differ. In addition, changes in the properties of a solvent caused by the addition of other substances, may affect the solubility of a solid in the system (p. 12).

Water is the most common solvent encountered in pharmaceutical practice, especially for preparations intended for internal use. Ethanol, usually in various mixtures with water, is also popular. Simple organic liquids such as ether, chloroform, acetone, and various glycols and oils, may be used in addition to water and alcohol in preparations intended for external use. (*See* Gunn and Carter, 1965, for solvents used in parenteral products.)

4. *pH*

Many drugs behave as weak acids or bases and their solubility is therefore affected by the pH of an aqueous solvent. For example, a weakly acidic drug such as acetylsalicylic acid (aspirin) will be more soluble in alkaline solution, since it will be converted to the more soluble salt. Conversely, the drug will be precipitated from aqueous solution if the pH is lowered by the addition of a strong acid. Similarly, a weakly basic drug will be more soluble in solutions of low pH and will precipitate if the pH is raised by the addition of an alkali.

5. *Additional Substances*

(*a*) *Common Ion Effect.* The solubility of a sparingly soluble electrolyte is decreased by the addition of a second electrolyte that possesses a similar ion to the first. This is known as the common ion effect.

In a saturated solution in contact with undissolved solid, the equilibrium may be represented as follows for a compound *AB*:

$$\underset{\text{undissolved solid}}{AB_{(s)}} \rightleftharpoons \underbrace{\underset{\text{undissociated molecule}}{AB} \rightleftharpoons \underset{\text{ions}}{A^+ + B^-}}_{\text{species in solution}}$$

If the salt is sparingly soluble, then the concentration of solute is sufficiently small to assume complete dissociation into ions. The overall equilibrium may then be represented by

$$AB_{(s)} \rightleftharpoons A^+ + B^- \tag{2.2}$$

From the Law of Mass Action, the equilibrium constant (K) for this reversible reaction is given by

$$K = \frac{[A^+][B^-]}{[AB]_{(s)}}$$

where the square brackets indicate concentration of the respective components. Furthermore, the concentration of a solid may be regarded as being constant

$$\therefore \quad K = \frac{[A^+][B^-]}{\text{constant}}$$

$$\therefore \quad K_s' = [A^+][B^-] \qquad (2.3)$$

where K_s' is a constant and is known as the solubility product of compound AB.

If each molecule of the salt contains more than one ion of each type, e.g. A_xB_y, then in the definition of the solubility product the concentration of each ion is expressed to the appropriate power; i.e.

$$K_s' = [A^+]^x[B^-]^y$$

These equations for the solubility product are only applicable to solutions of sparingly soluble salts.

If K_s' is exceeded by the product of the concentration of the ions, i.e. $[A^+][B^-]$, then the equilibrium shown above, Eqn (2.2), moves towards the left in order to restore the equilibrium, and solid AB is precipitated. The product $[A^+][B^-]$ will be increased by the addition of more A^+ ions produced by the dissociation of another compound, e.g. $AX \rightarrow A^+ + X^-$, where A^+ is the common ion. Solid AB will be precipitated and the solubility of this compound is therefore decreased. This is known as the common ion effect. The addition of common B^- ions would have the same effect.

The precipitating effect of common ions is, in fact, less than that predicted from Eqn (2.3). The reason for this is explained in the following section.

(*b*) *Effect of Indifferent Electrolytes.* The solubility of a sparingly soluble electrolyte may be increased by the addition of a second electrolyte that does not possess ions common to the first; i.e. an indifferent electrolyte.

The definition of the solubility product of a sparingly soluble electrolyte in terms of the concentration of ions produced at equilibrium, as indicated by Eqn (2.3), is only an approximation from the more exact thermodynamic relationship expressed by Eqn (2.4),

$$K_s = a_{A^+} \,.\, a_{B^-} \qquad (2.4)$$

where K_s is the solubility product of compound AB and a_{A^+} and a_{B^-} are known as the activities of the respective ions. The activity of a particular ion may be regarded as its 'effective concentration'. In general, this has a lower value than the actual concentration, because some ions produced by dissociation of the electrolyte are strongly associated with oppositely charged ions and do not contribute so effectively as completely unassociated ions to the properties of the system. At infinite dilution, the wide separation of ions prevents any interionic association, and the molar concentration (c_{A^+}) and activity (a_{A^+}) of a given ion (A^+) are then equal; i.e.

$$a_{A^+} = c_{A^+}, \quad \text{or} \quad \frac{a_{A^+}}{c_{A^+}} = 1$$

As the concentration increases, the effects of interionic association are no longer negligible, and the ratio of activity to molar concentration becomes less than unity; i.e.

$$\frac{a_{A^+}}{c_{A^+}} = f_{A^+}$$

or

$$a_{A^+} = c_{A^+} \,.\, f_{A^+}$$

where f_{A^+} is known as the activity coefficient of A^+. If concentrations and activity coefficients are used instead of activities in Eqn (2.4) then

$$K_s = (c_{A^+} \,.\, c_{B^-})(f_{A^+} \,.\, f_{B^-})$$

The product of the concentrations, i.e. $(c_{A^+} \,.\, c_{B^-})$, will be a constant (K_s') as shown by Eqn (2.3), and $(f_{A^+} \,.\, f_{B^-})$ may be equated to $f^2_{A^+B^-}$, where $f_{A^+B^-}$ is the mean activity coefficient of the salt AB, i.e.

$$K_s = K_s' f^2_{A^+B^-} \qquad (2.5)$$

Since $f_{A^+B^-}$ varies with the overall concentration of ions present in solution (the ionic strength), and since K_s is a constant, it follows that K_s' must also vary with the ionic strength of the solution in an inverse manner to the variation of $f_{A^+B^-}$. Thus, in a system containing a sparingly soluble electrolyte without a common ion, the ionic strength will have an appreciable value and the mean activity coefficient $f_{A^+B^-}$ will be less than one.

From Eqn (2.5) it will be seen that K_s' will, therefore, be greater than K_s. In fact, the concentration solubility product K_s' will become larger and larger as the ionic strength of the solution increases. The solubility of AB will therefore increase as the concentration of added electrolyte increases.

This argument also accounts for the fact that if no allowance is made for the variation in activity with ionic strength of the medium, the precipitating effect of common ions is less than that predicted from the Law of Mass Action.

(*c*) *Effect of Non-electrolytes on the Solubility of Electrolytes.* The solubility of electrolytes depends on the dissociation of dissolved molecules into ions. The ease of this dissociation is affected by the dielectric constant of the solvent, which is a measure of the polar nature of the solvent. Liquids with a high dielectric constant (e.g. water and formic acid) are able to reduce the attractive forces that operate between oppositely charged ions produced by dissociation of an electrolyte.

If a water-soluble non-electrolyte such as alcohol is added to an aqueous solution of a sparingly soluble electrolyte, the solubility of the latter is decreased because the alcohol lowers the dielectric constant of the solvent and ionic dissociation of the electrolyte becomes more difficult

(*d*) *Effect of Electrolytes on the Solubility of Non-electrolytes.* Non-electrolytes do not dissociate into ions in aqueous solution, and in dilute solution the dissolved species therefore consists of single molecules. Their solubility in water depends on the formation of weak intermolecular bonds (hydrogen bonds) between their molecules and those of water. The presence of a very soluble electrolyte (e.g. ammonium sulphate), the ions of which have a marked affinity for water, will reduce the solubility of a non-electrolyte by competing for the aqueous solvent and breaking the intermolecular bonds between the non-electrolyte and water. This effect is important in the precipitation of proteins (p. 60).

(*e*) *Effect of Complex Formation.* The apparent solubility of a solute in a particular liquid may be increased or decreased by the addition of a third substance which forms an intermolecular complex with the solute. The solubility of the complex will determine the apparent change in the solubility of the original solute. For example, the formation of the complexes between 3-aminobenzoic acid and various dicarboxylic acids has been shown to increase the apparent water solubility of the former compound (Wurster and Kilsig, 1965), and Kostenbauder and Higuchi (1956) have shown that soluble and insoluble complexes may be obtained by interactions between various amides and 4-hydroxybenzoic acid, salicylic acid, chloramphenicol, and phenol. Use is also made of complex formation as an aid to solubility in the preparation of solution of mercuric iodide (HgI_2). The latter is not very soluble in water but it is soluble in aqueous solutions of potassium iodide because of the formation of a water-soluble complex, $K_2(HgI_4)$.

(*f*) *Effect of Surface Active Agents.* These compounds are capable of forming large aggregates at certain concentrations in aqueous solutions. Organic compounds with low water solubilities are taken into the interior of these aggregates, and the apparent water solubilities of the organic compounds are increased. The phenomenon is termed solubilisation, and more information is given in Chapter 5.

Dissolution Rates

The dissolution of a solid in a liquid involves the transfer of mass from a solid to a liquid phase. The overall transfer process may be regarded as being composed of two consecutive stages. The first of these, which is an interfacial reaction that results in the liberation of solute molecules from the solid phase, is followed by the transport of solute away from the interfacial boundary under the influence of diffusion or convection. Like any complex reaction that involves consecutive stages, the overall rate of mass transfer in dissolution will be determined by the rate of the slowest stage. If the rates of the two consecutive stages are comparable in magnitude, then both stages will influence the overall rate of transfer.

The Noyes-Whitney Eqn (2.6) indicates that the rate of dissolution (dc/dt) is proportional to the surface area, S, of the solid and the concentration gradient ($C_s - C$), where C_s is the concentration of the substance in a thin saturated liquid film (boundary layer) adjacent to the solid surface, and C is the concentration in the surrounding bulk medium. K is a proportionality constant that is known as the dissolution rate constant.

$$\frac{dc}{dt} = KS(C_s - C) \qquad (2.6)$$

This equation assumes that the rate of mass transfer depends on the rate at which the solute diffuses from the thin boundary layer into the bulk solution. Therefore, K will depend on the diffusion coefficient of the solute and the thickness of the diffusion pathway, and it will be influenced by factors that influence the diffusion coefficient and the film thickness.

FACTORS THAT AFFECT DISSOLUTION RATES

1. *Factors Affecting the Complete System*

(*a*) *Temperature.* It has been shown previously that Le Chatelier's principle will apply to the process of dissolution. An increase in temperature will increase the solubility of a solid with a positive heat of solution. The solid will therefore dissolve at a more rapid rate on heating the system. When complete dissolution has been achieved the system can be cooled to the required temperature, and the substance will remain in solution provided its maximum solubility at the lower temperature is not exceeded. Care should be taken when using this means of increasing the rate of dissolution to ensure that precipitation of the solute does not occur on cooling.

Conversely, a decrease in temperature may be used to increase the dissolution rate of a substance with a negative heat of solution; e.g. paraldehyde.

(*b*) *Agitation.* The rate of transfer of solute from the boundary layer to the surrounding solute will depend on the concentration gradient between these two regions, as indicated by Eqn (2.6), and on the thickness of the diffusion pathway. This latter factor is included in the value of K in Eqn (2.6). Agitation will help to increase a dissolution rate by reducing the thickness of the diffusion pathway and by bringing fresh solvent into contact with the boundary layer, so producing a high value for $(C_s - C)$. The rate of dissolution may therefore be markedly affected by agitation or stirring, and particular care should be paid to this factor in the measurement of these rates. However, it should be borne in mind that the overall rate of mass transfer by dissolution will be independent of agitation if the interfacial reaction that involves the liberation of molecules from the solid phase into the solution is the rate determining stage.

An increase in dissolution rate may also be achieved by proper positioning of the solid in order to take advantage of the difference in densities of the solution and solvent. A solution is usually denser than its solvent, so that if the solid is supported by some means in the upper part of the liquid the denser solution will fall and be replaced by fresh solvent. This process is less efficient than continuous stirring but it is made use of in the extraction of soluble materials from crude drugs.

2. *Changes in the Characteristics of the Solid*

(*a*) *Surface Area.* The Noyes-Whitney Eqn (2.6) shows that the dissolution rate is increased by an increase in the surface area of the solid. Reduction in particle size is effective in creating an increase in surface area, as indicated by Eqn (2.7)—

$$S = \frac{6m}{d\rho} \tag{2.7}$$

where, d is the mean diameter of the particles, m is the mass of the particles, and ρ is the density of the particles. If a unit mass of powder with a density of 1 is considered, then

$$S = \frac{6}{d}$$

and a ten-fold reduction in the mean particle diameter will provide a similar increase in the surface area.

The porosity of the solid particles will also influence the area of contact between the solid and liquid phases. The rate of dissolution of material from the solid surfaces inside pores is less than that from a plane surface because the pathway for diffusion is longer in the former case. The effect of porosity ceases to be of importance when the pores are of molecular dimensions and become too small to allow access of solvent molecules.

(*b*) *Polymorphism.* A substance is said to exhibit polymorphism if it can exist in more than one type of structure, which may be stable or metastable. Polymorphism in solids gives rise to a difference in crystalline form between polymorphs of the same substance. This difference may produce a change in the dissolution rates of the polymorphs. For example, Wurster and Taylor (1965b) have shown that three crystalline forms of prednisolone exhibit different dissolution behaviours, and Tawashi (1968) has reported a marked difference in the dissolution rates of two polymorphic forms of aspirin. The pharmaceutical applications of polymorphism have been reviewed by Haleblian and McCrone (1969).

3. *Changes in the Characteristics of the Solvent*

(*a*) *Viscosity.* An increase in viscosity of the liquid phase will reduce the rate of diffusion of solutes. It is therefore to be expected that dissolution rates dependent on diffusion will be decreased by an increase in viscosity of the solvent, whereas those that are controlled by reactions at the interface will be little affected by changes in viscosity.

(*b*) *Surface Activity.* It has been postulated that increased dissolution rates obtained in the presence of surface active agents may be caused by a lower interfacial tension, which allows better wetting and penetration by the solvent (Taylor and Wurster, 1965). In addition, changes in the extent of etching of crystal surfaces caused by the presence of surface active agents may lead to increased dissolution rates (Westwood *et al.*, 1962).

The whole subject of dissolution rates has been reviewed by Wurster and Taylor (1965a).

Solubility of Solids in Mixtures of Miscible Liquids

The effect of the addition of a miscible liquid to a solution is of importance in certain pharmaceutical processes and systems. For example, resins are soluble in ethanol but not in water. The former solvent is, therefore, often used in the extraction of resins from crude drugs. The alcoholic solution is then concentrated by evaporation, and the resin is precipitated by pouring into an excess of water. The precipitate can then be collected and washed free from water-soluble impurities. Alcoholic solutions (tinctures) of resins are often used in dispensing practice. If these solutions are diluted with water then the resin is precipitated as a sticky mass. This type of precipitate should be avoided by slowly pouring the resinous tincture into an aqueous dispersion of a protective colloid (p. 61). The mixture should be continuously stirred during this process. The resin should then be precipitated as finely divided particles that are readily dispersible in the aqueous vehicle.

The Distribution of Solutes between Immiscible Liquids

If a substance, which is soluble in both components of a mixture of immiscible liquids, is dissolved in such a mixture, then, when equilibrium is attained at constant temperature, it is found that the solute is distributed between the two liquids in such a way that the ratio of the activities of the substance in each liquid is a constant. This is known as the Nernst distribution law, which can be expressed by Eqn (2.8)—

$$\frac{a_A}{a_B} = \text{constant} \tag{2.8}$$

where a_A and a_B are the activities of the solute in solvents A and B, respectively. When the solutions are dilute or when the solute behaves ideally, the activities may be replaced by concentrations (c_A and c_B),

$$\frac{c_A}{c_B} = K \tag{2.9}$$

where the constant K is known as the distribution or partition coefficient. In the case of sparingly soluble substances, K is approximately equal to the ratio of the solubilities (s_A and s_B) of the solute in each liquid; i.e.

$$\frac{s_A}{s_B} = K \tag{2.10}$$

In most other systems, however, deviation from ideal behaviour invalidates Eqn (2.10).

Association or dissociation of the solute molecules in either solvent should be taken into account, since Eqn (2.9) applies only to an equilibrium between solute molecules in the same state in both liquids. For example, if the solute exists as monomers in solvent A and as dimers in solvent B, the distribution coefficient is given by Eqn (2.11), in which the square root of the concentration of the dimeric form is used:

$$K = \frac{c_A}{\sqrt{c_B}} \tag{2.11}$$

If the dissociation into ions occurs in the aqueous layer, B, of a mixture of immiscible liquids, then the degree of dissociation (α) should be taken into account as indicated by Eqn (2.12)

$$K = \frac{c_A}{c_B(1 - \alpha)} \tag{2.12}$$

The solvents, in which the concentrations of the solute—numerators and denominators of Eqns (2.9), (2.11), and (2.12)—are expressed, should be indicated when partition coefficients are quoted. For example, a partition coefficient of 2 for a solute distributed between oil and water may also be expressed as a partition coefficient between water and oil of 0·5.

APPLICATIONS OF THE DISTRIBUTION LAW

1. *Extraction.* Extraction of substances from one phase into another is often used in analytical and organic chemistry and in the removal of active principles from crude drugs. Application of the distribution law to the process of extraction shows that it is more efficient to divide the extracting solvent into a number of smaller volumes that are used in successive extractions rather than to use the total amount of solvent in one single process. (*See* Chapter 22 for more information on extraction, and an example of the method of increasing extraction efficiency.)

2. *Partition Chromatography.* This is a technique used for the separation of components in a mixture. It depends on the difference in the distribution coefficients of the components between two immiscible liquids or between a liquid and a vapour phase. One liquid is maintained stationary by adsorption on to an inert solid support, which may be a powdered solid or strips of a porous material such as filter paper. If a powder is used, then the separation is carried out in a column packed with the powder and its adsorbed liquid phase. The other liquid or vapour is allowed to pass through the column and,

therefore, constitutes a mobile phase. Components of a mixture introduced into the system will become distributed between the mobile and stationary phases in accordance with their partition coefficients, and, provided there is a difference between these coefficients, the components will move at different rates along the column and will eventually become separated.

In paper chromatography, where a filter paper is used to support the stationary liquid phase, the flow of a mobile liquid phase may be made to occur in a vertical or horizontal direction through a strip of paper, or in a radial direction through a paper disc. Whatever the direction of flow the basic principles of the separation technique remain the same as those explained above. More information on the theories and techniques of chromatography is given by Heftmann (1961).

3. *Release of Drugs from Certain Dosage Forms.* Some common dosage forms such as suppositories and ointments are often formulated in water-immiscible bases. The rate of release of medicaments from these dosage forms into aqueous body fluids or secretions will depend on several factors. One of the most important of these is the partition coefficient of the medicament between the base and the body fluid. The effect of partition between water immiscible bases and body fluids is also made use of in the formulation of products intended to provide a prolonged release of drug.

4. *Passage of Drugs through Living Membranes.* The cell membrane is considered to behave as a lipoidal barrier surrounding the cells. One of the main routes of penetration of drugs into cells therefore involves the partition of the substances between these lipoidal layers and the aqueous body fluids with which they are in contact. The partition coefficient of the drug is therefore important in all processes that involve the transport and distribution of drugs throughout the body; e.g. the absorption of drugs from the gastro-intestinal tract, distribution of drugs between various tissues, and penetration of drugs to the sites where they can exert their pharmacological activity (*see* Chapter 6 for more information).

The uptake of preservatives and other substances by micro-organisms is also influenced by the partition coefficient of the compound between the cells and the surrounding aqueous phase (*see* Chapter 31).

5. *Preservation of Emulsions and Creams.* Emulsions and creams are systems comprised of two immiscible phases, one of which is dispersed as globules throughout the other. Micro-organisms usually multiply in the aqueous phase of this type of system, and preservatives must therefore be capable of exerting their activity in this phase. However, most preservatives are usually soluble in both oil and water and will be distributed between these two phases in an emulsion and cream. This distribution will affect the concentration of the substance in the aqueous phase and should be taken into account when deciding the overall concentration of preservative to be used in these systems (*see* Chapter 31).

6. *The Formulation of Solubilised Systems.* Use is often made of compounds known as solubilising agents as a means of increasing the apparent water solubility of organic compounds in the formulation of pharmaceutical preparations. The process is known as solubilisation and it may be regarded as a partition of the organic compound between the interior of the colloidal aggregates (micelles p. 64) formed by the solubilising agents and the surrounding aqueous phase. The activity of the solubilised compound is related to its concentration in the aqueous phase, and a knowledge of the partitioning effect therefore becomes necessary for the proper formulation of such preparations.

7. *Determination of Equilibrium Constants for the Formation of Intermolecular Complexes.* If an intermolecular complex is formed in one phase of an immiscible liquid mixture between a solute *A* that is only soluble in that phase and a second solute *B* that is soluble in both phases, then the partition coefficient of the latter will differ from the value that is observed in the absence of *A*. This change in partition coefficient may be used to determine the equilibrium constant for the formation of the intermolecular complex. This method has been used by Higuchi and his co-workers (1969) to determine the equilibrium constants for a variety of complexes involving compounds of pharmaceutical interest.

Colligative Properties of Solutions

These are the physical properties of a solution that depend on the proportion of dispersed solute particles that are present in the solution. The colligative properties arise from the attractive forces that are exerted by the solute on the solvent. For example, such attractive forces reduce the tendency of the solvent to escape from the liquid as a vapour, and the vapour pressure of the solvent is therefore reduced by the presence of a solute.

The other colligative properties are: (*a*) the elevation of boiling point, (*b*) the depression of freezing point, and (*c*) the osmotic properties. The last of these properties is the most important from a pharmaceutical point of view and the implications of this effect are given by Gunn and Carter (1965).

BIBLIOGRAPHY

BARROW, G. M. (1966) *Physical Chemistry.* 2nd ed. McGraw-Hill, London.

GUCKER, F. T. and SEIFERT, R. L. (1967) *Physical Chemistry.* English Universities Press, London.

HILDEBRAND, J. H. and SCOTT, R. L. (1962) *Regular Solutions.* Prentice Hall Inc., New Jersey.

HOOVER, J. E. (Ed.) (1970) *Remington's Pharmaceutical Sciences.* 14th ed. Chapter 19, Mack Publishing Co., Pennsylvania.

MARON, S. H. and PRUTTON, C. F. (1958) *Principles of Physical Chemistry.* 3rd ed. Macmillan, New York.

MARTIN, A. N., SWARBRICK, J. and CAMMARATA, A. (1969) *Physical Pharmacy.* 2nd ed. Henry Kimpton, London.

ROSE, J. (1961) *Dynamic Physical Chemistry.* Pitman, London.

REFERENCES

BUCKLEY, H. E. (1951) *Crystal Growth.* pp. 29–31. Wiley, New York.

GUNN, C. and CARTER, S. J. (1965) *Dispensing for Pharmaceutical Students.* 11th ed. Pitman Medical, London.

HALEBLIAN, J. and MCCRONE, W. (1969) Pharmaceutical applications of polymorphism. *J. pharm. Sci.*, **58**, 911–929.

HEFTMANN, E. (1961) *Chromatography.* Reinhold, New York.

HIGUCHI, T., RICHARDS, J. H., DAVIS, S. S., KAMADA, A., HOU, J. P., NAKANO, M., NAKANO, N. I., and PITMAN, I. H. (1969) Solvency and hydrogen bonding interactions in non-aqueous systems. *J. pharm. Sci.*, **58**, 661–671.

KOSTENBAUDER, H. B. and HIGUCHI, T. (1956) Formation of molecular complexes by some water soluble amides. I. Interaction of several amides with *p*-hydroxybenzoic acid, salicylic acid, chloramphenicol, and phenol. II. Effect of decreasing water solubility on degree of complex formation. *J. pharm. Sci.*, **45**, 518–522 and 810–813.

MCGLASHAN, M. L. (1968) *Physico-chemical Quantities and Units.* Monographs for teachers, No. 15, Royal Institute of Chemistry, London.

TAWASHI, R. (1968) Aspirin: Dissolution rates of two polymorphic forms. *Science*, **160**, 76.

TAYLOR, JR., P. W. and WURSTER, D. E. (1965) Dissolution kinetics of certain crystalline forms of prednisolone. II. Influence of low concentrations of sodium lauryl sulphate. *J. pharm. Sci.*, **54**, 1654–1658.

WESTWOOD, A. R. C., OPPERHAUSER JR., H., and GOLDHEIM, D. L. (1962) Anomalous solution and etching phenomena. *J. appl. Phys.*, 33, 1764–1766.

WURSTER, D. E. and KILSIG, D. O. (1965) Effect of complex formation on dissolution kinetics of *m*-aminobenzoic acid. *J. pharm. Sci.*, **54**, 1491–1494.

WURSTER, D. E. and TAYLOR JR., P. W. (1965a) Dissolution rates. *J. pharm. Sci.*, **54**, 169–175.

WURSTER, D. E. and TAYLOR JR., P. W. (1965b) Dissolution kinetics of certain crystalline forms of prednisolone. *J. pharm. Sci.*, **54**, 670–676.

3

Phase Equilibria

A PROPER understanding of certain systems and processes that are encountered in pharmaceutical practice necessitates a knowledge of the principles that govern the equilibria between solid, liquid, and gaseous phases.*

The Phase Rule

The conditions relating to physical equilibria between various states of matter are conveniently expressed by the Phase Rule, which was derived by J. Willard Gibbs in 1876. In order to understand this rule it is first necessary to explain what is meant by the terms 'phase', 'number of components', and 'degrees of freedom'.

PHASE

A phase is defined as any homogeneous and physically distinct part of a system that is separated from other parts of the system by definite boundaries. For example, ice, water, and water vapour are three separate phases; each is physically distinct and there are definite boundaries between them. Pure liquids or solutions constitute homogeneous phases, but two immiscible liquids (or solutions) constitute two phases since there is a definite boundary between them. A mixture of gases always constitutes one phase because the mixture is homogeneous and there are no bounding surfaces between the different gases in the mixture.

NUMBER OF COMPONENTS

The number of components of a system is the smallest number of independent chemical constituents necessary to express the concentration of all phases present in the system. For example, in the three-phase system ice, water, and water vapour, the number of components is one, since each phase can be expressed in terms of H_2O. A mixture of salt and water is a two component system since both chemical species are independent.

DEGREES OF FREEDOM

The number of degrees of freedom is the number of variable conditions such as temperature, pressure, and concentration that it is necessary to state in order that the condition of the system at equilibrium may be completely defined. The significance of the number of degrees of freedom of a system will be better understood after considering the specific equilibria that are discussed in the succeeding sections on phase equilibria.

The relationship between the number of phases, P, components, C, and degrees of freedom, F, for equilibria that are influenced only by temperature, pressure, and concentration is given by Eqn (3.1) which is a quantitative expression of the Phase Rule.

$$F = C - P + 2 \qquad (3.1)$$

The application of Eqn (3.1) to various systems of pharmaceutical interest will obviously depend on the number of components present in individual systems. For the convenience of discussing these systems, it is therefore better first to consider those with one component only, and then to move on to those with two components. The succeeding sections are therefore based on such a division, and in each case an attempt will be made to discuss a range of pharmaceutical systems that are examples of each category.

In these discussions the effects of temperature, pressure and composition on the phase equilibria will be indicated by graphs called phase diagrams, which show the variation of a transition temperature such as a boiling point or a melting point with pressure or composition. Representation of the simultaneous effect of three variables would require three axes. This can be achieved with three-dimensional models but if one variable is fixed the resulting planar diagram can be regarded as a section through such a model. The difficulties associated with the representation of three variables do not arise in systems containing one component because

---

* It is assumed that the student is familiar with the kinetic theory of matter and the properties of solids, liquids, and gases.

no variation in the compositions of these systems can occur. It is therefore sufficient to consider only the effects of variation in temperature and pressure.

SYSTEMS OF ONE COMPONENT

The phase diagram for the ice–water–water vapour system (Fig. 3.1) may be used to illustrate the interpretation of these diagrams for one-component systems. This particular diagram is also of importance in the understanding of the process of freeze drying.

In a diagram such as Fig. 3.1, the areas each correspond to a single phase. The number of degrees of freedom is therefore given from Eqn (3.1) as

$$F = 1 - 1 + 2 = 2$$

This means that temperature and pressure can be varied independently within these areas. For example, by varying the temperature and pressure, a mass of water under conditions corresponding to point w_1 in Fig. 3.1 may be converted to a mass at higher temperature and pressure at point w_2; i.e. this independent variation of temperature and pressure has not altered the number of phases in the system. However, if the conditions are such that the system corresponds to a point that lies on one of the lines *AO*, *BO*, or *CO*, then two phases now exist in equilibrium with each other, since these lines form the boundaries between different phases. The number of degrees of freedom is reduced, because, from Eqn (3.1) $F = 1 - 2 + 2 = 1$. This means that a single variable exists when equilibrium is established between two phases, and if the pressure is altered the temperature will assume a particular value or, conversely, if the temperature is altered the pressure will have a definite value.

Melting Points

The boundary *BO* represents the coexistence of liquid water and solid ice at various temperatures and pressures. *BO* therefore indicates the effect of pressure on the melting point of ice, and the negative slope of this line shows that the melting point decreases as the pressure increases. If at any point on this line the pressure is increased while the temperature is maintained constant, then all the ice will be converted to liquid water; i.e. only one phase will remain instead of the two original phases that were in equilibrium at the point on *BO*. Thus, in order to maintain equilibrium conditions between the two phases, the temperature and pressure must not be varied independently of each other.

Boiling Points

The boundary *CO*, which is known as the vapour pressure curve, represents the coexistence of liquid water and water vapour under various conditions.

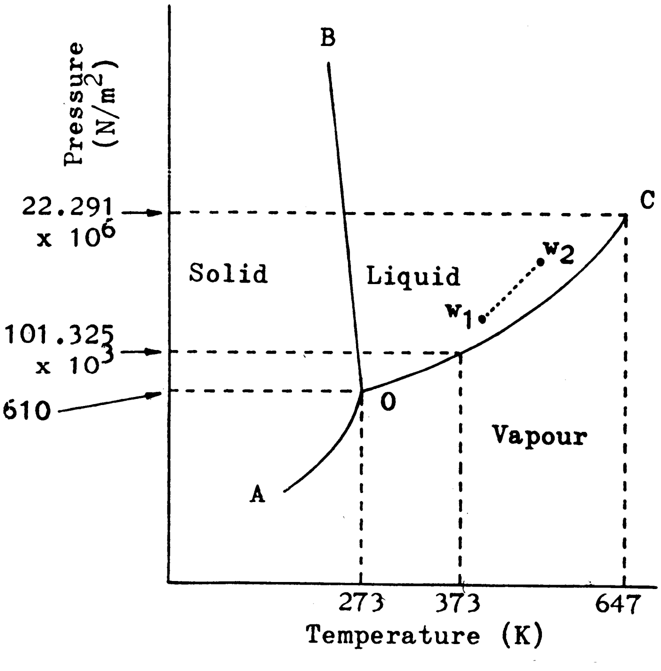


Fig. 3.1 Phase diagram for water at moderate temperatures and pressures (not drawn to scale)

The temperature and pressure again cannot be varied independently otherwise a change from a two-phase system to a single-phase system will occur. For example, if the pressure is kept constant at any point along *CO* while the temperature is increased, then all the water will be converted to vapour and only one phase will remain. *CO* therefore represents the effect of pressure on the boiling point of water and it has an upper limit at the critical temperature of water (647 K*). This is the temperature above which it is impossible to liquefy water vapour.

Equilibria that involve vapours are affected appreciably by variation in pressure. Tabulated boiling points should therefore be quoted at a definite pressure (usually atmospheric pressure, i.e. $1{\cdot}013\,25 \times 10^5$ N/m²), and experimentally obtained values should be converted to values at the appropriate pressure for the purposes of comparison with previously reported ones. Changes in the boiling point of a compound may be calculated from Eqn (3.2), which is known as the Clapeyron equation. (The derivation of this is given in most textbooks of physical chemistry—*see* Bibliography.)

$$\frac{dT}{dp} = \frac{T(V^{\mathrm{G}} - V^{\mathrm{L}})}{L_{\mathrm{V}}} \tag{3.2}$$

In Eqn (3.2), V^{G} and V^{L} are the volumes occupied by one mole of the gas and liquid, respectively, L_{V} is the latent heat of vaporisation of the liquid, T is the thermodynamic temperature, and dT/dp is the change in boiling point with pressure.

A similar Eqn (3.3) can be used to calculate the effect of pressure on melting points, and in this equation V^{S} is the volume occupied by one mole of the solid, L_{f} is the latent heat of fusion of the solid, dT/dp is the change in melting point with pressure, and the remaining terms are defined as before:

$$\frac{dT}{dp} = \frac{T(V^{\mathrm{L}} - V^{\mathrm{S}})}{L_{\mathrm{f}}} \tag{3.3}$$

Since the equilibrium at a melting point does not involve vapour, then the effect of pressure is very small compared with the effect on boiling points.

Triple Points

In Fig. 3.1 the boundary lines meet at *O*, which is the only point in the diagram where three phases may coexist in equilibrium and it is therefore termed a triple point. The application of the Phase Rule equation to the system at *O* shows that

$$F = 1 - 3 + 2 = 0$$

The system is therefore invariant; i.e. any change in pressure or temperature will result in an alteration of the number of phases that are present.

The triple point for water occurs at a temperature of 273·159 8 K and a pressure of 610 N/m². Thus, the triple point temperature is 0·009 8°C above the usual freezing point of water at $1{\cdot}013\,25 \times 10^5$ N/m². However, application of Eqn (3.3) will show that a change in pressure from 610 N/m² to $1{\cdot}013\,25 \times 10^5$ N/m² will produce a decrease in the freezing point of water of only 0·007 5°C. The difference between this calculated value and the observed value is caused by the common presence of air dissolved in water.

Sublimation and Sublimation (Freeze) Drying

The boundary *AO*, which is known as the sublimation pressure curve for ice, indicates the conditions for the coexistence of vapour and solid phases in equilibrium. A mass of ice may be converted directly into water vapour by heating, provided that the pressure is kept below the triple point pressure. This transition is particularly valuable in drying compounds that are sensitive to the higher temperatures usually associated with drying techniques. Removal of water by means of sublimation is termed sublimation or freeze drying and the importance of the triple point in this process should be appreciated at this stage. Further information on the actual drying process is given in Chapter 24.

Polymorphism

Some substances can exist in more than one type of crystal structure. This ability is known as polymorphism and the different structures are termed polymorphs. If the substance is an element then the phenomenon is called allotropy instead of polymorphism.

A reversible change from one polymorph to another frequently occurs at a definite temperature, and both structures can therefore exist in equilibrium at this temperature. Crystalline forms that exhibit this type of behaviour are said to be enantiotropic. For example, the rhombic crystal structure, which is the stable form of sulphur at ordinary temperatures, is converted to the monoclinic structure at 368·5 K, and the transition is reversed at this point on cooling from a higher temperature. Since each enantiomorph represents a separate phase, the number of degrees of freedom that exist when equilibrium is established between them is restricted

---

* The freezing point of water on the thermodynamic temperature scale is equal to 273·15°C. This value has been approximated to 273°C in this chapter unless stated otherwise.

to one, and the transition temperature is therefore affected by pressure. Thus, in a system containing a solid that is able to exist as two polymorphs, an additional line is required in the phase diagram to represent the boundary between the two solid forms. The inclusion of an extra boundary increases the number of triple points in a phase diagram. The sulphur phase diagram is the classical example of such a system and the Bibliography should be consulted for an explanation of this diagram. The phase diagram of water at very high pressures could, in fact, be used to illustrate the phenomenon of polymorphism, because ice may exist in several different forms. However, this would not be very satisfactory as a simple illustration owing to the rather complicated nature of the phase diagram of water under these conditions.

In some cases the change from one polymorphic form to another occurs in one direction only and reversion is not possible in a direct manner. Substances that exhibit this type of polymorphism are termed monotropic. For example, diamond can be converted directly into graphite but the reverse process is not directly possible.

The polymorphic changes between the crystalline forms of fatty acids and glycerides are nearly always monotropic. Theobroma oil, which is used in the preparation of suppositories, is a polymorphous, natural substance. It consists mainly of a single glyceride and usually melts over a narrow temperature range (34 to 36°C), which is just below normal body temperature. The four polymorphic forms of this substance are shown in Table 3.1 together with their melting points. If, during the course of preparation of suppositories, theobroma oil is heated to about 35°C or above and completely liquefied, then the resulting suppositories are too soft for proper administration and tend to melt at ordinary room temperatures. It has been pointed out (Riegelman, 1955) that the excessive heating will cause complete destruction of the nuclei of the stable β form. Consequently, the mass tends to supercool to about 15°C before crystallisation reoccurs in the form of the metastable α, β', and γ forms with a melting point of 22 to 24°C. If the initial heating is limited to about 33°C the mass is sufficiently fluid for pouring, but the nuclei of the stable β form are preserved and cause the separation of β crystals with a melting point of 34·5°C on cooling.

Table 3.1
The Polymorphic Forms of
Theobroma Oil

Polymorph	*m.p.* (°C)
Metastable γ form	18
Metastable α form	22
Metastable β' form	24
Stable β form	34·5

Polymorphism may also be exhibited by liquids. For example, cholesteryl acetate melts to produce a turbid liquid which becomes clear at a higher transition temperature. The turbid and clear forms of polymorphic liquids have different optical properties and the turbid forms have been referred to as liquid crystals, although the terms anisotropic or mesomorphic liquids are preferable. The transition temperature for the change from mesomorphic liquid to clear liquid is pressure dependent.

The different crystal structures of polymorphic forms of the same substance will cause a difference in the thermodynamic activities of the polymorphs (Higuchi and co-workers, 1963). This is of importance in pharmacy, since many drugs exhibit polymorphism and their activities will govern their stabilities and their rates of solution. Thus, one polymorph may be more stable than others. In addition, one may show a greater rate of solution and may therefore be absorbed from the gastro-intestinal tract at a greater rate than other forms, and so produce a higher plasma concentration. For example, the effect of polymorphism on the availability of methylprednisolone and sulphathiazole has been investigated by Higuchi and his co-workers (1963, 1967), and Aguiar and his co-workers (1967) have shown that the polymorphic state of chloramphenicol palmitate has a significant influence on the blood levels of chloramphenicol in humans. If the existence of polymorphism is unrecognised, then the possibility of variation in the availability of a given drug from successive doses may arise. The pharmaceutical applications of polymorphism have been reviewed by Haleblian and McCrone (1969).

SYSTEMS OF TWO COMPONENTS

Table 3.2 shows the effect of the number of phases on the degrees of freedom in a two-component system. When one phase only is present there are

Table 3.2
The Degrees of Freedom in
Two-component Systems

P	*F*
1	3
2	2
3	1
4	0

three degrees of freedom; i.e. temperature, pressure, and composition. Thus, the behaviour of a two component system may be represented completely only by a three-dimensional diagram showing the relations between the three variables. However, it is more convenient to use separate two-dimensional diagrams which show the relation between two of the variables while the third is kept constant; e.g. diagrams showing the variation of pressure with composition at constant temperature, or the variation of composition with temperature at constant pressure.

Solid-Vapour Systems of Two Components

The conversion of an anhydrous salt to a hydrated form, the transition of one hydrated form to a higher hydrate, and the phenomena of deliquescence, hygroscopicity, efflorescence, and exsiccation are examples of the equilibria in systems containing a solid and water vapour.

HYDRATION AND DEHYDRATION OF SALTS

Each hydrated form of a salt exerts a definite vapour pressure at a given temperature. The relation between this value and that of the vapour pressure of water vapour in the atmosphere surrounding the salt is of importance in deciding the type of hydrate formed under various conditions.

EFFLORESCENCE AND EXSICCATION

If the vapour pressure of a hydrated salt is greater than the pressure exerted by the water vapour in the surrounding atmosphere then the salt will attempt to attain equilibrium with its surroundings, and therefore tend to lose water to form a lower hydrate or an anhydrous salt. This phenomenon is known as efflorescence.

The pressure of water vapour in the atmosphere is about $13{\cdot}33 \times 10^2$ N/m^2 at 293 K, and therefore hydrates with vapour pressures greater than this will tend to exhibit efflorescence and be unstable, provided that the lower hydrate that is formed still exerts a vapour pressure greater than the surrounding atmosphere. If this is not so, then water will be taken up from the atmosphere by the lower hydrate as fast as it is formed and the final equilibrium will depend on the rates at which water is lost or taken up by the two hydrates. For example, the behaviour of the various forms of sodium carbonate may be represented by the following scheme—

$Na_2CO_3.10\,H_2O$ v.p. $= 32 \times 10^2$ N/m^2 at 293 K

$\downarrow$ spontaneous dehydration; i.e. efflorescence.

$Na_2CO_3.H_2O$ v.p. $= 16 \times 10^2$ N/m^2 at 293 K

$\uparrow\downarrow$ efflorescence not observable because anhydrous salt is rapidly hydrated.

Na_2CO_3 (anhyd.) v.p. $= 0$

Since the vapour pressure exerted by the decahydrate is much greater than that of normal atmosphere it loses water by the process of efflorescence and is converted to the monohydrate. The vapour pressure of the latter is still above that of the atmosphere, but further apparent loss of water does not occur since the anhydrous salt is rehydrated at a faster rate than dehydration of the monohydrate.

The vapour pressure of hydrated salts, and therefore the rate of efflorescence, increases with rise in temperature. The process of accelerating the rate of efflorescence by increasing the temperature in order to remove water of crystallisation from a hydrated salt is known as exsiccation, although this term is also used where water is not normally lost by efflorescence. For example, the pentahydrate of copper sulphate may be converted to the trihydrate by heating to 303 K. Two further molecules of water are removed at 373 K to yield the monohydrate, and the remaining molecule of water is removed at 473 K to yield the anhydrous salt.

Since the instability that arises from efflorescence is caused by the loss of water vapour, the common method of minimising such deterioration involves the use of containers that prevent the loss of water vapour. The additional precautions of using well-filled containers with a minimum amount of atmosphere above the efflorescent material and storage in a cool place are also advisable.

DELIQUESCENCE AND HYGROSCOPICITY

Both of these terms are used to indicate that a material takes up water vapour from the atmosphere and is converted to a more hydrated form. In the case of a hygroscopic substance the more hydrated state is still a solid but deliquescence implies the eventual formation of a liquid phase; i.e. a solution. In both phenomena, however, the final more hydrated state must still exert a lower vapour pressure than that of the water vapour in the surrounding atmosphere. If this is not so then the newly formed hydrated state will immediately lose water by efflorescence and revert to the initial state.

Thus, for a liquid phase to be produced by deliquescence, it is necessary that the vapour pressure exerted by a saturated solution of the deliquescent material should be less than $13{\cdot}33 \times 10^2\ \mathrm{N/m^2}$. The following scheme showing the behaviour of sodium hydroxide may be used as an example of deliquescence.

$NaOH.H_2O$ v.p. is very low at 293 K

$\downarrow$ Deliquescence

Saturated solution of NaOH v.p. $= 1{\cdot}33 \times 10^2\ \mathrm{N/m^2}$ at 293 K

$\downarrow$ Further deliquescence

Unsaturated solution of NaOH v.p. = v.p. of atmospheric moisture

Other deliquescent materials include potassium hydroxide, sodium lactate and potassium carbonate, while examples of hygroscopic materials include exsiccated sodium sulphate, ammonium chloride, and squill.

Storage precautions for pharmaceutical preparations that are deliquescent or hygroscopic are aimed at the maintenance of a moisture-free atmosphere inside the container. The closures of the latter should therefore prevent the access of water vapour, and official monographs usually direct that such substances should be stored in 'well-closed' containers. In addition, a well-filled container limits the volume of atmosphere in the container and, therefore, further reduces the uptake of moisture by the product. In certain cases, where the product is particularly susceptible to moisture, a drying agent may be placed inside the container. The drying agent is usually contained in small packets, made from a material that is pervious to water vapour, in order to prevent contact between the agent and the product. Silica gel is often used in this way and it may contain an indicator to show when its drying properties are no longer satisfactory. Anhydrous cobaltous chloride, which is blue, may be used as an indicator, since it is converted to a pink hydrate when the silica gel has adsorbed its maximum amount of water vapour.

Liquid-Liquid Systems of Two Components; Solutions of Liquids in Liquids

The degree of miscibility of two liquids may be used as a basis for the classification of these systems into the following types:

1. Completely miscible liquids.
2. Partially miscible liquids.
3. Immiscible liquids.

1. COMPLETELY MISCIBLE LIQUIDS

A solution of one liquid in another in contact with vapour from the liquid mixture constitutes a two-phase system of two components. The Phase Rule therefore indicates that such a system will possess two degrees of freedom; i.e.

$$F = 2 - 2 + 2 = 2.$$

Thus, the system will be completely defined by two variables. For example, if the temperature and composition are fixed, the vapour pressure must have a definite value, or if the pressure and composition are fixed then the equilibrium will only be maintained at a particular temperature.

In diagrams involving vapour pressures exerted by solutions of liquids in liquids it is necessary to consider the application of Raoult's law. This states that the partial vapour pressure exerted by each component is proportional to its molar concentration in the solution. The law may be expressed by Eqn (3.4):

$$p = p_A + p_B = p_A^0 x_A + p_B^0 x_B \qquad (3.4)$$

where p is the total vapour pressure above a liquid mixture containing x_A and x_B mole fractions of components A and B, respectively, p_A^0 and p_B^0 are the vapour pressures exerted by the pure components, and p_A and p_B are the partial vapour pressures exerted by the components in the liquid mixture.

Ideal Solutions. Raoult's law is obeyed by only a few solutions of liquids in liquids. However, it is convenient to consider an ideal solution, to which this law applies, and to discuss real solutions in terms of the deviations in their behaviour from that of an ideal solution.

A graph showing the variation in the partial pressure of each component in an ideal solution, with the composition of the solution at constant temperature, should produce a straight line which passes through the origin, since $p_A = p_A^0 x_A$ and $p_B = p_B^0 x_B$. The line for each component can be represented on the same diagram as shown in Fig. 3.2, where $O_A p_A^0$ and $O_B p_B^0$ indicate the variations in partial pressures exerted by components A and B, respectively, with the composition of the mixture. From Eqn (3.4), and by applying simple geometry, it can be shown that the line $p_A^0 p_B^0$ will indicate the variation in the total vapour pressure with composition.

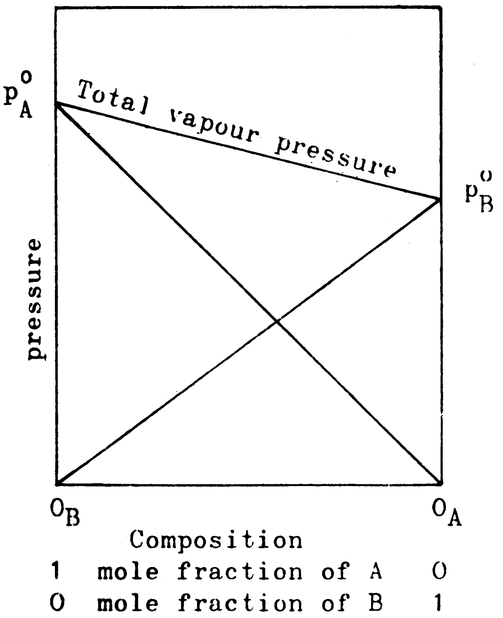


Fig. 3.2 Vapour-composition diagram for an ideal solution

Real Solutions. As previously stated, only a few actual solutions show ideal behaviour. The components of these ideal solutions have a similar chemical structure; e.g. benzene and toluene, *n*-hexane and *n*-heptane, ethyl bromide and ethyl iodide.

Most systems show varying degrees of deviation from Raoult's law, depending on the nature of the liquids and on the temperature. When interactions such as hydrogen bonding, salt formation, or hydration occur between the components of a solution, then the vapour pressure of each component is lowered with respect to the behaviour of an ideal solution and the system is said to exhibit a negative deviation from Raoult's law; e.g. chloroform and acetone, pyridine and acetic acid, water and nitric acid.

In most systems, the vapour pressures are greater than those of an ideal solution and positive deviations from Raoult's law are therefore exhibited. This type of behaviour occurs when the components differ in their polarity, length of hydrocarbon chain, and degree of association; e.g. carbon tetrachloride and cyclohexane, benzene and ethanol, water and ethanol. The degree of deviation from Raoult's law decreases as the temperature increases, since the effects of the differences in the natures of the components are reduced at higher temperatures. Conversely, a decrease in temperature may lead to a decrease in miscibility of the two components and phase separation may occur.

The occurrence of deviations from ideal behaviour allows the classification of solutions of liquids in liquids into three types.

(*a*) Systems where the total vapour pressure is always intermediate between those of the pure components; i.e. there is neither a maximum nor minimum in the vapour pressure-composition diagram as shown in Fig. 3.3(*a*). These systems are known as zeotropic mixtures and examples include carbon tetrachloride and cyclohexane, and water and methanol.

(*b*) Systems that exhibit a maximum value in the vapour pressure-composition diagrams as shown in Fig. 3.3(*b*). These are known as azeotropic mixtures with a maximum vapour pressure or minimum boiling point and examples include benzene and ethanol, and water and ethanol.

(*c*) Systems that exhibit a minimum value in the vapour pressure-composition diagrams as shown in Fig. 3.3(*c*). These are known as azeotropic mixtures with a minimum vapour pressure or maximum boiling point. and examples include chloroform and acetone, pyridine and acetic acid, and water and nitric acid.

The effects of these different types of behaviour on the results of distillation of liquid mixtures is of importance in various pharmaceutical fields. However, before these distillations can be considered, it is necessary to take into account the composition of the vapour that is in equilibrium with the liquid mixtures of different compositions. The previous vapour pressure diagrams show only the relation between vapour pressure and composition of the liquid phase. If these diagrams are drawn to show the variation in vapour pressure with both vapour and liquid compositions, the results may be represented by Fig. 3.4(*a*), (*b*), and (*c*), respectively.

The upper curves in these figures represent the variation in total vapour pressure with composition of the liquid phase, and the lower curves represent the variation in total vapour pressure with composition of the vapour phase. The different areas correspond to the existence of liquid, vapour, or liquid plus vapour, as shown in the diagrams.

When equilibrium is established between liquid and vapour phases the vapour pressure must be constant. It can therefore be seen that the compositions of the vapours (v_1 to v_5) that are in equilibrium with liquids of compositions given by l_1 to l_5 are obtained by drawing horizontal tie-lines through the points on the liquid composition curves that correspond to l_1 to l_5. The points at which these lines intersect the vapour curves provide the compositions

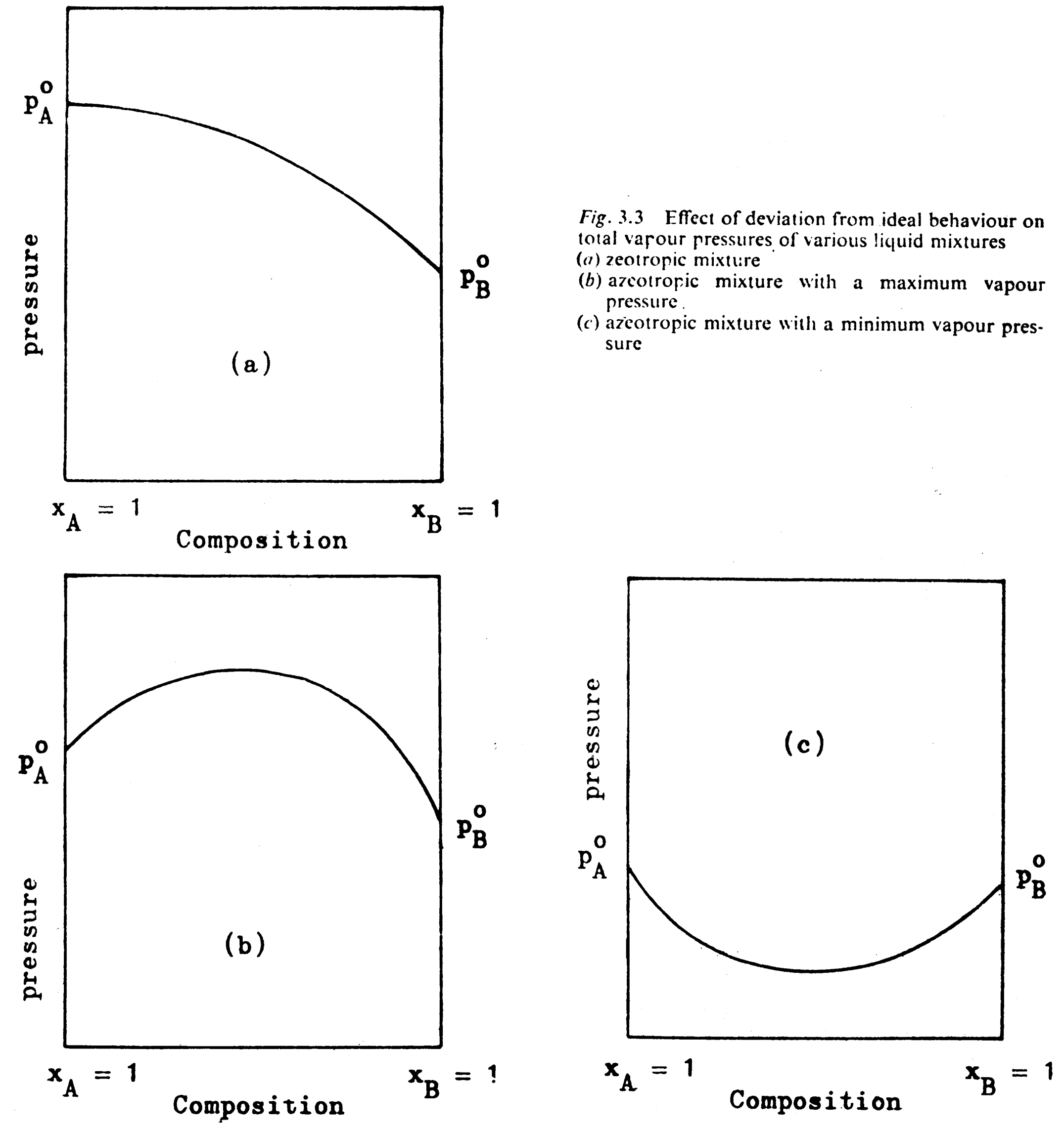


Fig. 3.3 Effect of deviation from ideal behaviour on total vapour pressures of various liquid mixtures
(*a*) zeotropic mixture
(*b*) azeotropic mixture with a maximum vapour pressure.
(*c*) azeotropic mixture with a minimum vapour pressure

v_1 to v_5. It can be seen from Fig. 3.4(*a*) that the vapour phase in equilibrium with a particular liquid composition is richer in the more volatile component *A*; i.e. the component with the higher vapour pressure. This is known as Konowaloff's Rule.

The point *M* in Fig. 3.4(*b*) and (*c*) corresponds to the formation of an azeotrope, which is a mixture with a lower or higher vapour pressure than is exerted by any other composition in the system. It will be observed that the liquid and vapour have identical compositions at this point.

Distillation of Solutions of Liquids in Liquids

(*a*) *Zeotropic Mixtures.* For the purpose of explaining the effects of distillation it is more convenient to use a phase diagram (Fig. 3.5) that shows the variation in boiling point with composition of the liquid and vapour phases at constant pressure. It should be observed that the upper and lower curves in Fig. 3.5 represent the vapour composition and the liquid composition, respectively, and that the areas corresponding to liquid and vapour phases are

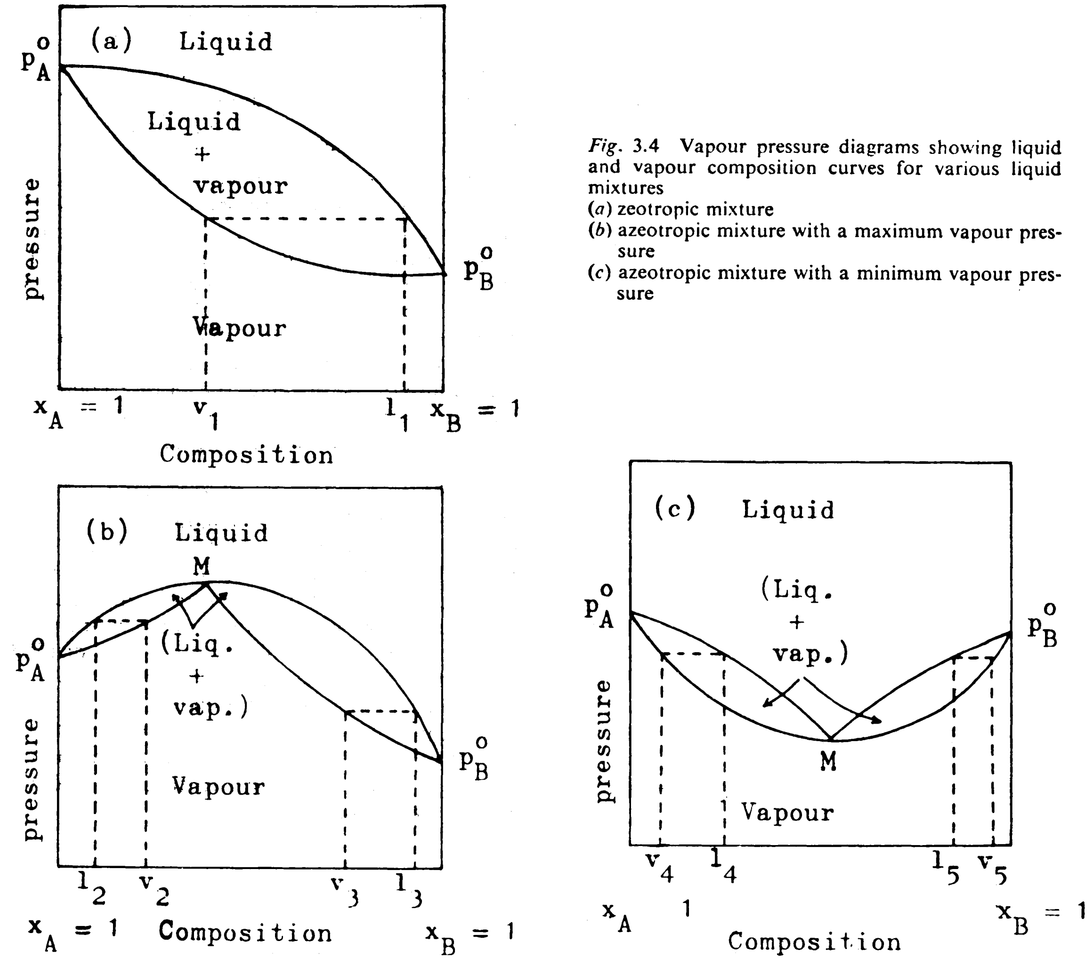


Fig. 3.4 Vapour pressure diagrams showing liquid and vapour composition curves for various liquid mixtures
(*a*) zeotropic mixture
(*b*) azeotropic mixture with a maximum vapour pressure
(*c*) azeotropic mixture with a minimum vapour pressure

transposed when compared with the vapour pressure diagrams shown in Fig. 3.4(*a*). Konowaloff's Rule can still be seen to apply in the boiling point diagram, since a liquid with a composition corresponding to l_1 will boil at temperature T_1 and be in equilibrium with vapour of composition l_2. This vapour is therefore richer in component *A*, which has the lower boiling point (T_A) and is therefore the more volatile component of the liquid mixture.

It can also be seen from Fig. 3.5 that if the vapour of composition l_2 is removed and condensed it will give a liquid of composition l_2. If this liquid is subsequently heated it will boil at temperature T_2 to provide a vapour of composition l_3 that is even richer in component *A*; i.e. the composition of the distillate will approach closer to pure *A* as more stages of heating and condensation are involved. Conversely, as vapour that is richer in *A* is removed from the distillation flask the composition of the liquid remaining in the flask gradually approaches pure *B*. Thus, the components of a zeotropic mixture may be separated completely by the process of fractional distillation, which involves the occurrence of many individual stages of vaporisation and condensation in a distillation column. (*See* Chapter 23 for more information on distillation as a unit operation.)

If the Phase Rule is applied to the two-component system in the distillation flask where two phases (i.e. liquid and vapour) are present, it can be shown that two degrees of freedom exist:

$$F = 2 - 2 + 2 = 2$$

Since the pressure is kept constant, the temperature will therefore change as the composition varies in order to maintain the same number of phases; i.e.

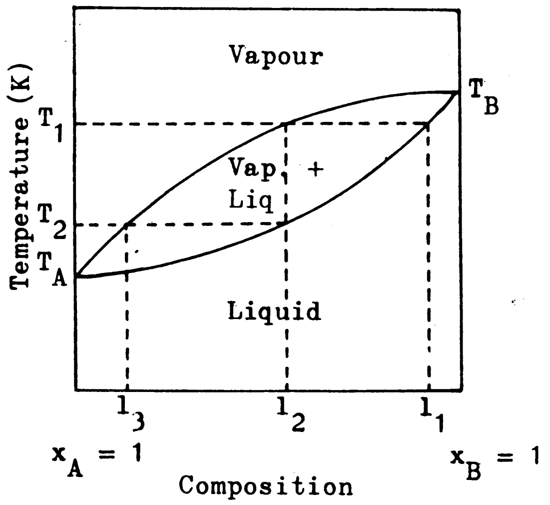


Fig. 3.5 Boiling point-composition diagram for a zeotropic system

the boiling point of the liquid remaining in the flask increases as its composition approaches pure *B*.

(*b*) *Azeotropic Mixtures with a Maximum Boiling Point.* The boiling point diagram for this type of system may be represented by Fig. 3.6(*a*). The same sequence of events and the same reasoning that was involved in the distillation of a zeotropic mixture may be applied to Fig. 3.6(*a*), taking a liquid with a composition on either side of that corresponding to *M* as the starting point. It will be found that a complete separation into pure *A* and pure *B* cannot be achieved whatever the initial composition of the liquid mixture. This is because the liquid and vapour composition curves are coincident at *M*, and a liquid with a composition corresponding to this point boils at a maximum temperature for the system and produces a vapour with the same composition. These mixtures can therefore be separated only by fractional distillation into pure *A* or pure *B*, as the distillate, and a constant boiling mixture of composition *M*, which remains in the distillation flask. The nature of the distillate (i.e. *A* or *B*) depends on the composition of the initial mixture with respect to that of the constant boiling azeotropic mixture.

(*c*) *Azeotropic Mixtures with a Minimum Boiling Point.* The boiling point diagram for this type of system is shown in Fig. 3.6(*b*). Fractional distillation of these mixtures will allow separation of the mixture into pure *A* or pure *B* only, and a constant boiling mixture with a composition corresponding to *M*. Pure *A* or *B* will remain in the distillation flask after the constant boiling mixture has been completely removed at a minimum boiling point. The nature of the pure liquid obtained in the flask (i.e. *A* or *B*) will depend on the composition of the initial mixture with respect to that of the constant boiling azeotrope.

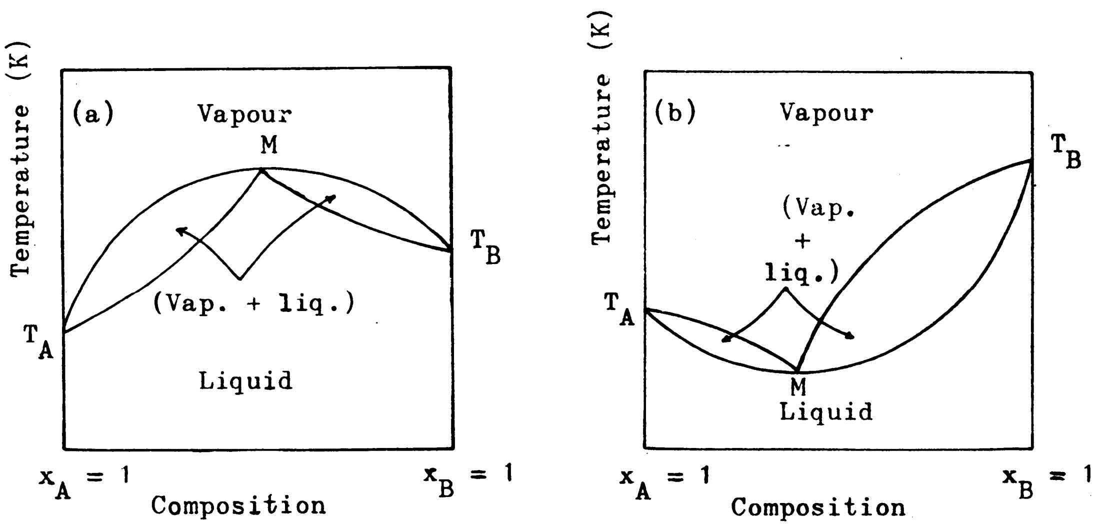


Fig. 3.6 Boiling point-composition diagrams for azeotropic mixtures
(*a*) maximum boiling point azeotrope (*b*) minimum boiling point azeotrope

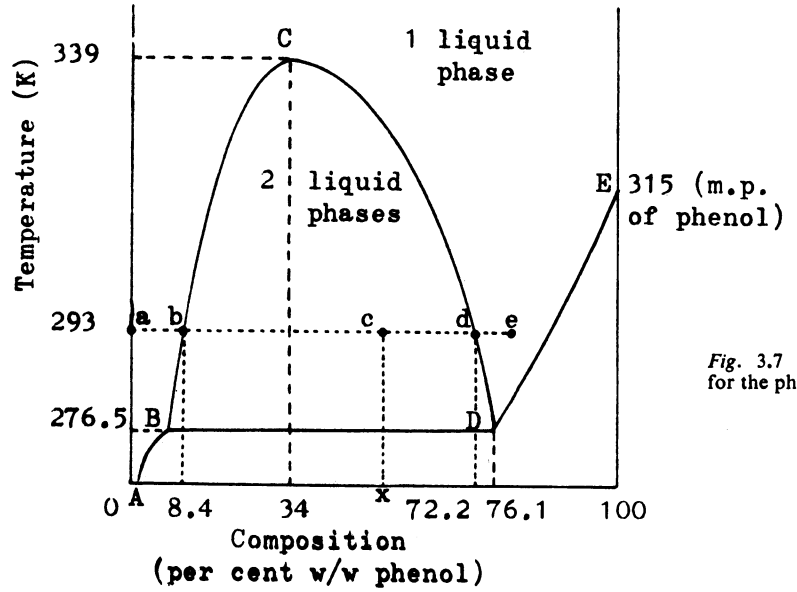


Fig. 3.7 Temperature-composition diagram for the phenol-water system (at 101 325 N/m^2)

2. PARTIALLY MISCIBLE LIQUIDS

These systems may be divided into the following types for the convenience of discussion.

(*a*) *Systems Showing an Increase in Miscibility with Rise in Temperature.* A positive deviation from Raoult's law arises from a difference in the cohesive forces that exist between the molecules of each component in a liquid mixture. This difference becomes more marked as the temperature decreases and the positive deviation may then result in a decrease in miscibility sufficient to cause the separation of the mixture into two phases. Each phase consists of a saturated solution of one component in the other liquid. Such mutually saturated solutions are known as conjugate solutions.

The equilibria that occur in mixtures of partially miscible liquids may be followed either by shaking the two liquids together at constant temperature and analysing samples from each phase after equilibrium has been attained, or by observing the temperature at which known proportions of the two liquids, contained in sealed glass ampoules, become miscible, as shown by the disappearance of turbidity.

Phenol-Water System. The temperature-composition diagram of phenol and water at constant pressure (Fig. 3.7) is convenient to use in the explanation of the effects of partial miscibility in systems that show an increase in miscibility with rise in temperature.

The areas shown in Fig. 3.7 each correspond to the existence of various phases as shown in the diagram. The most important part of the diagram, for the purpose of the present discussion, is that indicated by the line *BCD*, which separates a single-phase system of one liquid from a two-phase system of two mutually saturated liquids. If gradually increasing amounts of phenol are added to water at 293 K, the composition moves along the line *abcde*. Between *a* and *b* there is only one liquid phase, and application of the Phase Rule shows that there are three degrees of freedom:

$$F = 2 - 1 + 2 = 3$$

This means that temperature and composition must be specified in order to define the system completely at constant pressure. The aqueous solution is eventually saturated at a composition corresponding to *b* (containing 8·4 per cent phenol). The line *BC* therefore represents the effect of temperature on the solubility of phenol in water. If more phenol is added, then a second layer separates. This is a saturated solution of water in liquid phenol. Thus, at a total composition corresponding to point *c* (i.e. containing *x* per cent phenol) two conjugate solutions will exist as separate phases. The compositions of these phases will correspond to points *b* and *d* respectively (i.e. one solution contains 8·4 per cent phenol in water and the other contains 27·8 per cent water in phenol). The relative amounts of these solutions at 293 K are given by

$$\frac{\text{Amount of saturated solution of phenol in water}}{\text{Amount of saturated solution of water in phenol}} = \frac{cd}{bc}$$

Application of the Phase Rule to the system at c shows that there are two degrees of freedom:

$$F = 2 - 2 + 2 = 2$$

Thus, the system is completely defined if the temperature is specified at constant pressure. For example, at 293 K the system consists of one phase with a composition corresponding to b and another with a composition corresponding to d.

If more phenol is added, the point d is reached where the system consists of a saturated solution of phenol in water. Thus, the line DC represents the effect of temperature on the solubility of water in phenol. Further addition of phenol produces an unsaturated solution of water in phenol (i.e. at e).

It can be seen that BC and DC meet at C; i.e. above C only one liquid phase can exist. The two components are therefore miscible in all proportions above C, which is known as the upper critical temperature (upper CST). The composition at C corresponds to a mixture containing 34 per cent phenol and such a mixture will remain separated into two layers up to a higher temperature than any other. Application of the Phase Rule to the system at C shows that this point is invariant; i.e. the upper CST is fixed at 339 K for the phenol-water system at atmospheric pressure (101 325 N/m^2).

The remaining part of the phenol-water phase diagram is of importance with regard to the formulation of a solution containing a large proportion of phenol. Such a solution is used in dispensing practice in preference to solid phenol, since the latter is deliquescent and therefore difficult to weigh accurately. In order to be satisfactory for dispensing purposes the liquid should remain homogeneous at normal room temperatures; i.e. it should not readily deposit crystalline phenol when the room temperature falls during cool periods. It can be seen from Fig. 3.7 that a solution containing 76·1 % w/w phenol will meet this requirement most satisfactorily since this mixture will withstand the lowest possible temperature (276·5 K) before solid phenol is deposited. On the basis of his investigation into this section of the phenol-water diagram Mulley (1959) has suggested that a solution containing 76·1 % w/w phenol, which corresponds to an 80% w/v solution in water, should be used in dispensing. The percentage w/v concentration allows more rapid calculation of the weight of phenol in a given volume of solution than the official solution known as Liquefied Phenol BP 1968, which contains 80% w/w phenol in water. In addition, the official formulation deposits phenol at a higher temperature (about 283 K).

(*b*) *Systems Showing a Decrease in Miscibility with Rise in Temperature.* A few mixtures, which probably involve compound formation, exhibit a lower critical solution temperature (lower CST), e.g. triethylamine plus water, paraldehyde plus

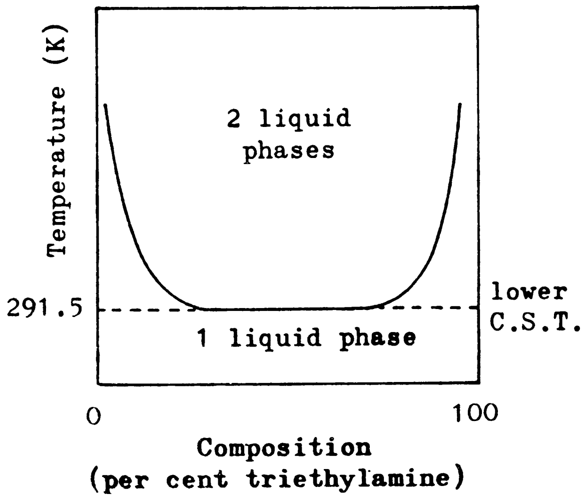


Fig. 3.8 Temperature-composition diagram for the triethylamine-water system (at 101 325 N/m^2)

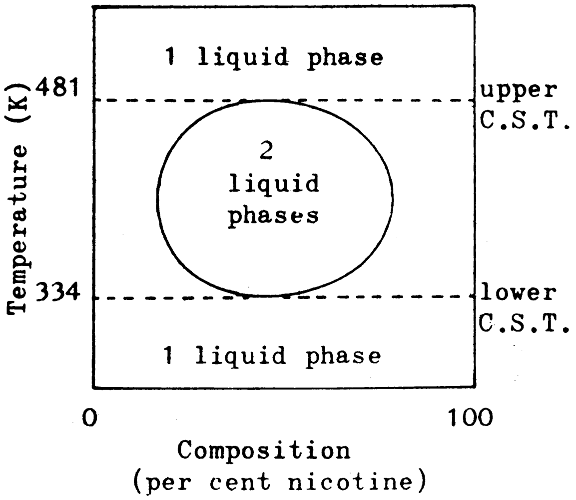


Fig. 3.9 Temperature-composition diagram for the nicotine-water system (at 101 325 N/m^2)

water. The formation of a compound produces a negative deviation from Raoult's law, and miscibility therefore increases as the temperature falls, as shown in Fig. 3.8.

The effect of temperature on miscibility is of use in the preparation of paraldehyde enemas, which usually consist of a solution of paraldehyde in normal saline. Cooling the mixture during preparation allows more rapid solution, and storage of the enema in a cool place is recommended. (N.B. Enemas should be warmed to body temperature before use.)

(*c*) *Systems Showing Upper and Lower CSTs.* The decrease in miscibility with increase in temperature in systems having a lower CST is not indefinite. Above a certain temperature, positive deviations from Raoult's law become important and miscibility starts to increase again with further rise in temperature. This behaviour produces a closed-phase diagram as shown in Fig. 3.9, which represents the nicotine-water system.

In some mixtures where an upper and lower CST are expected, these points are not, in fact, observed since a phase change by one of the components occurs before the relevant CST is reached. For example, the ether-water system is expected to exhibit a lower CST, but water freezes before the temperature is reached.

The Effects of Added Substances on Critical Solution Temperatures

It has already been stated that a CST is an invariant point at constant pressure. These temperatures are very sensitive to impurities or added substances. In general, the effects of additives may be summarised by Table 3.3.

The increase in miscibility of two liquids caused by the addition of a third substance is referred to as blending. This is made use of in the formulation of Solution of Cresol with Soap BP 1968, which contains 50 per cent cresol. Cresol is only partially miscible with water but the soap in this preparation decreases the upper CST and produces complete miscibility at ordinary temperatures.

Table 3.3
The Effects of Additives on CST

Type of CST	*Solubility of additive in each component*	*Effect on CST*	*Effect on miscibility*
Upper	Approx. equally soluble in both components	Lowered	Increased
Upper	Readily soluble in one component but not in other	Raised	Decreased
Lower	Approx. equally soluble in both components	Raised	Increased
Lower	Readily soluble in one component but not in other	Lowered	Decreased

3. IMMISCIBLE LIQUIDS

In a mixture containing two immiscible liquids each liquid exerts its own vapour pressure independently of the other. The total vapour pressure, p, is therefore equal to the sum of the separate vapour pressures of the pure compounds (p_A^0 and p_B^0, respectively), i.e.

$$p = p_A^0 + p_B^0$$

The application of the Phase Rule to the distillation of such a system, which involves two liquids in equilibrium with a vapour, shows that only one degree of freedom exists,

$$F = 2 - 3 + 2 = 1$$

i.e. the boiling point is constant at constant pressure and is independent of the composition. Since the liquid mixture will boil when its total vapour pressure is equal to atmospheric pressure, it follows that the boiling point is lower than that of either pure component, even the one with the lowest boiling point. After one of the components has been completely removed by distillation the boiling point will rise to that of the remaining component.

The constant boiling mixture will produce a vapour in which the number of molecules of each component (n_A and n_B, respectively) is proportional to its vapour pressure if ideal behaviour occurs, i.e.

$$\frac{n_A}{n_B} = \frac{p_A^0}{p_B^0} \tag{3.5}$$

But

$$n_A = \frac{m_A}{M_A} \quad \text{and} \quad n_B = \frac{m_B}{M_B} \tag{3.6}$$

where m_A and m_B are the masses of each component in the vapour and M_A and M_B are their molecular weights, and from Eqns (3.5) and (3.6)

$$\frac{m_A}{M_A} \times \frac{M_B}{m_B} = \frac{p_A^0}{p_B^0}$$

$$\frac{m_A}{m_B} = \frac{p_A^0 M_A}{p_B^0 M_B} \tag{3.7}$$

Equation (3.7) indicates that the ratio of A to B by weight in the distillate is proportional to the ratios of their partial pressures and their molecular weights. Thus, if an immiscible mixture of water and an organic compound is distilled, a high proportion of the distillate will be comprised of the organic compound, since water has a low molecular weight. This process is known as steam distillation and it is often used in pharmaceutical chemistry as a means of purifying organic compounds that do not react and are immiscible with water. In practice, the process may also be applied to liquids that are partially miscible with water (e.g. aniline). It is especially useful in the purification of liquids with high boiling points, e.g. essential oil of almond, and in the extraction of volatile oils, e.g. clove and eucalyptus oils, which are obtained at the lower distillation temperature without decomposition or loss of aroma.

Liquid-Gas Systems of Two Components; Solutions of Gases in Liquids

A saturated solution of a gas in a liquid is produced when equilibrium is attained between the dissolved gas and that which remains undissolved above the liquid. The amount of gas that will dissolve depends on the temperature and the pressure of the gas, provided that no reaction occurs between the gas and the liquid.

The effect of pressure is indicated by Henry's law, which states that 'at constant temperature, the solubility of a gas in a liquid is directly proportional to the pressure of the gas above the liquid'. The law may be expressed by Eqn (3.8)

$$x = kp \qquad (3.8)$$

where, x is the mole fraction of the dissolved gas, p is the partial pressure of the gas above the solution, and the proportionality constant k is known as Henry's law constant.

The effect of gas pressure is made use of in preparing effervescent solutions containing carbon dioxide, which is maintained in solution by increasing the pressure of carbon dioxide in the atmosphere above the liquid in the container. It should be borne in mind that the solubility of carbon dioxide does not follow Henry's law, because a reaction occurs (i.e. $CO_2 + H_2O \rightleftharpoons H_2CO_3$). However, the equilibrium of this reaction will be influenced by the solubility of CO_2, which will, in turn, be influenced by the pressure.

Equation (3.8) also applies to the solubility of each gas in a solution of several gases in the same liquid, provided that x and p represent the mole fraction and partial pressure, respectively, of a particular gas. Henry's law is most applicable at high temperatures and low pressures. As the pressure rises the value of k may vary.

The solubility of most gases in liquids decreases as the temperature rises. This provides a means of removing dissolved gases. For example, carbon dioxide-free Water for Injections may be prepared by boiling Water for Injections and preventing the access of atmosphere during cooling. The presence of electrolytes may also decrease the solubility of a gas in water by a 'salting out' process, which is caused by the marked attraction exerted between the electrolyte and water.

Solid-Solid Systems of Two Components; Solutions of Solids in Solids

When two solids are melted together and the resultant liquid is cooled, then the components may be deposited either independently or as a single homogeneous phase, which is known as a solid solution. Both components participate in the lattice structure of the solid solution.

Solid solutions may be classified in a similar manner to solutions of liquids in liquids, into completely and partially miscible mixtures and immiscible mixtures. The Bibliography should be consulted for phase diagrams of these systems.

The production of alloys often involves the formation of solid solutions; e.g. gold/silver, gold/platinum. However, Goldberg and his co-workers (1965, 1966a, b, and c) have investigated the usefulness of solid solutions of urea and slowly soluble drugs such as griseofulvin and chloramphenicol. They suggest that when in contact with water, urea is dissolved rapidly from these systems and the remaining drug is in a molecular state of subdivision and therefore dissolves more rapidly than the original compound.

Solid-Liquid Systems of Two Components; Solutions of Solids in Liquids

The effect of pressure on a solid–liquid system is small. The 'reduced Phase Rule' is therefore often used in connection with these systems and this is expressed by Eqn (3.9),

$$P + F' = C + 1 \qquad (3.9)$$

where, P and C are defined as before, and F' is the number of degrees of freedom in addition to pressure. Any system in which a vapour phase is not taken into account is known as a condensed system. Equilibria in these systems are therefore represented by temperature-composition diagrams at constant pressure, and the greatest value of F' is two.

The behaviour of two component solid–liquid systems can be classified into three types.

1. Systems that show the formation of a eutectic mixture.
2. Systems that show the formation of a compound with a congruent melting point (i.e. the compound, which consists of one component solvated by the other, yields a liquid with the same composition as the compound on melting).
3. Systems that show the formation of a compound with an incongruent melting point (i.e. the compound undergoes fusion on heating to a certain temperature and produces a liquid and a new solid phase, the composition of which is different from that of the original compound).

Systems of the first type, which involve the formation of a eutectic mixture, have found certain applications of pharmaceutical interest. The other types are of less importance. No further comment will be made regarding these, and the Bibliography should be consulted if more information is required.

Formation of Eutectic Mixtures

The ice-potassium chloride system, whose behaviour is represented by the phase diagram shown in Fig. 3.10, may be used as an example of the type of system that involves the formation of a eutectic.

In Fig. 3.10, *A* and *B* represent the melting points of ice and KCl, respectively. If KCl is added to water, the freezing point of the latter is reduced as indicated by *AC*, which therefore represents the effect of composition on the temperature at which ice separates from the system. Similarly, if water is added to KCl the melting point of the latter is lowered, and *BC* therefore represents the effect of composition on the temperature at which solid KCl separates from the system.

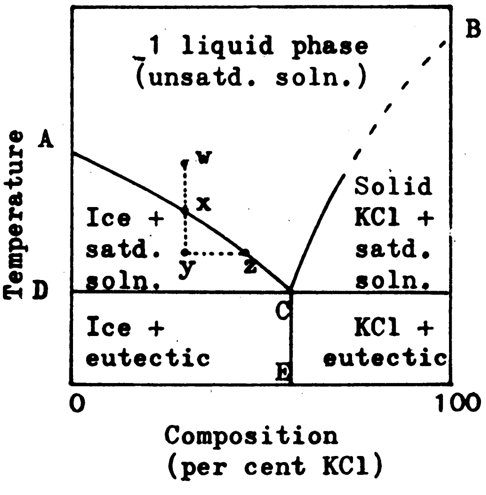


Fig. 3.10 Temperature-composition diagram for the KCl-water system

At *C*, both solid components can exist in equilibrium with a liquid of definite composition. Application of the 'reduced Phase Rule' shows that the system is invariant at this point since there are no degrees of freedom.

$$F' = 2 + 1 - 3 = 0$$

This means that the mixture will freeze completely at a constant temperature *D*, which is lower than the freezing points of either pure component.

Further understanding of the phase diagram may be obtained by considering the effects of cooling a solution of KCl in water represented by point *w* in Fig. 3.10. If the solution is cooled to point *x* on the freezing curve *AC*, then some ice will separate out. Further cooling to *y* will produce more ice and the remaining solution will become more concentrated since it will contain all the original KCl. The composition of this remaining solution will correspond to point *z*. Therefore, as the temperature falls, the composition of the remaining liquid moves along *AC*. A limit is reached when the remaining solution is saturated with KCl, and, on cooling this solution, ice and KCl separate out in the same ratio in which they exist in the saturated solution, and the temperature remains constant at *D* until all the saturated solution has solidified. The mixture, which separates at this temperature, is termed a eutectic (or a cryohydrate, if one of the components is water) and its composition is given by *E*. Although the eutectic has a definite melting point, the following evidence suggests that it is an intimate mechanical mixture and not a compound.

(*a*) The components can be separated mechanically.
(*b*) The addition of each component raises the melting point of the eutectic. The melting point of a compound would be lowered on admixture with another substance.
(*c*) A heterogeneous structure can be seen under a microscope.
(*d*) X-ray analysis reveals the existence of two phases.

As shown in Fig. 3.10, the areas in the phase diagram each correspond to the existence of various phases or mixtures of phases.

The phase diagram for the water-KCl system may be used to explain the principle of freezing mixtures prepared from ice and salt. If salt is added to ice and a little water, some of the salt will dissolve in the water to produce a system comprised of ice, salt, and solution. Such a system is in stable equilibrium at the eutectic point only. The system will therefore tend to move towards this point and ice will melt and salt will continue to dissolve in the resultant water. Both of these processes are accompanied by absorption of heat and the temperature therefore falls until one of the solid components has been used up completely. If the initial proportions of ice and salt are chosen satisfactorily the eutectic temperature will be reached.

It has been suggested that eutectic mixtures may be useful as a means of increasing the rates of solution of slowly soluble drugs in aqueous body fluids (Goldberg, Gibaldi and Kanig, 1965). It was thought that the rapid solution of the second component (e.g. urea) in eutectic mixtures with these drugs would present the drug in a very fine crystalline form that would be more rapidly soluble than the usual forms of the drug. However, subsequent studies have suggested that the increased dissolution rates of the drugs in the presence of urea is likely to be caused by the formation of solid solutions of these drugs with urea and not by eutectic formation (Goldberg and co-workers, 1966c).

BIBLIOGRAPHY

BARROW, G. M. (1966) *Physical Chemistry*. 2nd ed. McGraw-Hill, London.

GUCKER, F. T. and SEIFERT, R. L. (1967) *Physical Chemistry*. English Universities Press, London.

MARON, S. H. and PRUTTON, C. F. (1958) *Principles of Physical Chemistry*. 3rd ed. Macmillan, New York.

MARTIN, A. N., SWARBRICK, J. and CAMMARATA, A. (1969) *Physical Pharmacy*. 2nd ed. Henry Kimpton, London.

ROSE, J. (1961) *Dynamic Physical Chemistry*. Pitman, London.

WILLIAMS, V. R. and WILLIAMS, H. B. (1967) *Basic Physical Chemistry for the Life Sciences*. Freeman and Co., San Francisco.

REFERENCES

AGUIAR, A. J., KRC JR., J., KINKEL, A. W., and SAMYN, J. C. (1967) Effect of polymorphism on the absorption of chloramphenicol palmitate. *J. pharm. Sci.*, **56**, 847–853.

GOLDBERG, A. H., GIBALDI, M., and KANIG, J. L. (1965) Increasing dissolution rates and gastrointestinal absorption of drugs via solid solutions and eutectic mixtures. I. Theoretical considerations and discussion of the literature. *J. pharm. Sci.*, **54**, 1145–1148.

GOLDBERG, A. H., GIBALDI, M., and KANIG, J. L. (1966a) II. Experimental evaluation of a eutectic mixture: Urea—acetaminophen system. *J. pharm. Sci.*, **55**, 482–487.

GOLDBERG, A. H., GIBALDI, M., and KANIG, J. L. (1966b) III. Experimental evaluation of griseofulvin—succinic acid solid solution. *J. pharm. Sci.*, **55**, 487–492.

GOLDBERG, A. H., GIBALDI, M., KANIG, J. L., and MAYERSOHN, M. (1966c) IV. Chloramphenicol—urea system. *J. pharm. Sci.*, **55**, 581–583.

HALEBLIAN, J. and MCCRONE, W. (1969) Pharmaceutical applications of polymorphism. *J. pharm. Sci.*, **58**, 911–929.

HIGUCHI, W. I., BERNARDO, D. P., and MEHTA, S. C. (1967) Polymorphism and drug availability. II. Dissolution rate behaviour of the polymorphic forms of sulphathiazole and methylprednisolone. *J. pharm. Sci.*, **56**, 200–207.

HIGUCHI, W. I., LAU, P. K., HIGUCHI, T., and SHELL, J. W. (1963) Polymorphism and drug availability. Solubility relationships in the methylprednisolone system. *J. pharm. Sci.*, **52**, 150–153.

MULLEY, B. A. (1959) Phase equilibria of phenol—water mixtures and the formulation of liquefied phenol. *Drug Stand.*, **27**, 108–109.

RIEGELMAN, S. (1955) *American Pharmacy*, 4th ed., Chap. 18. (R. A. Lyman and J. B. Sprowls, Eds). Lippincott Co., Philadelphia.

4

Surface and Interfacial Phenomena

AN INTERFACE is a boundary between two phases, while a surface is strictly a boundary between a solid or liquid phase and a vacuum. However, the latter term is often used to describe the interfaces between a gas and a solid or a gas and a liquid. Surfaces and interfaces can be conveniently classified according to the types of phase they separate; i.e. solid–liquid, solid–solid, and liquid–liquid interfaces and solid-gas and liquid-gas surfaces.

The Role of Surface Molecules of Solids and Liquids in Interfacial Phenomena

The kinetic theory of matter indicates that appreciable forces of attraction exist between molecules in liquids and in solids.* However, the surface molecules of a solid or liquid differ in state from those molecules within the bulk of that phase. This difference will be appreciated from a consideration of Fig. 4.1. The points *A* and *B* represent molecules in the bulk and surface, respectively, of a solid or liquid phase at a given instant, while the circles represent the zones in which molecular attractions between *A* and *B* and their surrounding molecules are significant. Since the zone around *A* contains, on average in any given direction, equal numbers of molecules all exerting similar forces, then the resultant attraction experienced by *A* will be zero. In a liquid surface, part of the sphere of molecular attraction around *B*, i.e. that part that projects above the surface, is occupied by a relatively small number of gas or vapour molecules. In addition to their low numbers, the attractive forces exerted by these molecules are relatively weak. Consequently, the attractions that exist between *B* and the underlying bulk molecules are not counterbalanced. The surface molecules, therefore, experience a resultant inward attraction and the transference of molecules from the bulk to the surface layer involves the expenditure of energy in order to overcome this inward attraction. It is known from the First Law of Thermodynamics that energy cannot be created or destroyed but can only be converted from one form to another. In this particular example the expended energy is converted into the so-called surface free energy of the system.

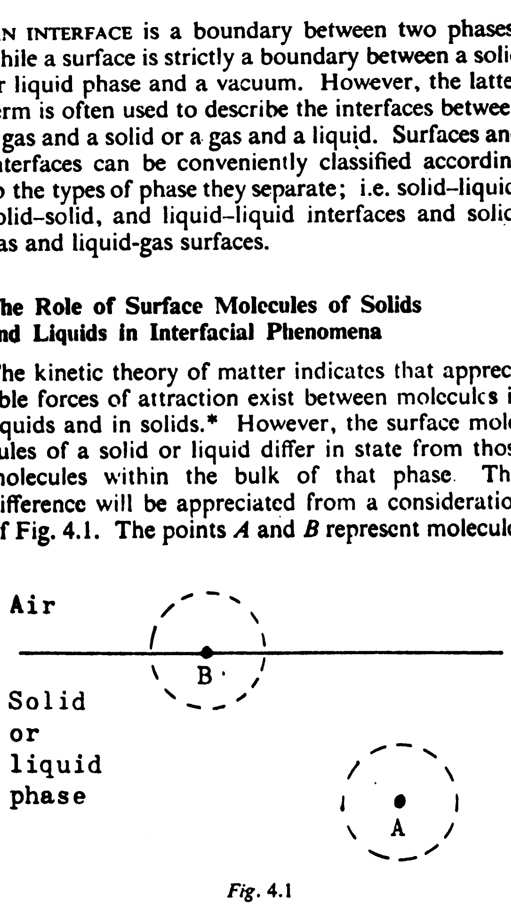


Fig. 4.1

In solids it is easier to explain the existence of surface free energy in terms of the unsatisfied forces that project from the surface than on the basis of an inward attraction exerted on the surface molecules. This is because the molecules in a solid are not mobile with respect to each other, as they are in a liquid. If we consider a solid suspended in a vacuum, it is readily appreciated that the projecting attractive forces associated with *B* are unsatisfied. The surface molecules therefore possess a greater amount of free energy than those in the bulk, and the transference of a molecule from the bulk to the surface of a solid again involves the expenditure of energy, some of which is converted to extra surface free energy.

If a solid block or a bulk liquid is subdivided into many small particles or drops by grinding or spraying, respectively, then the total surface area of the solid or liquid is increased tremendously; i.e. much greater proportions of the molecules are now in the surfaces of the solid or liquid and the total surface free energy of each of these systems is therefore increased. This extra energy is supplied by the work expended in the processes of grinding and spraying. In fact, surface free energy is defined

---

* It is assumed that the student is familiar with the basic concepts of the kinetic theory of matter: *see* Bibliography.

as the work required to extend a surface by unit area. The unit in the SI system is joule/metre^2.

Similar arguments may be applied to the solid–liquid, solid–solid, and liquid–liquid interfaces, and interfacial free energy is defined in a similar manner.

A consideration of thermodynamics also indicates that all systems will tend to react spontaneously in order to reach the state with the lowest total free energy. The existence of surface free energy and the methods by which systems react in order to decrease their total free energies to a minimum give rise to the various surface and interfacial phenomena that are discussed in the following sections of this chapter.

Surface and Interfacial Tension

It has been established above that surfaces possess free energy, and that systems will react spontaneously in order to reach a state with the lowest total free energy.

Since the total surface free energy is directly proportional to the surface area of a solid or a liquid, then one of the methods by which a system can reduce its surface free energy is by contracting the surface area of any solid or liquid within the system to a minimum. This is particularly obvious in liquids in which the molecules are mobile with respect to each other. This mobility is responsible for the tendency of liquid drops suspended in a gas or in an immiscible liquid to assume the minimum ratio of surface area to volume, i.e. a spherical shape, in the absence of other influences such as gravity. The tendency of mobile surfaces to contract can be readily explained by the fact that the surface molecules are continually moving into the bulk at a faster rate than other molecules can move outwards to replace them because of the resultant inward attraction experienced by surface molecules. The number of molecules in a newly formed surface therefore decreases until the surface area reaches a minimum; i.e. until the surface free energy is reduced as much as possible by this means.

Since the liquid surface shows this tendency to contract, it is usual to consider a surface tension, which is defined as the contractile force in newtons, operating in the plane of the surface normal to a line one metre in length. Surface tension is, in fact, the mathematical equivalent of surface free energy and has the same dimensions and magnitude; i.e. surface free energy (joule/metre^2) is equal to surface tension (newton/metre). Although surface free energy is the fundamental property, surface tension is more widely used in liquids since it is usually more convenient to consider problems in terms of a directional surface tension rather than the energy expended per unit increase in surface area.

Pressures Across Curved Liquid Menisci

One of the consequences of surface tension is that an excess pressure (p) is exerted on the concave side of a curved liquid meniscus compared with that exerted on the convex side. Consider a spherical bubble of radius r m divided by an imaginary line into two hemispheres. Surface forces will act around the circumference of the bubble and tend to pull the two hemispheres together. For the bubble to exist, this inward attractive force, which is equal to $2\pi r\gamma$ newtons (where γ N/m is the surface tension of the liquid) must be counteracted by a force that tends to push the hemispheres apart; i.e. the pressure on the inside concave surface of the bubble must be greater than that on the outer convex surface. If this excess pressure is denoted by p newtons/metre^2, then the total force pushing the hemispheres apart is given by $\pi r^2 p$ newtons since the area over which the pressure is exerted is that which separates the two hemispheres. Thus, at equilibrium—

$$\pi r^2 p = 2\pi r\gamma \quad \text{newtons}$$

and, therefore, by rearrangement—

$$p = \frac{2\gamma}{r} \quad \text{newtons/metre}^2 \qquad (4.1)$$

Equation (4.1) applies to any curved liquid meniscus; e.g. to a liquid film around a bubble or to the meniscus of a bulk liquid. Often the meniscus is not part of a sphere; then Eqn (4.2) may be used, where r_1 and r_2 are the principal radii of curvature of the surface.

$$p = \left(\frac{1}{r_1} + \frac{1}{r_2}\right)\gamma \quad \text{newtons/metre}^2 \qquad (4.2)$$

The Shape of Liquid Surfaces and the Wetting of Solids by Liquids

Liquid drops in contact with solid surfaces can assume different shapes, as shown in Fig. 4.2.

The degree of wetting of a solid by a liquid can be conveniently expressed in terms of the contact angle (θ), since a decrease in the value of θ results in a greater contact between the solid and the liquid. The particular value of θ depends on the relative values of the tensions associated with the liquid–solid interface (γ^{LS}) and the gas–solid and gas–liquid surfaces (γ^{GS} and γ^{GL}, respectively) as shown

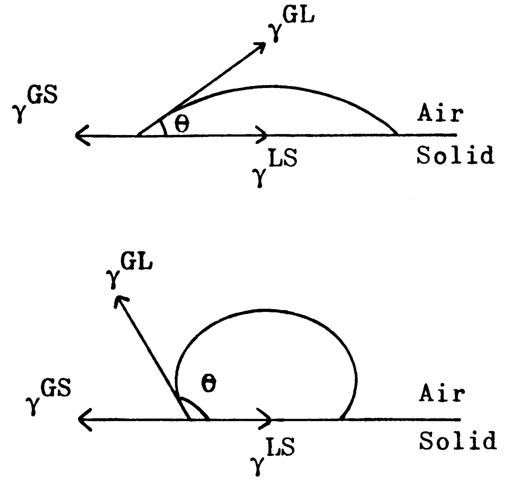


Fig. 4.2 Examples of the shapes of liquid drops in contact with solids

by Eqn (4.3).

$$\gamma^{GS} = \gamma^{LS} + \gamma^{GL} \cos \theta \qquad (4.3)$$

This equation is obtained by using the cosine rule to equate the opposing horizontal components of the tensions shown in Fig. 4.2; i.e. the component acting towards the left-hand side of a horizontal line drawn in the plane of the paper is $\gamma^{GS} \cos 0\ (= \gamma^{GS})$ and the components acting towards the right-hand side are $\gamma^{LS} \cos 0 + \gamma^{GL} \cos \theta\ (= \gamma^{LS} + \gamma^{GL} \cos \theta)$. From Eqn (4.3) it follows that if γ^{GS} is greater than γ^{LS} then $\cos \theta$ must be positive, i.e. $\theta < \pi/2$ radians, and conversely, if γ^{GS} is less than γ^{LS} then θ must lie between $\pi/2$ and π radians.

Marked variation between the values of contact angles in different systems can be observed from visual examination of the shapes of liquid menisci in glass tubes. For example, in the case of liquids, such as water, that wet glass the meniscus is concave upwards (i.e. θ is very small) whereas liquids, such as mercury, that do not wet glass exhibit a meniscus that is concave downwards (i.e. θ is large).

The Mechanism of Capillary Rise

Consider a capillary tube placed vertically in a liquid that wets the walls of the tube and that is held in a large container so that the liquid surface in this container is flat. The initial state of the system may be represented by Fig. 4.3(a). However, Eqn (4.1) indicates that there is an excess pressure on the concave side of the liquid meniscus in the capillary tube. This means that the pressure, p_A, at a point A immediately below the meniscus is less than the pressure, p_B, at a point B immediately above the meniscus by an amount equal to $2\gamma/r$; i.e.

$$p_B - p_A = \frac{2\gamma}{r} \text{ newtons/metre}^2 \qquad (4.4)$$

This difference in pressure will be appreciable since r, the radius of curvature of the meniscus inside the tube, is small. The difference in pressure, p_C and p_D, at points C and D, which are above and below the meniscus in the large container, respectively, is equal to $2\gamma/r_0$. However, in this case the pressure difference is negligible since the radius of curvature, r_0, at this point is so large, i.e.

$$p_C = p_D \text{ newtons/metre}^2 \qquad (4.5)$$

In addition, the pressures at B and C are both equal to the atmospheric pressure,

$$\therefore\ p_B = p_C \text{ newtons/metre}^2 \qquad (4.6)$$

Thus, from Eqns (4.5) and (4.6)

$$p_B = p_D \text{ newtons/metre}^2 \qquad (4.7)$$

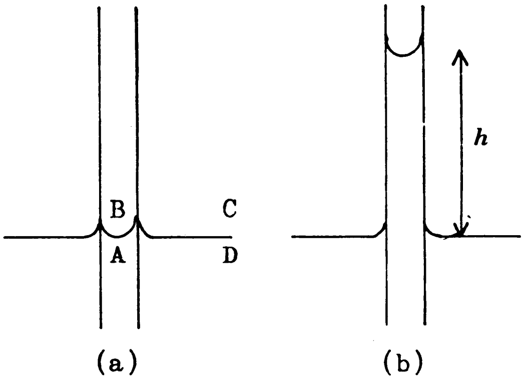
and from Eqns (4.4) and (4.7)

$$p_D - p_A = 2\gamma/r \text{ newtons/metre}^2 \qquad (4.8)$$

The difference in pressures within the liquid, as indicated by Eqn (4.8), produces a driving force which causes the liquid to rise up the capillary tube. Equilibrium is reached when the hydrostatic pressure of the column of liquid is equal to $2\gamma/r$. If at equilibrium the height of the column is h m, as shown in Fig. 4.3(b) then

$$2\gamma/r = h\rho g \text{ newtons/metre}^2 \qquad (4.9)$$

where ρ is the density of the liquid in kg/m^3 and g is the gravitational constant (9·807 m/s^2).



(a) (b)

Fig. 4.3 Stages in the rise of a liquid up a capillary tube

Measurement of Surface Tension

Three of the most common methods of measuring surface tension are described below, together with the relevant correction factors and practical precautions that must be taken into account to obtain accurate results. Details of other methods are given in the references cited in the Bibliography.

1. THE CAPILLARY RISE METHOD

This is one of the most accurate methods and is based on the measurement of the height to which a liquid rises in a capillary tube. Equation (4.10), which is a rearranged form of Eqn (4.9) can be used for approximate determinations.

$$\gamma = \frac{h\rho g r}{2} \quad \text{newtons/metre} \qquad (4.10)$$

For more accurate results certain modifications of this equation should be taken into account to allow for—

(*a*) a change in the value of the contact angle (θ) between the liquid and the walls of the tube from the zero value assumed in Eqn (4.10) to a finite value,

(*b*) the hydrostatic pressure exerted by the small volume of liquid contained in the meniscus above the point from which the height h is usually measured, and

(*c*) the difference in the density of the liquid (ρ^{L}) inside the capillary tube and that of the vapour (ρ^{V}) outside the tube.

The corrected equation is—

$$\gamma = \frac{\left(h + \frac{r}{3}\right)(\rho^{L} - \rho^{V})gr}{2\cos\theta} \quad \text{newtons/metre} \qquad (4.11)$$

Experimental Precautions

(*a*) The outer vessel must have a large diameter compared with the capillary, otherwise $p_C \neq p_D$.

(*b*) The capillary tube should be uniform in diameter and it should be circular in cross-section. If this is not so, then corrections must be made for ellipticity.

(*c*) To obtain the lowest value for the contact angle (θ) it is better to allow the meniscus to fall to rest than to rise up the tube to an equilibrium position. In addition, it is useful if the capillary walls are thoroughly wetted by the liquid before beginning the measurement. This can be done by passing liquid vapour through the cooled capillary, or by evacuating the capillary just before filling.

(*d*) The difference in the height of the menisci should be measured with a cathetometer, preferably with two telescopes, and care is necessary in lighting and viewing the meniscus in the large vessel.

(*e*) Adequate temperature control is necessary.

2. THE DROP WEIGHT METHOD

The mass, m kg, of an ideal drop of liquid having a surface tension γ newton/metre, falling from a tube of external radius r m, is given by Tate's equation (1864), Eqn (4.12),

$$mg = 2\pi r\gamma \quad \text{newtons} \qquad (4.12)$$

where g is the gravitational constant (9·807 m/s^2). In practice, the drop that falls from a tube is not ideal since the drop is constricted at the point where the break occurs to a diameter less than that of the tip of the tube. In addition, only part of the liquid protruding from the tip actually breaks away, and this falls in the form of more than one drop.

Harkins and Brown (1919) introduced a correction factor which takes the shape of the drop into account, and Eqn (4.13) then becomes

$$mg = 2\pi r\gamma \,.\, f\left(\frac{r}{V^{-3}}\right) \quad \text{newtons} \qquad (4.13)$$

The correction factor $f(r/V^{-3})$ is a function of the radius of the tip r and the cube root of the volume, V, of the drop. It is, in fact, the fraction of an ideal drop that actually falls, and Harkins and Brown give tables and graphs showing the value of their correction factor for various values of the actual ratio r/V^{-3}

An error is often introduced by attempting to obtain 'relative' values of surface tension by comparing drop weights of two different liquids. The results can be used in this way only if the surface tensions and densities of the two liquids and the tips of the tubes are chosen so that the values of r/V^{-3} are within the narrow limits where the correction curve is approximately horizontal.

Experimental Precautions

(*a*) The tip should be sharply ground with no imperfections in the outer circumference.

(*b*) The drops should be allowed to form slowly, especially in the later stages, otherwise the weight of the drop is greater than expected due to the effects of liquid streaming into the drop during the final, relatively slow stages of detachment.

(*c*) Adequate temperature control is necessary.

(*d*) It is normal practice to collect 20 to 30 drops in a determination, and calculate the average weight of one drop. Since the time of formation and collection of a single drop may take several minutes, it is necessary to prevent loss in weight of the collected sample brought about by evaporation.

The degree of attention paid to the above practical precautions has led to the development of various types of apparatus. These range from the simple type of stalagmometer shown in Fig. 4.4 to the more sophisticated apparatus used by Harkins and Brown (1919).

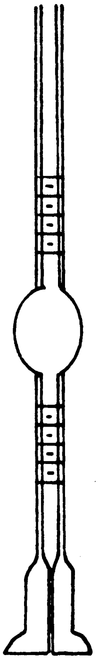


Fig. 4.4 Simple stalagmometer

3. DU NOUY TENSIOMETER

This instrument (Fig. 4.5) measures the force required to detach a platinum wire ring from the surface of a liquid. This force is applied by means of a torsion wire attached to a scale calibrated directly in units of surface tension. This calibration takes into account the fact that the surface tension forces act along the inner and outer circumferences of the wire ring but it is only suitable for approximate readings since it is based on the following Eqn (4.14),

$$P = 4\pi R\gamma \text{ newtons} \tag{4.14}$$

where, P is the maximum pull on the whole ring, R is the radius of the ring (from the centre of the ring to the centre of the wire) in metres, and γ is the surface tension of the liquid in newtons/metre.

For more accurate results it is necessary to multiply the scale reading observed on the tensiometer by a correction factor (F) which was introduced by Harkins and Jordan (1930) and depends on R^3/V and also on the ratio R/r, where V is the volume of liquid raised above the surface and r is the radius of the wire. Harkins and Jordan give tables of F for various values of R^3/V and R/r. The value of V may be calculated by initial determination of the scale deflections necessary to return the boom to its zero position when known weights are placed on the ring in air. The constant ratio of weight:scale reading may then be used to determine the weight of liquid raised by the ring for any scale reading. The corresponding volume can then be obtained from a knowledge of the density of the liquid.

Experimental Precautions

(*a*) The wire of the ring should be in the horizontal plane.
(*b*) The vessel containing the liquid should be large enough for the curvature of the surface not to affect the shape of the liquid drop raised by the ring.
(*c*) The surface of the liquid should be free from wave motion.
(*d*) There should be no motion of the ring except for a very slow upward movement.
(*e*) Temperature control should be adequate, and there should be no evaporation with consequent cooling of the surface.

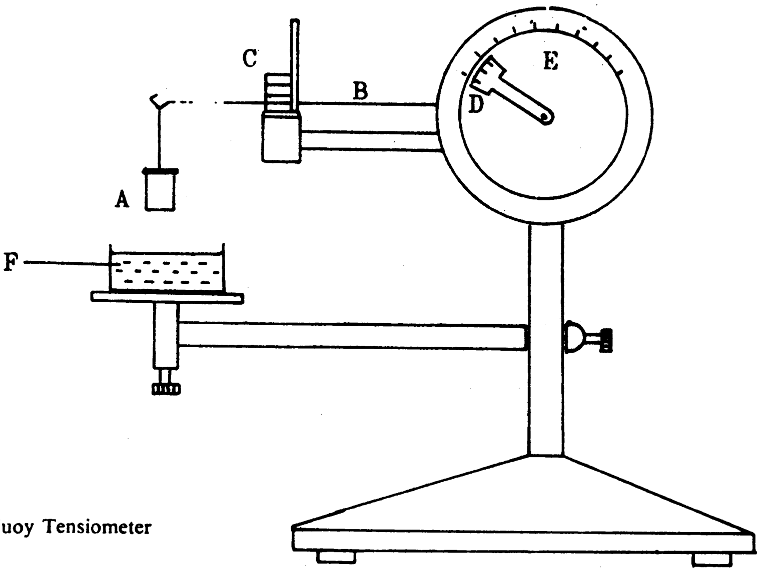
Effect of Temperature on Surface Tension

When a liquid in equilibrium with its vapour is heated, the interface between the two phases disappears at the critical temperature; i.e. there is no surface tension at this temperature. Thus, there is a general tendency for the surface tension of liquids to decrease as the temperature rises, provided no changes in molecular composition occur. This decrease arises from thermal expansion of the liquid in so far as this reduces the number of molecules that lie within the zone of molecular attraction (*see* Fig. 4.1).

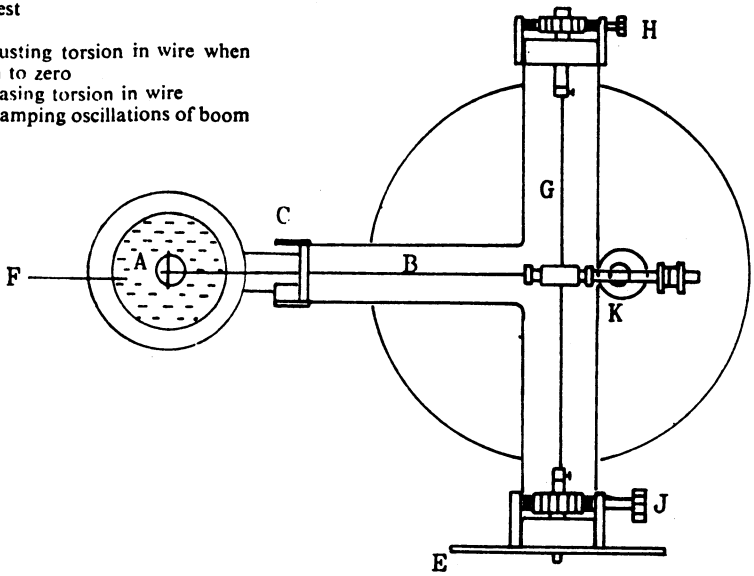
The relationship between temperature and surface tension is expressed by Eqn (4.15),

$$\gamma = \gamma_0\left(1 - \frac{T}{T_c}\right)^{1.2} \text{newtons/metre} \tag{4.15}$$

where γ is the surface tension at temperature T, γ_0 is the surface tension at thermodynamic zero, and



Side view



Plan view

Fig. 4.5 The Du Nuoy Tensiometer
A. Platinum loop
B. Light boom
C. Mirror for aiding zero adjustment of boom
D. Vernier
E. Scale calibrated in units of surface tension
F. Liquid under test
G. Torsion wire
H. Screw for adjusting torsion in wire when bringing boom to zero
J. Screw for increasing torsion in wire
K. Dash-pot for damping oscillations of boom

T_c is the critical temperature of the liquid; i.e. the temperature at which the interface between the liquid and its saturated vapour disappears.

Orientation of Molecules of Pure Liquids at Surfaces and Interfaces

The inward attraction experienced by the surface molecules of a liquid will tend to result in a particular orientation of these molecules with respect to the plane of the surface. This orientation arises from the difference in the strengths of attraction between the various groups in the surface molecules and the underlying molecules. Consequently, those groups with the strongest associated attractive forces will tend to be directed inwards, and those with the weakest associated fields will tend to project from the surface. This orientation offers a further means by which the total surface free energy of a system is reduced to a minimum; i.e. in addition to the tendency for the surface to assume a minimum area for a given volume.

The tendency towards a definite orientation will be disrupted by thermal effects so that the surface and underlying layers of a liquid long-chain alcohol or carboxylic acid, for example, may be represented by Fig. 4.6(*a*). The rod-like ends of the diagrammatic molecules represent groups with relatively low attractive forces (e.g. hydrocarbon chains) whereas the globular ends represent more polar groups (e.g. —OH and —COOH groups) with greater associated forces.

The type of orientation of the surface molecules of a given liquid is markedly influenced by the nature of the other phase with which it is in contact. For example, if either the alcohol or carboxylic acid shown in Fig. 4.6(*a*) is in contact with a more polar phase such as water or glass, which will exert a relatively strong attractive force, then the interfacial molecules of the organic liquid will tend to become orientated with their polar groups directed outwards as shown in Fig. 4.6(*b*).

Effects of Solutes on the Surface Tension of Liquids

The composition of a newly formed surface of a solution is the same as that of the bulk phase. However,

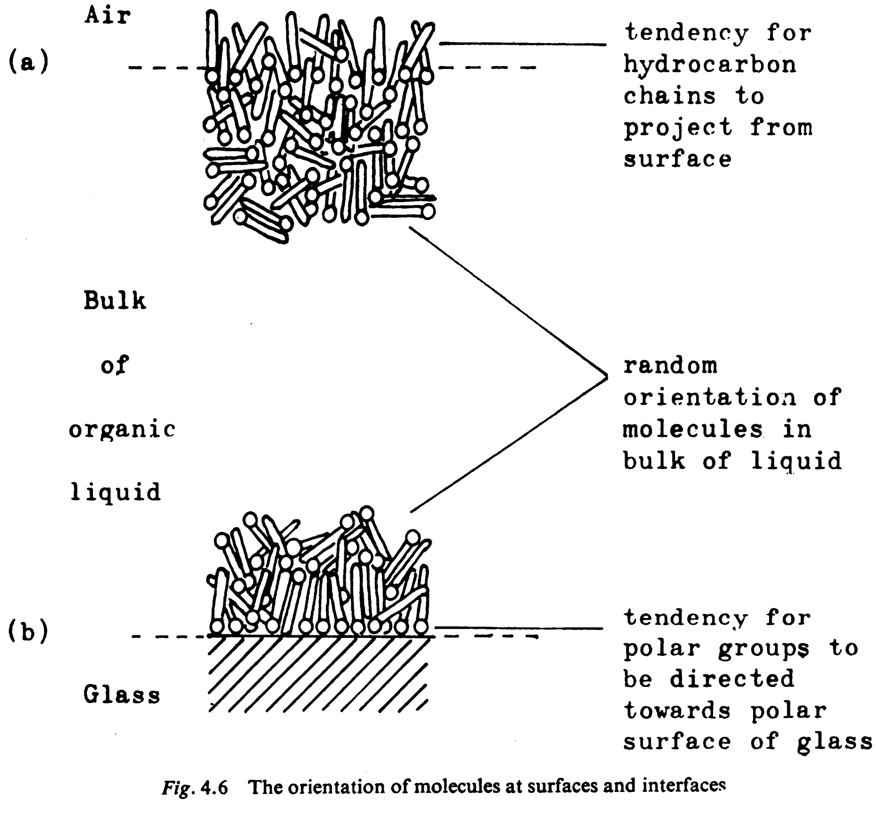


Fig. 4.6 The orientation of molecules at surfaces and interfaces the unequal forces that exist between the underlying molecules and the solute and solvent molecules in the fresh surface will eventually produce a surface layer with a different composition from that of the bulk solution, since those surface molecules that are attracted most strongly by the underlying liquid will be removed most rapidly from the surface and replaced by other molecules. The rate at which an equilibrium is established depends on the nature of the solute and solvent and may vary between values of the order of 0·1 s for small molecules or ions in aqueous solution to several days for some colloidal solutions.

The surface properties of the solution, including the surface tension, will depend on the composition of the surface layer and will therefore be affected by the distribution of the solute between this layer and the bulk solution. The quantitative relationship between solute distribution and surface tension is expressed by Gibbs adsorption isotherm Eqn (4.16)

$$\Gamma = \frac{-c}{RT} \cdot \frac{d\gamma}{dc} \tag{4.16}$$

where Γ is the surface excess concentration of the solute, which may be regarded as the concentration of the solute per unit area of the surface layer minus the solute concentration in the bulk of the solution, c is the overall solute concentration, R is the gas constant, T is the thermodynamic temperature, and the term $d\gamma/dc$ is a measure of the change in surface tension with solute concentration. The surface excess may be positive or negative depending on the distribution of the solute between the surface and bulk of the phase, thus giving rise to so-called positive and negative adsorption, respectively.

Equation (4.16), which is only applicable to dilute solutions where solute concentrations may be substituted for activities, has been verified experimentally. Such verification can be carried out only by methods in which sufficient time can be allowed for adsorption equilibrium to be attained. Such conditions did not apply in some early investigations. However, good agreement has been obtained between observed and calculated surface excess concentrations by McBain and Swain (1936) using a trough from which the surface layers were removed for analysis by a rapidly moving microtome blade, and by radio-tracer techniques (Dixon *et al.*, 1954).

Equation (4.16) indicates that solutes that tend to concentrate in the surface layers will cause a reduction in the surface tension of the solvent, i.e. $d\gamma/dc$ will be negative. Conversely, negatively adsorbed solutes will cause an increase in the surface tension.

In solutions containing negatively adsorbed solutes there is little change in the composition of the surface layer when compared with that of the pure solvent. Consequently, only small increases in surface tension are observed in these cases.

In solutions containing positively adsorbed solutes, the composition of the surface layer may change markedly from that of the pure solvent, depending on the nature of the solute and solvent and the concentration of the solution. For example, there may be a slight lowering of surface tension with increase in concentration as in the case of many mixtures of organic liquids, since the surface tensions of the majority of organic liquids are fairly similar ($<35 \times 10^{-3}$ newtons/metre). However, the greater difference between the surface tension of water (about 72×10^{-3} newtons/metre) and organic liquids leads to a considerable decrease in the surface tension of aqueous solutions of organic compounds even at low concentrations.

Surface Active Agents

Solutes that cause a marked decrease in the surface tension of the solvent are termed surface active agents or surfactants. These substances are of importance in a wide variety of fields as emulsifying agents, detergents, solubilising agents, wetting agents, foaming agents, antifoaming agents, flocculants and deflocculants.

All types of soluble surface active agent contain—

(*a*) a lipophilic (or hydrophobic) group; i.e. a group, such as a long hydrocarbon chain, that has little affinity for aqueous solvents, and

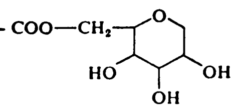
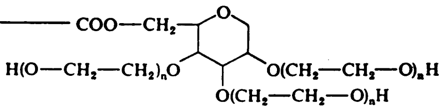
(*b*) a hydrophilic (or lipophobic) group, i.e. a group that has an affinity for water. To have such an affinity the group must possess an appreciable polar character, e.g. an ion or a group with a large permanent dipole. A molecule or ion that possesses this type of structure is termed an amphipathic molecule or ion.

A suitable balance between the opposing hydrophilic and lipophilic characteristics of the surface active agent is necessary to ensure that surface active properties are obtained. For example, substances would be too soluble in either water or oil and, therefore, of little use as surface active agents if there were an excessive predominance of hydrophilic or lipophilic properties, respectively.

The classification of water-soluble surface-active agents is based on the nature of the polar group, and various examples of the different classes are shown in Table 4.1.

The attractive forces between water molecules are appreciable. These attractions involve the formation of hydrogen bonds between adjacent molecules that

Table 4.1
The Classification of Synthetic Surface Active Agents

Class	*Surface active agent*	*Chemical formula* (in aqueous solution)		*Surface inactive ion*
		Lipophilic group	*Hydrophilic group*	
1. *Anionic*				
(*a*) Alkali soaps	Potassium stearate	$C_{17}H_{35}$	COO^-	K^+
(*b*) Organic sulphates	Sodium lauryl sulphate (sodium dodecyl sulphate)	$C_{12}H_{25}$	OSO_3^-	Na^+
(*c*) Organic sulphonates	Sodium cetyl sulphonate (sodium hexadecane sulphonate)	$C_{16}H_{33}$	SO_3^-	Na^+
2. *Cationic*				
(*a*) Quaternary ammonium compounds	Cetyl trimethyl ammonium bromide or cetrimide (hexadecyl trimethyl ammonium bromide)	$C_{16}H_{33}$	$\overset{+}{N}(CH_3)_3$	Br^-
(*b*) Pyridinium compounds	Dodecyl pyridinium chloride	$C_{12}H_{25}$	$\overset{+}{N}C_5H_5$	Cl^-
3. *Ampholytic* Amino-acids	N-dodecyl -alanine	$C_{12}H_{25}$ $C_{12}H_{25}$ $C_{12}H_{25}$	(i) in alkaline solution (e.g. NaOH)—anionic $NH-CH_2-CH_2-COO^-$ (ii) in acid solution (e.g. HCl)—cationic $\overset{+}{N}H_2-CH_2-CH_2-COOH$ (iii) at iso-electric point—zwitterion $\overset{+}{N}H_2-CH_2-CH_2-COO^-$	Na^+ Cl^- none
4. *Non-ionic*				
(*a*) Alcohol-polyethylene glycol ethers	Polyethylene glycol 1000 monocetyl ether (cetomacrogol 1000)	$CH_3-(CH_2)_n$ $n = 15$ or 17	$(O-CH_2-CH_2)_m-OH$ $m = 20-24$	none
(*b*) Fatty acid-polyethylene glycol esters	Polyethylene glycol 40 monostearate	$CH_3-(CH_2)_{16}$	$CO-(O-CH_2-CH_2)_{40}-OH$	none
(*c*) Fatty acid-polyhydric alcohol esters	Sorbitan mono-oleate	$C_{17}H_{33}$		none
	Polyoxyethylene sorbitan mono-oleate	$C_{17}H_{33}$		none

A more detailed classification of surface active agents is given by Schwartz and Perry (1949) and Moilliet *et al.* (1961). In addition, Lower (1962) gives a list of the chemical structures and trade names of surface active agents produced in the UK in 1961.

tend to become arranged in a particular manner. Water may therefore be said to possess a relatively high degree of structure. The presence of a solute will cause the replacement of some intermolecular water bonds by attractive forces between the solute and water. If these latter forces are appreciable, then the solute will remain in solution. However, in a surface active agent, although the forces of attraction between the hydrophilic groups and water are appreciable, those between the hydrocarbon chains and water are much weaker than the hydrogen bonds between water molecules. The system therefore tends to react in order to limit the degree of contact between water and the hydrophobic groups by the following methods—

(*a*) The hydrocarbon chains of molecules or ions in true solution tend to assume a coiled configuration, thus reducing the area of contact that would exist if these chains were extended.

(*b*) Although some molecules of the surface active agent are in true solution the majority are adsorbed at the surface, the hydrophilic groups remaining in contact with water, while the hydrocarbon chains tend to project from the surface. This type of adsorption is, of course, responsible for the marked effect of surface active agents on the surface properties of water.

Traube's rule (1891) states that in a homologous series the concentration required to produce an equal lowering of surface tension decreases by a factor of three for every CH_2 group added to the hydrocarbon chain. Thus, the differences in the free energies of adsorption of a homologous series of surface active agents are approximately equal. This fact has been interpreted as evidence for suggesting that the adsorbed molecules lie with their hydrocarbon chains parallel to the surface since in this orientation each CH_2 group has a similar position with respect to the surface. However, this orientation has been criticised, and alternative suggestions involving either a coiled hydrocarbon chain in the adsorbed molecules (Ward, 1946), or an orientation in which the hydrocarbon chains project from the surface and show a random hindered rotation around the C—C bonds (Aronow and Witten, 1958), have been made, together with explanations of the significance of the factor of three observed by Traube.

In addition to their marked effect on the surface properties of the solvent, surface active agents also show unusual changes in their physical properties at reasonably well-defined concentrations. These changes are attributed to the association of the amphipathic molecules or ions into aggregates of colloidal dimensions, which are known as micelles. The formation of these aggregates and the properties of the colloidal solutions are dealt with in more detail in Chapter 5.

ADSORPTION AT SOLID–GAS AND SOLID–LIQUID INTERFACES

The irregular shapes of solid particles indicate the absence of a surface tension similar to that operating in liquids. In addition, it is impossible to extend the surface of a solid without upsetting the lattice structure of the component crystals. The changes in energy that result when the crystal lattices are altered make the estimation of changes in true surface free energy difficult and are also responsible for the difference in the ways in which the surface energies of solids and liquids are manifested. In general, the phenomena that do not involve the disruption of crystal lattices, e.g. adsorption and wetting, are thought to depend solely on the existence of surface free energy in solids.

If a solid comes into contact with a gas or a liquid there is an accumulation of gas or liquid molecules at the interface; i.e. the densities of the gas or liquid at the interface are greater than their bulk densities. This phenomenon is known as adsorption. (Since adsorption is a surface phenomenon it should be distinguished from absorption, which involves penetration of the absorbing material by the molecules of the absorbed substance.) The occurrence of adsorption at the solid–gas interface is easy to perceive since it causes a decrease in the pressure of the gas. However, the effects of adsorption at the solid–liquid interface are less obvious because the concentration of molecules in the bulk of a liquid is relatively great so that the detection of this type of adsorption is more difficult. In adsorption at the interface between a solid and liquid that contains two components (e.g. a solution or a mixture of miscible liquids), then the effects of adsorption are again readily demonstrable because the different degrees of adsorption of the two components result in a concentration change in the bulk of the solution.

It has already been pointed out that adsorption is one of the phenomena that occur at interfaces as a result of the tendency of a system to attain the lowest state of free energy. The surface (or interfacial) free energy that arises from the residual attractive forces on the surface molecules is reduced by the interaction of these forces with those of the adsorbed molecules. The irregular shapes of solid particles and the roughness of solid surfaces introduces a further factor which complicates any quantitative treatment of phenomena that occur at solid surfaces. This factor is the variation in the activity of different sites on the solid surface; e.g. the activity of a site on a plain surface will be different from that of a similar site at the bottom of a crack in the surface. Thus, the attractive forces exerted by the solid surface will be heterogenous. the adsorbed molecules will first be attracted to those sites with the highest activity, and subsequent adsorption will occur at the sites of decreasing activities.

In the majority of its practical applications the process of adsorption involves, in fact, the displacement of previously adsorbed molecules by others. For example, the adsorption of gases often involves the displacement of air by another gas, and adsorption from solution often involves the displacement of solvent molecules by a more strongly adsorbed solute. However, a system that contains two components only, i.e. the adsorbing material (the adsorbent) and the substance that is adsorbed (the adsorbate), is more simple to consider from a theoretical point of view. For this reason the adsorption of gases by evacuated solids will first be discussed, followed by an account of adsorption at the solid–liquid interface.

Adsorption at the Solid–Gas Interface

The interactions between the solid adsorbent and the adsorbed gas or vapour molecules may range from a surface reaction, which is similar to a normal chemical reaction, to weak attractions similar to those responsible for the condensation of gases and vapours. It is therefore convenient to classify adsorption systems into chemisorption and physical adsorption on the basis of the interactions that are involved. The characteristic properties of these types of adsorption are listed in Table 4.2 in order to allow comparisons to be made.

Table 4.2
The Characteristics of Physical Adsorption and Chemisorption

	Physical adsorption	*Chemisorption*
1. Adsorption forces	Weak physical forces (van der Waals forces)* Heat of adsorption is usually <50 kJ/mole. May be regarded as a surface condensation	Involves transfer or sharing of electrons between adsorbent and adsorbed molecules. Heat of adsorption is usually about 60 to 420 kJ/mole. May be regarded as a surface reaction.
2. Specificity	Non-specific; i.e. will occur to some degree in any system.	Specific; i.e. only occurs when reaction is possible between adsorbent and adsorbate.
3. Reversibility	Reversible; i.e. adsorbate can be removed easily from surface in an unchanged form.	Often irreversible; i.e. adsorbate is removed with difficulty usually in a changed form: e.g. oxygen adsorbed by carbon is removed as carbon dioxide.
4. Effect of temperature	Process is exothermic; i.e. amount of adsorption decreases with rise in temperature.†	Surface reaction only proceeds above a certain temperature.‡ Reaction is usually exothermic.¶
5. Number of adsorbed layers	Monomolecular layer formed at low pressures followed by additional layers as pressure increases. Condensation of vapour in capillaries of porous solids may occur.	Restricted to formation of a monolayer.
6. Rate of adsorption	Usually rapid at all temperatures.	Usually proceeds at a finite rate which increases rapidly with rise in temperature.

* The appendix gives information on the various forces involved in physical adsorption.

† The exothermic nature of the process is indicated by the thermodynamic equation $\Delta G = \Delta H - T\Delta S$, where ΔG, ΔH, and ΔS are the changes in free energy, heat content, and entropy, respectively, that occur during a physicochemical process. For a spontaneous process such as adsorption ΔG must be negative; i.e. the process must lead to a decrease in the free energy of the system. In addition, the adsorbed molecules will have less freedom than those in the gaseous state; i.e. there is a decrease in the disorderly orientations of the gas molecules on adsorption and ΔS will be negative. Consequently, the term $-T\Delta S$, where T is the thermodynamic temperature, will be positive, and, therefore, from the above equation, H must be negative and the process is exothermic.

‡ Physical adsorption at low temperatures may give way to chemisorption at higher temperatures.

¶ It is often stated that all adsorptions are exothermic. However, some chemisorptions may be endothermic (Thomas, 1961).

ADSORPTION ISOTHERMS

It is normal practice to express the results of an adsorption process in the form of a graph showing the variation in the amount of gas or vapour adsorbed, x, by a given mass, m, of the adsorbent, with the equilibrium pressure, p, of the gas at constant temperature. These graphs are known as adsorption isotherms. In vapour adsorption it is more common to use the relative vapour pressure, p/p_0, where p_0 is the saturated vapour pressure, instead of simply the equilibrium pressure.

The majority of isotherms for the adsorption of gases or vapours by solids may be classified on the basis of their shapes into Types I–V as shown by Fig. 4.7.

Many attempts have been made to obtain a relationship between the amount of adsorption and the pressure of gas in the system. Since these relationships are quantitative expressions of the shapes of graphs given in Fig. 4.7, the equations themselves are also referred to as adsorption isotherms. The most common are those suggested by Freundlich, Langmuir, and by Brunauer, Emmett, and Teller.

The Freundlich adsorption isotherm, which is expressed by Eqn (4.17), was the earliest of these adsorption isotherms and it was deduced empirically from its agreement with a large number of experimental results.

$$\frac{x}{m} = kp^{1/n} \tag{4.17}$$

In Eqn (4.17), x and m are defined as before, and n and k are constants for a particular system at constant temperature. The power of p is usually written in the form of a fraction ($1/n$, where $n > 1$) to indicate that the increase in the amount of adsorption at the low pressures, where this equation is usually valid, is less rapid than the increase in the pressure itself. In fact, the value of $1/n$ is often between 0·3 and 0·5.

The agreement of experimental data with Eqn (4.17) may be checked by plotting $\log x/m$ against $\log p$. A straight line with a slope of $1/n$ and an intercept of $\log k$ should be obtained if the equation is applicable.

Langmuir (1916) attempted to explain adsorption in terms of a dynamic equilibrium determined by

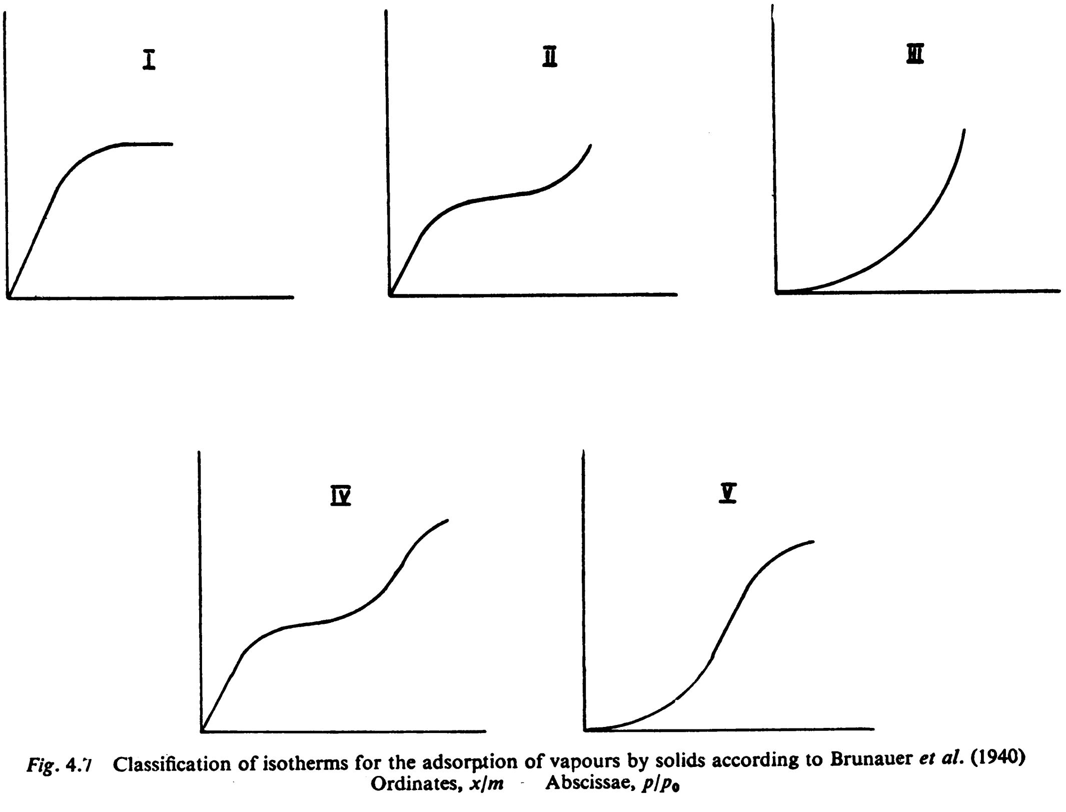


Fig. 4.7 Classification of isotherms for the adsorption of vapours by solids according to Brunauer *et al.* (1940)
Ordinates, x/m - Abscissae, p/p_0

the rates at which gas molecules were adsorbed and desorbed. The rate of adsorption at any instant was considered to be proportional to the amount of 'free' surface available at that instant and to the pressure of the gas in the system. The rate of desorption was considered to be proportional to the fraction of the surface onto which the gas was already adsorbed. In addition, it was assumed that the interaction between adsorbed molecules and the adsorbent involved forces similar to short-ranged chemical bonds, so that the adsorption of only a monolayer was envisaged.

This approach resulted in the Langmuir adsorption isotherm which is expressed by Eqn (4.18), where x, m, and p are defined as before, and a and b are constants for a given system at constant temperature.

$$\frac{x}{m} = \frac{abp}{1 + ap} \tag{4.18}$$

In the original derivation of Eqn (4.18), the constant a is the ratio of the rate constants for the adsorption and desorption processes, and the constant b is the amount of gas required to produce an adsorbed monolayer over the whole surface of the adsorbent. However, the Langmuir adsorption isotherm is often successfully applied to systems where the assumptions made by Langmuir are obviously invalid. In these cases, the constants a and b can no longer be defined as above and are arbitrary constants only.

The applicability of the Langmuir adsorption isotherm to experimental data may be checked by rearranging Eqn (4.18) to the form shown in Eqn (4.19).

$$p \cdot \frac{m}{x} = \frac{1}{b} \cdot p + \frac{1}{ab} \tag{4.19}$$

It can be seen from Eqn (4.19) that a graph of pm/x against p should give a straight line with a slope of $1/b$ and an intercept of $1/ab$ if the equation is obeyed.

The Langmuir adsorption isotherm is capable of application only to isotherms represented by Type I in Fig. 4.7. Many attempts have been made to derive an equation that is applicable to all the different types of adsorption isotherm. One of the more successful approaches is due to Brunauer *et al.* (1938). These workers considered that multilayers of gas were adsorbed and assumed that the forces involved in physical adsorption are the same as those responsible for the condensation of gases. Since these forces have a very short range the effect of the solid was considered to be appreciable only on the first layer of adsorbed molecules. The possibility of multilayer formation depended, therefore, on the degree of interaction between the gas molecules. Equation (4.20), which is the result of this approach, is known as the BET adsorption isotherm.

$$\frac{p}{x(p_0 - p)} = \frac{1}{x_m h} + \frac{h - 1}{x_m h} \cdot \frac{p}{p_0} \tag{4.20}$$

where x = the amount of gas adsorbed per unit weight of adsorbent at the equilibrium pressure p, x_m = the amount of gas adsorbed per unit weight of adsorbent when the surface of the latter is covered with a monolayer of adsorbed molecules, p_0 = the saturated vapour pressure of the gas, and h = a constant, which is a function of the difference between the heat of adsorption of the gas in the first layer (E_a) and the heat of condensation of the gas (E_c) as shown by Eqn (4.21).

$$h = e^{-E_a - E_c/RT} \tag{4.21}$$

The agreement of experimental data with Eqn (4.20) may be verified by plotting $p/x(p_0 - p)$ against p/p_0. A straight line with a slope of $h - 1/x_m h$ and an intercept of $1/x_m h$ is obtained if the equation is applicable.

The derivation of Eqn (4.20) neglects the existence of lateral interactions between molecules adsorbed in any layer and assumes that the energy of adsorption is independent of the amount already adsorbed in that layer. In addition, it is assumed that the energy of interaction between two adjacent layers of adsorbed molecules is the same, irrespective of their positions relative to the solid surface.

Since Eqn (4.20) reduces to the Langmuir equation for systems in which adsorption is restricted to monolayer formation, it is therefore applicable to isotherms of Type I in Fig. 4.7. In addition, the BET equation can be applied to isotherms of Type II, in which the further rise after the first plateau is attributed to the adsorption of layers beyond the monolayer level. However, the agreement of Eqn (4.20) with the experimental results for this type of isotherm is satisfactory only when the relative pressure (p/p_0) is within the range of 0·05 to 0·35.

It has been suggested that the failure of Eqn (4.20) below a relative pressure of 0·05 arises from the non-uniformity of the surface sites, which results in a variation in the value of E_a for different parts of the surface, and which cannot be accounted for satisfactorily in a quantitative manner. Furthermore, the poor agreement above a relative pressure of 0·35 has been ascribed to the effect of the walls of narrow pores in the solid on the thickness of the adsorbed films. This effect can be taken into account by rather complicated modifications of Eqn (4.20) (Gregg, 1965).

Except for Type IV isotherms at low pressures, where they resemble Type II isotherms, the BET equation appears to be of little use in describing quantitatively the isotherms of Types III to V. However, modifications of the equation are available which allow a reasonable agreement with a Type IV isotherm (Gregg, 1965).

THE CAPILLARY CONDENSATION THEORY

In adsorption of vapours by porous solids, it is considered that much of the adsorption at high relative pressures is caused by condensation of the vapour in the fine capillaries of the solid. The occurrence of condensation at pressures lower than normal saturated vapour pressures follows from a consideration of the Kelvin equation, which indicates the relationship between the equilibrium vapour pressure, p, in a capillary of radius r and the normal saturated vapour pressure, p_0. This relationship is represented by Eqn (4.22) in which γ is the surface tension of the liquid, V is its molecular volume, R is the gas constant, and T is the thermodynamic temperature.

$$\ln \frac{p}{p_0} = \frac{-2\gamma V}{rRT} \tag{4.22}$$

Equation (4.22) indicates that condensation of a vapour will occur at pressure p in all those capillaries with a radius less than r in a porous solid. This theory is supported by the fact that at pressures equal to or just below the saturated vapour pressure the amounts of adsorbed vapours for different substances calculated as liquid volumes are the same for a given adsorbent (Gurvich, 1915). Thus, the efficiency of a porous solid in adsorbing vapour at high relative pressures is determined mainly by the total volume of the pores that are accessible to the vapour.

Adsorption at the Solid–Liquid Interface

The adsorption of solutes from solution at the solid–liquid interface has been studied to a lesser extent than gas adsorption, although the phenomenon has been used practically for centuries; e.g. in the clarification of liquids. It is only in more recent years that the theories of adsorption from solution have become more developed, and the importance of the solvent recognised.

The isotherms for adsorption from solution have been classified into four main groups depending on the shape of the initial part of the isotherm (Giles *et al.*, 1960). Each of these groups is then subclassified on the basis of the shape of the isotherm at higher concentrations. The classification is shown in Fig. 4.8.

The initial curvature of the *S* curves indicates that adsorption becomes easier as the equilibrium concentration rises. This curve is characteristic of systems in which the solute molecule is monofunctional, i.e. the solute molecule has a single point of strong attachment on to the adsorbent, and there is an appreciable intermolecular interaction within the adsorbed layer. These conditions usually result in the vertical orientation of adsorbed molecules. The *S* curve may also be indicative of strong competition by the solvent or other species for the adsorbent sites.

The *L* curve is so-called because it is similar in shape to the Langmuir isotherm, a type of curve most commonly obtained for adsorption from dilute solution. The shape of the curve shows that the amount of adsorbed solute increases as the concentration increases until a limiting 'plateau' value is reached. In systems producing this type of curve the adsorbate molecules are likely to be adsorbed flat on the surface of the adsorbent, or vertically if there is little competition from the solvent for the adsorption sites.

The *H* or High Affinity curve is characteristic of systems in which there is a strong interaction between the adsorbent and adsorbate, which results in complete removal of solute from solutions of low concentration. Such a system may involve a chemisorption mechanism or the adsorption of micellar aggregates. (*See* Chapter 5 for information about structure of micelles.)

The *C* curve represents a constant partition of the solute between the adsorbent and solution. It seems to occur in systems that involve fibrous materials such as wool, where the penetration of the solute increases as the concentration in solution increases.

From a consideration of the above comments it can be seen that the shape of an isotherm for a particular system may provide information about the interaction between the adsorbate and adsorbent, the orientation of the adsorbed molecules, and the competition offered by the solvent.

Although attempts have been made to find a universal equation applicable to all solution isotherms, a suitable one has not been developed due to the wide variety and complexity of systems.* Also, as pointed out by Kipling (1965), 'many equations have proved inadequate because the role of the solvent has been neglected'. Equations analogous to the Freundlich, Langmuir, and BET isotherms (i.e. Eqns 4.17, 4.18, and 4.20) have been used for adsorption from solution by introducing a

---

* The Gibbs Adsorption Isotherm applies to adsorption at all interfaces. However, it involves a further term (interfacial tension) and therefore does not describe adsorption solely in terms of the concentration of the solution.

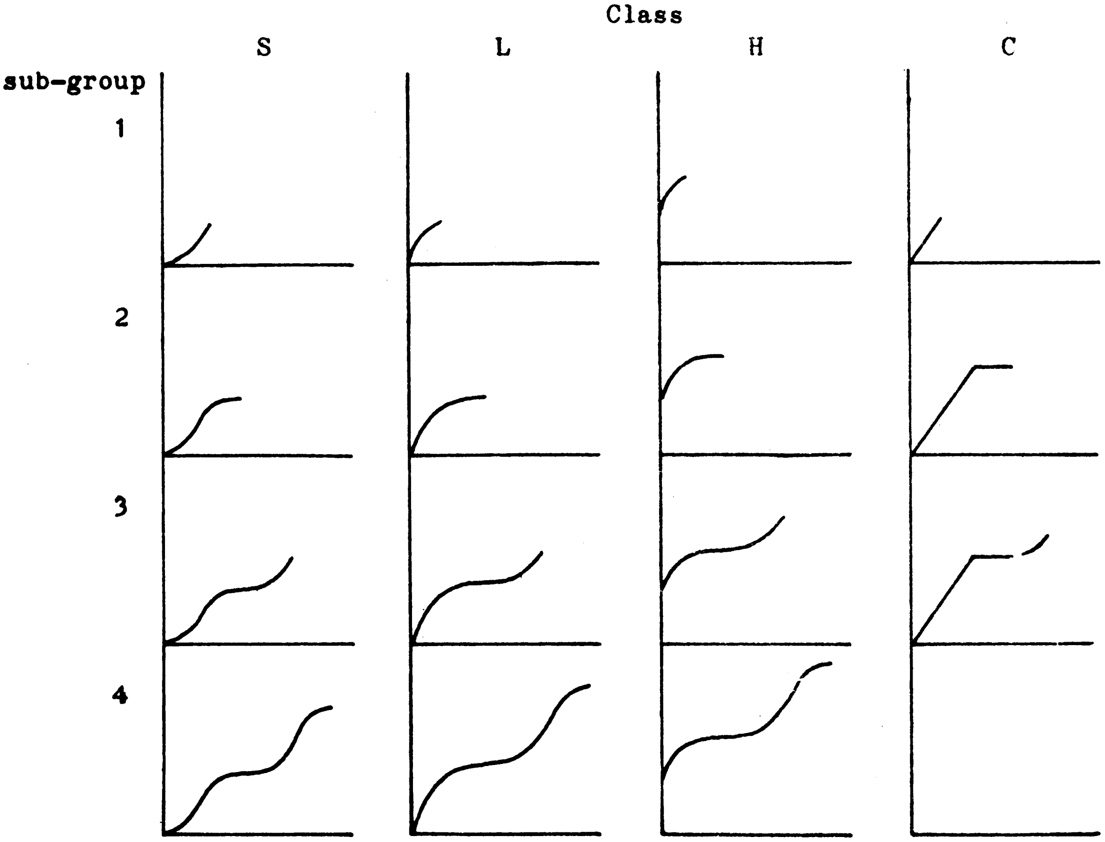


Fig. 4.8 Classification of isotherms for the adsorption of solutes from solution according to Giles and co-workers (1960) Ordinates—amount of solute adsorbed per unit weight of adsorbent Abscissae—equilibrium concentration of solution

term for solute concentration instead of pressure. For example, the Langmuir isotherm, which has probably found the greatest application to solution data, becomes

$$\frac{x}{m} = \frac{abc}{1 + ac} \qquad (4.23)$$

where, *c* is the equilibrium concentration of the solute, *x* is the amount of solute adsorbed by a given mass (*m*) of adsorbent, and *a* and *b* are constants for a given system at constant temperature. Although the use of these equations for adsorption from solution may not be entirely satisfactory from a theoretical point of view, they often fit the isotherm obtained for the adsorption of solutes from dilute solution (Kipling, 1965).

Kipling (1951) and Kipling and Tester (1952) have introduced the concept of a 'composite isotherm'. This takes into account the adsorption of all the components of a system (e.g. the solvent as well as the solute) by combining the individual isotherms for each component. However, in the case of dilute solutions the 'composite isotherm' is often indistinguishable from the usual solute isotherm and, therefore, there is little advantage to be gained from using it in place of the simpler Langmuir isotherm.

FACTORS AFFECTING ADSORPTION FROM SOLUTION

1. *Solute Concentration*

An increase in the concentration of the solute will cause an increase in the amount of adsorption that occurs at equilibrium until a limiting value is reached. It should be borne in mind that for most adsorptions from solution the *relative* amount of solute removed from solution is greater in dilute solutions.

2. *Temperature*

It has been pointed out previously that most adsorption processes are exothermic. Application of Le Chatelier's principle indicates that the amount of adsorption will decrease as the temperature rises.

3. *Surface Area of Adsorbent*

Since adsorption is a surface phenomenon, an increase in the surface area of the adsorbent, e.g. by reducing the particle size or increasing the porosity, will lead to an increase in the amount of adsorption.

4. *Removal of Adsorbed Impurities*

The removal of such impurities will increase the efficiency of the adsorbent. Adsorbents, the efficiencies of which are increased by these cleaning processes or by an increase in their surface area, are often termed activated adsorbents. For example, activated charcoal is obtained by heating charcoal in steam at high temperatures (e.g. 500 to 900 °C) in order to remove the impurities that remain from the initial preparation of the charcoal. Alumina can be activated by heating at high temperatures (e.g. 500 to 900 °C) to remove adsorbed water. The activated alumina is often deactivated by the addition of calculated amounts of water to give grades of known activities.

In these activation procedures care must be taken to obtain maximum activity. For example, the use of high temperatures in the activation of alumina may result in a decrease in the total surface area by a 'sintering' process. In addition, an inactive form of alumina (α-alumina) is produced at temperatures in excess of 1000 °C.

5. *Adsorbent-Solute Interactions and Solvent Competition*

The adsorption of a solute from dilute solution involves breaking solute-solvent bonds and adsorbent-solvent bonds and the formation of adsorbent-solute bonds. The strength of the various interactions between the components of a system, i.e. adsorbent, solute and solvent, will depend on their mechanisms, which, in turn, depend on the structures of the components. It is quite possible that an adsorbent will have a strong affinity for a particular type of solute, which will therefore be adsorbed preferentially from a mixed solution. This gives rise to what is known as selective adsorption. For example, charcoal will adsorb the dye magenta from solution. However, if saponin is added to the system the magenta is released, since the saponin is preferentially adsorbed.

The solvent can influence the amount of adsorption in a variety of ways, the effects of which are the basis of the main difference between adsorption from solution and the adsorption of gases. For example, if the solvent is inert, i.e. has no appreciable affinity for the adsorbent or the solvent, then the adsorption of the solute will be at a maximum. However, the amount of solute adsorbed will decrease if—

(*a*) the solvent has an affinity for the adsorbent, since there will be a competition between the solute and the solvent for the adsorbent surface. They may be adsorbed at the same or different sites on the surface. The first mechanism will involve a direct competition and the second may lead to a decrease in the number of sites available for solute adsorption because of the overlapping by solvent molecules adsorbed at adjacent sites;

(*b*) the solvent has an appreciable affinity for the solute, since there will be a competition between the solvent and the adsorbent for the solute. Thus, a change in solvent from one in which the solute is sparingly soluble to one in which it is appreciably soluble will lead to a decrease in the amount of adsorption. The removal of adsorbed material by the use of another solvent is known as elution; or

(*c*) a combination of the above two effects occurs.

6. *pH of the Solution*

In solutes that show incomplete dissociation into ions in solution the effect of a change in pH will depend on whether the ionised or unionised species is the most strongly adsorbed.

APPLICATIONS OF ADSORPTION IN PHARMACY AND ALLIED FIELDS

1. *Decolorising Agents*

The removal of traces of coloured impurity is often achieved by the use of an adsorption process. For example, the decolorisation of a liquid may be effected by shaking or heating with approximately one per cent of activated charcoal. After standing, the charcoal plus adsorbed impurity is removed by filtration. Care must be exercised when carrying out decolorisation processes in order to ensure that active ingredients are not also removed by adsorption. This is particularly likely to happen if alkaloidal solutions are decolorised by the use of activated charcoal.

An attempt is made to standardise the adsorptive capacity of decolorising charcoal in the BP by comparing the amount of bromophenol blue remaining in alcoholic solution after adsorption has occurred. Two tests for the adsorptive capacity of Charcoal BPC are given. One involves the adsorption of chloroform vapour and the other involves adsorption of phenazone from aqueous solution. The results of tests such as these should be treated cautiously, since it does not necessarily follow that the decolorising powers of various samples of charcoal for the removal of other solutes will parallel the results obtained with bromophenol blue or phenazone.

2. *Desiccants and Drying Agents*

Water is strongly adsorbed by alumina and silica gel. The last traces of water are therefore often removed from organic solvents by passage through a column of alumina or silica gel. These materials are also used as desiccants to maintain a dry atmosphere inside electronic apparatus, and small packets are sometimes included inside the containers of pharmaceutical preparations that may be affected adversely by a high humidity.

3. *Adsorption Chromatography*

This is a separation technique based on the difference in affinity of an adsorbent for various solutes. The different affinities result in a variation in the partition of the solutes between the liquid/solid interface and the liquid phase. If the latter moves through a stationary column of the adsorbent, then the rates of movement of the solutes will vary; for example, a strongly adsorbed solute will spend more time adsorbed on to the stationary phase than a weakly adsorbed solute and, consequently, the former will move through the column at a slower rate than the latter. Eventually some degree of separation will be obtained.

The applications of chromatography to compounds of pharmaceutical interest are many, since it is a standard method of separation and purification in most laboratories. The student is therefore recommended to read one of the available textbooks on this subject.

4. *Surface Area Determination*

Many of the properties of a powder are influenced by its surface area; e.g. rate of solution, rate of oxidation, hygroscopicity, sedimentation behaviour, resistance to gas flow, bulk density and associated packing problems.

Measurement of the adsorption of gases or solutes provides one of the commonest means of determining the specific area, S_a, i.e. the surface area per unit weight (or unit volume) of the powder. The majority of the adsorption methods are based on Eqn (4.24) and involve the determination of the following two parameters.

(*a*) The monolayer capacity, x_m, i.e. the amount of adsorbate required to produce a complete monolayer on the adsorbent surface.

(*b*) The surface area occupied by each molecule in the monolayer (A_0). This involves a knowledge of the orientation of the adsorbed molecules with respect to the surface.

$$S_a = x_m \,.\, N_A \,.\, A_0 \qquad (4.24)$$

In Eqn (4.24), x_m is expressed in moles and N_A is Avogadro's constant ($6{\cdot}023 \times 10^{23}$ mole^{-1}).

The accurate determination of x_0 is difficult in the case of adsorption from solution because adsorbed solvent molecules may prevent close packing of the adsorbed solute in the formation of a monolayer; i.e. the latter may consist of solute and solvent molecules. In addition, the orientation of the adsorbed solute will vary with the type of solvent, and A_0 is difficult to evaluate, especially in complex asymmetric molecules such as dyes. (These compounds are often used in surface area determinations because of the ease of analysis of coloured solutions.) These difficulties are either avoided or are not so marked in the adsorption of small gas molecules, e.g. nitrogen and argon. Absolute values for the specific surface area are therefore more reliable when obtained from gas adsorption measurements, although reasonable agreement has been obtained between such values and those determined by the adsorption of 4-nitrophenol from solution (Giles and Nakhwa, 1962). The common application of adsorption from solution is in the measurement of specific surface areas on a relative basis, e.g. for quality control. This use arises from the experimental simplicity of the method compared with the practical difficulties associated with the measurement of gas adsorption.

5. *Medicinal Adsorbents*

Kaolin and charcoal are both used as gastrointestinal adsorbents for the removal of toxic materials, and charcoal is also useful as a general antidote for poisoning by compounds such as atropine, strychnine, oxalic acid, and phenolphthalein. It should be borne in mind that these adsorbents may also take up vitamins, drugs, enzymes, and trace materials. Their continued use may therefore lead to unwanted effects such as vitamin and mineral deficiencies, incomplete medication, and interference with digestive processes.

Kaolin is useful in external preparations such as poultices for dressing boils, suppurating wounds and ulcers. Kaolin is also used for its drying properties in dusting powders for application to moist lesions.

6. *Miscellaneous Applications*

(*a*) Activated charcoal has been used for removing pyrogens from injections (Gunn and Carter, 1965).

(*b*) The concentration of penicillin and streptomycin may be effected by adsorption from dilute solution followed by recovery.

(*c*) The pore sizes of some filters are often slightly larger than the normal dimensions of retained particles. However, penetration of filters is prevented by several mechanisms, one of which involves adsorption of the particles onto the walls of the pores.

(*d*) The stability of colloids is often dependent on the adsorption of ions onto their surfaces, and the mechanism of action of protective colloids is thought to involve the adsorption of hydrophilic material onto hydrophobic particles (*see* Chapter 5).

(*e*) The stability of emulsions depends on the adsorption of the emulsifying agent at the oil/water interface (*see* Chapter 5).

(*f*) Pharmacological activity may involve adsorption of drug molecules at receptor sites in the cell (Ariens and Simonis, 1964).

(*g*) The rheological properties of suspensions are markedly affected by the adsorption of various materials especially surface active agents, at the liquid/solid interface (*see* Chapter 7).

(*h*) The adsorption of drugs by solid excipients included in many pharmaceutical formulations may have an important effect on the rate of drug release from these formulations and on the rate of absorption of the drugs (Wurster and Polli, 1961 and 1964; Sorby, 1965; Sorby *et al.*, 1966).

Electrical Effects at Interfaces

The existence of a difference in electrical potential across a solid–liquid interface is demonstrated by the following phenomena.

(*a*) *Electrophoresis*

This involves the movement of dispersed particles through a liquid medium under the influence of an electrical field.

(*b*) *Electro-osmosis*

This involves the movement of a liquid relative to a fixed solid under the influence of an applied field.

(*c*) *Streaming Potential*

This is the potential difference set up across a fixed porous plug of solid when a liquid is forced through it.

(*d*) *Sedimentation Potential*

This is the potential difference set up between the top and bottom of a suspension of solid particles in a liquid when the particles settle under the influence of gravity.

The above phenomena are termed electrokinetic phenomena since they are concerned with a potential difference associated with the relative movement between two phases separated by an interface.

The difference in electrical potential across an interface indicates that there must be a particular distribution of charges near to the interface. This distribution is referred to as an electrical double layer and its initial structure was proposed by Helmholtz

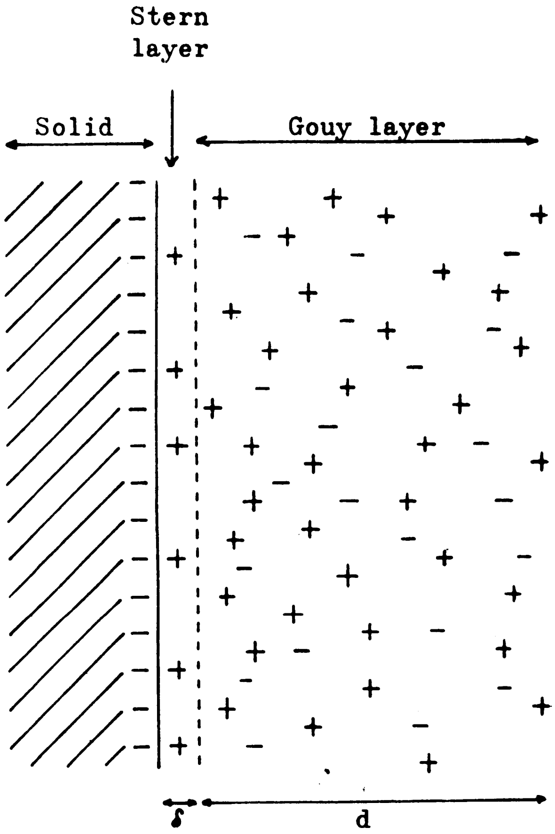


Fig. 4.9 The electrical double layer at a flat negatively charged solid surface

in 1879. Various modifications have been suggested since then, and the present theory is discussed below.

The Electrical Double Layer

Consider a solid surface carrying negative charges in contact with an aqueous solution containing positive and negative ions. The negatively charged solid surface will influence the distribution of ions in the neighbouring layers of the solution. Thus, positive ions will be attracted towards the solid surface and negative ions repelled away from it. However, this distribution will also be affected by thermal agitation which will tend to redisperse the ions in solution. The resulting effect creates a diffuse layer of solution, in which the concentration of positive ions gradually decreases on moving away from the interface, and the concentration of negative ions gradually increases. However, this diffuse layer contains a sufficient excess of positive ions to neutralise the negative charges on the solid surface. This type of distribution of charges is referred to as an electrical double layer. The surface charges form the fixed part of this double layer and the layers of liquid containing the excess oppositely charged ions constitute the diffuse part. The separation of charge in the electrical double layer gives rise to a decrease in potential, ψ, across the double layer from its value at the surface of the solid, ψ_0, to zero at the edge of the diffuse layer where a condition of electroneutrality is reached.

Sometimes thermal agitation is insufficient to overcome the attractive forces between the solid surface and oppositely charged ions. For example, strong attractions will exist if the ions in solution carry several charges (e.g. multivalent ions) or if additional forces are involved in the adsorption process (e.g. as in the adsorption of surface active agents). Strong adsorption of these types of oppositely charged ions will result, and these are held in the so-called Stern layer. The remaining oppositely charged ions are distributed in a diffuse layer (the Gouy layer) similar to that described previously. The structure of the electrical double layer is shown diagrammatically in Fig. 4.9 in which δ and d

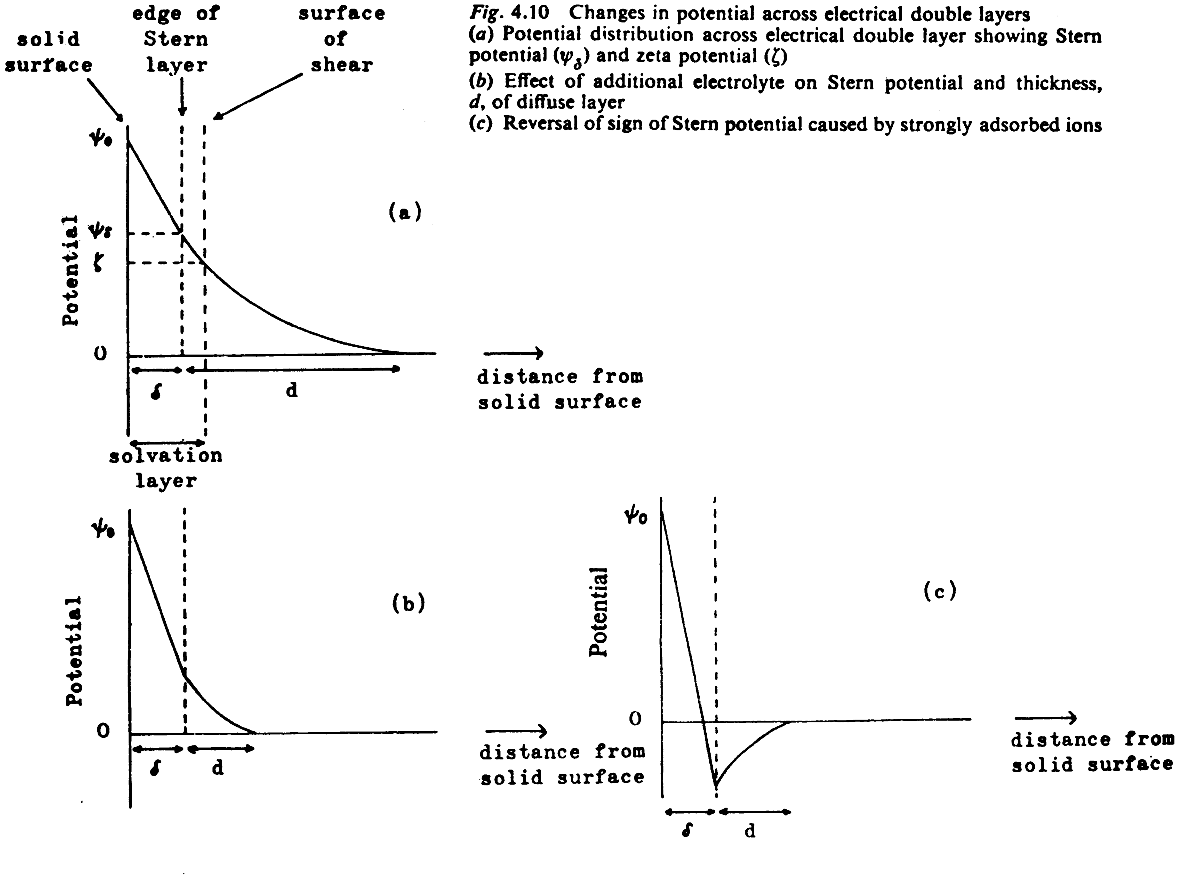


Fig. 4.10 Changes in potential across electrical double layers
(*a*) Potential distribution across electrical double layer showing Stern potential (ψ_δ) and zeta potential (ζ)
(*b*) Effect of additional electrolyte on Stern potential and thickness, *d*, of diffuse layer
(*c*) Reversal of sign of Stern potential caused by strongly adsorbed ions

represent the thicknesses of the Stern and Gouy layers respectively.

The distribution of ions shown in Fig. 4.9 will affect the potential at varying distances from the surface of the solid. This change is illustrated by Fig. 4.10(*a*). It can be seen that the potential decreases linearly across the Stern layer from the surface potential, ψ_0, to the Stern potential, ψ_δ, which is the value at the boundary between the Stern and Gouy layers, and then falls more and more slowly until it is zero at the edge of the Gouy layer.

A layer of liquid molecules will also be adsorbed on to the solid. This solvating layer is strongly held to the surface and is therefore of importance in electrokinetic phenomena since the outer surface of this layer represents the boundary of relative movement between the solid and the liquid. This surface is therefore termed the surface of shear. The potential at this point is termed the zeta, ζ, or electrokinetic potential, and it may be calculated from measurements of the electrokinetic phenomena mentioned previously. The thickness of the solvating layer of liquid molecules is ill-defined and the zeta potential therefore represents a potential at an unknown distance from the solid surface. Although the solvating layer is only one molecule thick in most cases, it is usually thicker than the Stern layer, unless the latter contains large adsorbed ions. The zeta potential is therefore usually lower than the Stern potential as shown in Fig. 4.10(*a*).

If ions are added to the solution, the diffuse part of the double layer is compressed; i.e. its thickness, *d*, decreases as shown in Fig. 4.10(*b*). In addition, if the oppositely charged ions are strongly adsorbed at the surface in the Stern layer then the Stern potential will also decrease. In fact, if sufficient strongly adsorbed ions are held in the Stern layer the surface charges may be more than balanced and the sign of the Stern potential will be reversed as shown in Fig. 4.10(*c*). The effects of electrolytes on the Stern potential will also be paralleled by their effects on the zeta potential.

BIBLIOGRAPHY

ADAM, N. K. (1941) *The Physics and Chemistry of Surfaces.* 3rd ed. Oxford University Press, London.

ADAMSON, A. W. (1967) *Physical Chemistry of Surfaces.* 2nd ed. Interscience, New York.

BIKERMAN, J. J. (1958) *Surface Chemistry.* 2nd ed. Academic Press, New York.

DAVIES, J. T. and RIDEAL, E. K. (1963) *Interfacial Phenomena.* 2nd ed. Academic Press, New York.

GREGG, S. J. (1965) *The Surface Chemistry of Solids.* 2nd ed. Chapman and Hall, London.

KIPLING, J. J. (1965) *Adsorption from Solutions of Non-electrolytes.* Academic Press, New York.

MOILLIET, J. L., COLLIE, B., and BLACK, W. (1961) *Surface Activity.* 2nd ed. Spon, London.

TROTMAN-DICKENSON, A. F. and PARFITT, G. D. (1966) *Chemical Kinetics and Surface/Colloid Chemistry.* Pergamon Press, London.

VERWEY, E. J. W. and OVERBEEK, J. TH. G. (1948) *Theory of the Stability of Lyophobic Colloids.* Elsevier, Amsterdam.

REFERENCES

ARIENS, E. J. and SIMONIS, A. M. (1964) A molecular basis for drug action. *J. Pharm. Pharmac.*, **16**, 137–157.

ARONOW, R. H. and WITTEN, L. (1958) Theoretical derivation of Traube's rule. *J. chem. Phys.*, **28**, 405–409.

BRUNAUER, S., DEMING, L. S., DEMING, W. S., and TELLER, E. (1940) On a theory of the van der Waals adsorption of gases. *J. Amer. chem. Soc.*, **62**, 1723–1732.

BRUNAUER, S., EMMETT, P. H., and TELLER, E. (1938) Adsorption of gases in multimolecular layers. *J. Amer. chem. Soc.*, **60**, 309

DIXON, J. K., JUDSON, C. M., and SALLEY, D. J. (1954) *Monomolecular Layers.* p. 63, American Association for the Advancement of Science, Washington, D.C.

GILES, C. H., MACEWEN, T. H., NAKHWA, S. N., and SMITH, D. (1960) Studies in adsorption. **Part XI.** A system of classification of solution adsorption isotherms, and its use in diagnosis of adsorption mechanisms and in measurement of specific surface areas. *J. chem. Soc.*, 3973-3993.

GILES, C. H. and NAKHWA, S. N. (1962) Studies in adsorption. Part XVI. The measurement of specific surface areas of finely divided solids by solution adsorption. *J. appl. Chem., Lond.*, **12**, 266–273.

GREGG, S. J. (1965) *The Surface Chemistry of Solids.* 2nd ed. pp. 55–60. Chapman and Hall, London.

GUNN, C. and CARTER, S. J. (1965) *Dispensing for Pharmaceutical Students.* 11th ed. Pitman Medical, London.

GURVICH, L. (1915) *J. Russ. Phys. Chem. Soc.*, chem. part, **47**, 805. (Taken from Bikerman, J. J. (1958) *Surface Chemistry.* 2nd ed. p. 203. Academic Press, New York.)

HARKINS, W. D. and BROWN, F. E. (1919) Determination of surface tension (free surface energy) and the weight of falling drops—surface tension of water and benzene by the capillary height method. *J. Amer. chem. Soc.*, **41**, 499.

HARKINS, W. D. and JORDAN, H. F. (1930) Determination of surface tension. *J. Amer. chem. Soc.*, **52**, 1751.

KIPLING, J. J. (1951) Adsorption of non-electrolytes from solution. *Q. Rev. chem. Soc.*, **5**, 60–74.

KIPLING, J. J. (1965) *Adsorption from Solutions of Non-electrolytes*. p. 25. Academic Press, London.

KIPLING, J. J. and TESTER, D. A. (1952) Adsorption from binary mixtures: Determination of individual adsorption isotherms. *J. chem. Soc.*, 4123–4133.

LANGMUIR, I. (1916) *J. Amer. chem. Soc.*, **38**, 2219.

LOWER, E. S. (1962) Dictionary of British surface active agents. Parts 1–7 and Additional Entries. *Mfg Chem.*, 33, 141–144, 181–184, 241–244, 285–288, 333–336, 371–374, 421–424 and 478.

MCBAIN, J. W. and SWAIN, R. S. (1936) Measurements of adsorption at the air-water interface. *Proc. R. Soc.*, **A154**, 608.

MOILLIET, J. L., COLLIE, B., and BLACK, W. (1961) *Surface Activity*. 2nd ed. Spon, London.

SCHWARTZ, A. M. and PERRY, J. W. (1949) *Surface Active Agents. Their chemistry and technology. Vol. 1.* Interscience, New York.

SORBY, D. L. (1965) Effect of adsorbents on drug absorption. I. Modification of promazine absorption by activated attapulgite and activated charcoal. *J. pharm. Sci.*, **54**, 677–683.

SORBY, D. L., PLEIN, E. M., and BENMAMAN, J. D. (1966) Adsorption of phenothiazine derivatives by solid adsorbents. *J. pharm. Sci.*, **55**, 785–794.

TATE, T. (1864) *Phil. Mag.*, **27**, 176.

THOMAS, J. M. (1961) Textbook errors, 29. The existence of endothermic adsorption. *J. chem. Educ.*, **38**, 138–139.

TRAUBE, I. (1891) *Justus Liebigs Annln. Chem.*, **265**, 27.

WARD, A. F. H. (1946) Thermodynamics of monolayers on solutions. I. Theoretical significance of Traube's rule. *Trans. Faraday Soc.*, **42**, 399–407.

WURSTER, D. E. and POLLI, G. P. (1961) Investigation of drug release from solids. IV. Influence of adsorption on the dissolution rate. *J. pharm. Sci.*, **50**, 403–406.

WURSTER, D. E. and POLLI, G. P. (1964) Investigation of drug release from solids. V. Simultaneous influence of adsorption and viscosity on the dissolution rate. *J. pharm. Sci.*, **53**, 311–314.

5

Disperse Systems

A DISPERSE system may be defined as a system in which one substance (the disperse phase) is dispersed as particles throughout another (the dispersion medium or continuous phase). This type of system can be classified on the basis of the physical states of the two substances. Since matter can exist as gas, liquid, or solid, then nine different types of disperse system are possible in the same way that nine different types of solution are possible (*see* Table 2.1, Chapter 2). With the exception of dispersions of gases in gases, these systems can be further classified on the basis of the size of the dispersed particles as shown in Table 5.1. It should be borne in mind that the sub-division given in the table is not rigid and there is no sharp demarcation between the particle sizes in these systems.

Since true solutions have been discussed in Chapters 2 and 3 the present discussion will be restricted to colloidal dispersions and coarse dispersions. Some examples of these systems are given in Table 5.2.

The Stability of Disperse Systems

The dispersion of one phase as small particles throughout another produces a considerable area of contact between the two phases. Any free energy associated with the interface between the phases will be decreased if the particles aggregate or coalesce because of the reduction in interfacial area that accompanies such aggregation. Since any system will tend to react spontaneously to decrease its free energy to a minimum, it follows that disperse systems are often unstable. Thus, if the interfacial free energy has a positive value, then the particles will eventually aggregate rather than remain in contact with the dispersion medium. Dispersions that exhibit this behaviour are termed lyophobic dispersions. In other systems, known as lyophilic dispersions, an affinity exists between the dispersed particles and the dispersion medium, and this contributes to the stability of these systems.

Colloidal dispersions are usually more stable (i.e. will remain as dispersions longer) than coarse dispersions, since the larger particles in the latter settle more rapidly under the influence of gravity and, unlike colloidal systems, the maintenance of their dispersions is not aided by Brownian movement (*see* below). Some types of disperse system are of particular importance in pharmacy. The factors that contribute to their stability will therefore be discussed together with the preparations and properties of the dispersions.

Colloidal Dispersions of Solids in Liquids

The majority of these systems of interest to pharmacists are dispersions in an aqueous medium. Unless otherwise stated, the discussion in the remainder of this section will be restricted to aqueous systems. The specific terms 'hydrophilic' and 'hydrophobic' will therefore be used instead of the general terms 'lyophilic' and 'lyophobic'. A hydrophilic colloid is a system in which the dispersed particles have an affinity for the aqueous dispersion medium, and a hydrophobic colloid is a system in which the particles exhibit little or no such affinity.

Mechanisms of Stabilisation

The affinity of hydrophilic colloidal particles for an aqueous dispersion medium is sufficient to render these dispersions thermodynamically stable. In fact, hydrophilic materials form colloidal dispersions spontaneously on addition to water.

However, spontaneous coalescence of hydrophobic colloids indicates that these systems are unstable and that attractive forces exist between the particles. Thus, if the latter are to remain in a dispersed state for any considerable time then a stabilising mechanism that reduces the rate of coalescence must operate.

The present theory for the stabilisation of lyophobic colloids was developed independently by Derjaguin and Landau (1941) and Verwey and Overbeek (1948), and is therefore referred to as the DLVO theory. It has also been well explained by

Table 5.1
Types of Disperse System

System	*Dispersed particles*	*Particle size*	*Notes*
True solutions	Small molecules or ions	Usually less than 1×10^{-6} mm	Mixtures of gases are restricted to this type of system
Colloidal dispersions	Single large molecules (or ions) or aggregates of small molecules (or ions)	Larger than those in true solution and have an upper size limit of about 1×10^{-3} mm	
Coarse dispersions	Aggregates of molecules	Larger than those in colloidal dispersions	Particle size ranges from about 1×10^{-3} mm to an upper limit, which depends on the system

other authors (Overbeek, 1952; van Olphen, 1963; Sheludko, 1966). To understand this theory it is necessary first to consider the repulsive and attractive forces that operate between approaching particles, and then to examine the combined effects of these opposing forces.

Repulsive Forces between Approaching Particles

The existence of electrical charges on dispersed particles is demonstrated by the process of electrophoresis (*see* later). These charges originate either from ionisation of surface groups or from preferential adsorption of specific ions, known as peptising ions, from solution. The former mechanism applies frequently in hydrophilic colloids, and the latter in hydrophobic colloids.

Table 5.2
Examples of Disperse Systems
(true solutions omitted)

Dispersed phase	*Dispersion medium*	*Examples*	
		Colloidal	*Coarse*
Liquid	Gas	Fog	Spray
Solid	Gas	Smoke	Dust
Gas	Liquid	Foam	Foam
Liquid	Liquid	Oil globules $<1 \times 10^{-3}$ mm in water	Emulsion
Solid	Liquid	Colloidal gold in water	Suspension of kaolin in water
Gas	Solid	Solid foam	Solid foam
Liquid	Solid	Mineral oil in wax	Solid emulsion
Solid	Solid	Colloidal gold in glass	Solid suspension

The presence of charges on the dispersed particles influences the distribution of positive and negative ions in the layers of solution that surround each particle. This distribution is also influenced by thermal motions in the solution. The resultant effect is that each particle is surrounded by an electrical double layer with a structure similar to that described in Chapter 4 for plane solid–liquid interfaces.

When two particles approach each other the diffuse parts of their electrical double layers overlap. This leads to a redistribution of the charge in each layer, and a repulsion will be exerted between the particles. Work must therefore be performed in order to overcome this repulsion and bring the particles close together. The amount of work necessary to bring the particles from an infinite distance apart to a given distance of separation may be calculated. This amount of work is equal to the repulsive potential at that given distance of separation. The continuous line in Fig. 5.1 shows the variation in the repulsive potential with distance of separation of the particles.

Attractive Forces between Approaching Particles

The repulsive forces that arise from the overlapping of electrical double layers still operate even when particles aggregate. An attractive force must therefore exist between particles, and this force must be comparable in its magnitude and range of operation to the repulsive force if the latter is to be overcome during an aggregation process. It is considered that the attraction is provided by van der Waals forces. Although the Appendix indicates that between a pair of atoms these forces are weak and only of short range, it should be pointed out that they are also additive. Thus, the total attraction between two particles will be equal to the sum of all the attractive forces between every atom of one particle and all those of the other. This additive effect not only produces an appreciable force but also increases the range over which this force is exerted. The broken line in Fig. 5.1 represents the variation in attractive potential with distance of separation of the particles.

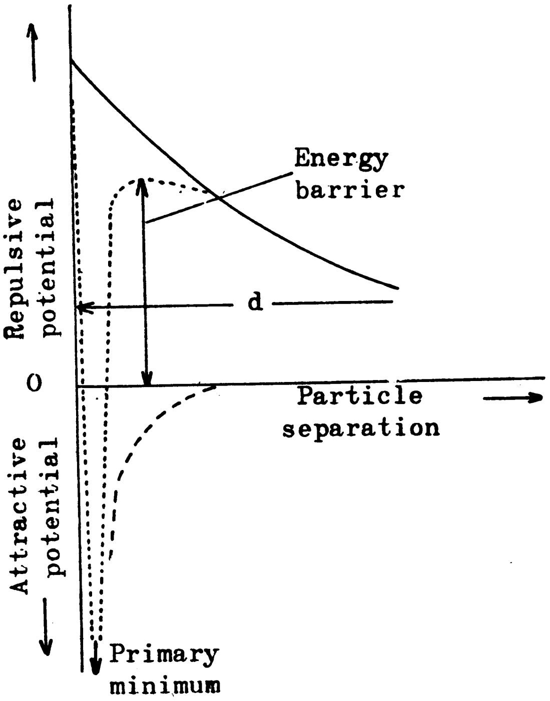


Fig. 5.1 Variation in net potential of interaction (. . . . .) arising from the combined effects of repulsive potential (———) and attractive potential (- - - -) between two approaching particles

Combined Effects of Repulsive and Attractive Forces between Approaching Particles

The dotted line in Fig. 5.1 represents the variation in net potential of interaction with the distance of separation of the particles. This curve is obtained by addition of the repulsive and attractive potentials at each distance, assuming that these separate potentials have opposite signs; e.g. the repulsive potential is regarded as being positive and the attractive potential as negative.

The net interaction curve indicates that as two particles approach each other the repulsive potential predominates initially, the attractive potential then predominating as the distance between the particles decreases.

The sharp rise of the net interaction curve when the particles are very close is caused by additional short-range repulsive forces that have not been taken into account in the previous discussion. These forces may arise from the contact of points on the irregular surface of the particles (the Born repulsion) and, also, from the presence of strongly adsorbed molecular layers of water on each particle. These layers must be desorbed before the particles can approach to closer distances than the thicknesses of these layers.

Since the repulsive potential predominates at the longer distances of separation, coalescence will occur only if the approaching particles possess sufficient activation energy to overcome this repulsion barrier, and pass into the region where the attractive potential predominates. The height of the energy barrier in Fig. 5.1 will determine its effectiveness in preventing the aggregation of dispersed particles. If this barrier is reduced, then the stability of the colloidal dispersion will decrease.

Effect of Electrolytes

An increase in the concentration of ions in the dispersion medium has the effect of compressing the diffuse part of the electrical double layer; i.e. the thickness, *d*, of this layer is decreased and the repulsive potential decays more rapidly. Thus, the energy barrier that opposes aggregation, decreases as the concentration of added electrolyte increases, and finally disappears as indicated by Fig. 5.2. When the barrier ceases to exist, rapid coalescence of the particles occurs and large aggregates become visible in the system.

The degree of compression of the double layer depends not only on the concentration of added electrolyte but also on the valency of the ion of opposite charge to the colloidal particle. This is known as the Schulze-Hardy rule. Thus, a negatively charged colloid would be coagulated rapidly by a certain concentration of aluminium chloride in the dispersion medium but a greater concentration of barium chloride would be required, since barium cations are only divalent whereas aluminium cations are trivalent. In fact, calculations show that the flocculation value of an electrolyte should be inversely proportional to the sixth power of the valency of the ion of opposite charge. Experimental results show good agreement with these calculations.

The coagulating effect of electrolytes may be used to assess the original stability of a colloidal dispersion. Thus, the greater the concentration of electrolyte required to produce rapid coagulation, the greater the stability of the original system.

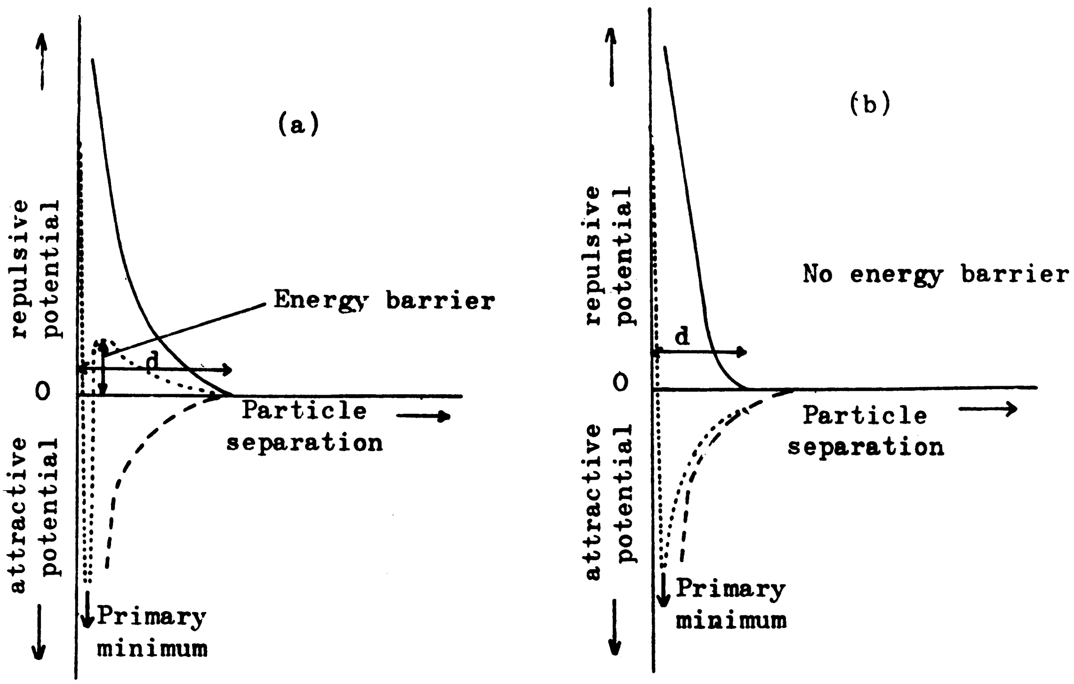


Fig. 5.2 The variation in thickness, *d*, of the electrical double layer and height of energy barrier to coagulation caused by (*a*) intermediate, and (*b*) higher concentrations of electrolyte (*See* Fig. 5.1 for key)

Preparation of Colloids

HYDROPHILIC COLLOIDS

As previously stated, the affinity of hydrophilic colloids for water leads to the spontaneous formation of a colloidal dispersion when the material is placed in contact with water. For example, acacia will readily disperse in water, and gelatin will disperse in hot water. This simple method of dispersion is a general one for the formation of lyophilic colloids.

HYDROPHOBIC COLLOIDS

The preparative methods for hydrophobic colloids may be divided into those methods that involve the breakdown of larger particles into particles of colloidal dimensions (dispersion methods) and those in which the colloidal particles are formed by aggregation of smaller particles such as molecules (condensation methods)

Dispersion Methods

The breakdown of coarse material may be effected by several means.

Colloid Mills. These mills cause the dispersion of coarse material by shearing in a narrow gap between a static cone and a rapidly rotating cone. A diagram of a colloid mill is given in Fig. 5.3.

Electrical Dispersion (Bredig's method). Certain metals may be dispersed by the passage of an electric arc between electrodes made of the metal and immersed in the dispersion medium.

Ultrasonic Irradiation. The passage of ultrasonic waves through a dispersion medium produces alternating regions of cavitation and compression in the medium. The cavities collapse with great force and cause the breakdown of coarse particles dispersed in the liquid.

Peptisation. Because the charges necessary for stabilising colloidal dispersions may originate from the preferential adsorption of specific ions at the surface of the particles, a finely divided solid may be converted into a colloidal dispersion by the addition of such ions to the dispersion medium. This process is known as peptisation.

Condensation Methods

These involve the rapid production of supersaturated solutions of the colloidal material under conditions in which it is deposited in the dispersion medium as colloidal particles and not as a precipitate. The supersaturation is often obtained by means of a

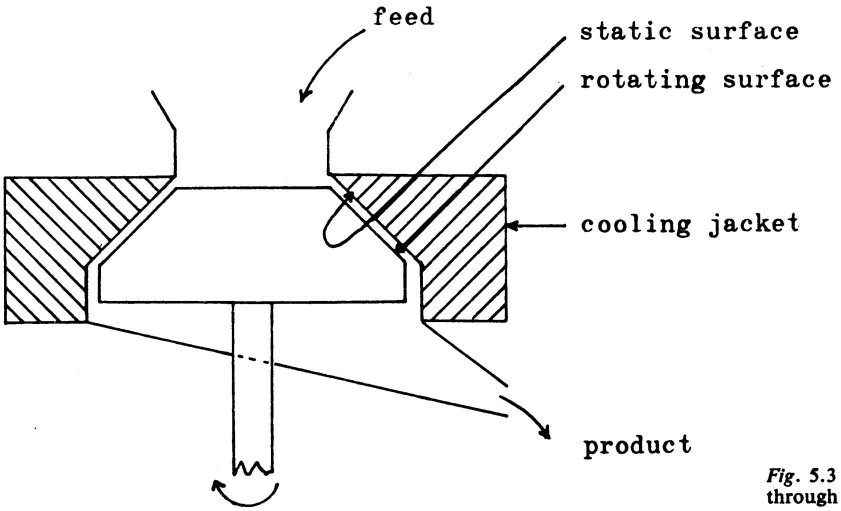


Fig. 5.3 Cross-section through a colloid mill

chemical reaction that results in the formation of the colloidal material. For example, colloidal sulphur may be obtained by passing hydrogen sulphide through a solution of sulphur dioxide, and colloidal arsenious sulphide may be obtained by passing hydrogen sulphide through a solution of arsenic trioxide.

A change in solvent may also cause the production of colloidal particles by condensation methods. If an alcoholic solution of a resin (resinous tincture) is poured slowly into water, a colloidal dispersion of resin particles may be obtained.

Purification of Colloids

DIALYSIS

Colloidal particles are too large to diffuse through the pores of certain membranes such as parchment or cellophane. The smaller particles in true solution are able to pass through these membranes. Use is made of this difference in diffusibility to separate micromolecular impurities from colloidal dispersions. The process is known as dialysis.

A colloidal dispersion may become diluted during dialysis because water will pass through the membrane under the influence of the osmotic pressure of the colloid. Such dilution may be prevented by applying to the colloid a pressure that is equal to or greater than its osmotic pressure. This procedure is referred to as dialysis under pressure.

ELECTRODIALYSIS

An electrical potential may be used to increase the rate of movement of ionic impurities through a dialysing membrane and so provide a more rapid means of purification. This method is of little use if any of the impurities are uncharged and care should be taken to ensure that the electrical potential does not affect the stability of the colloid.

ULTRAFILTRATION

Colloidal particles are too small to be retained by ordinary filter papers. However, filtration through a dialysing membrane may be carried out, but because the rate of flow of the liquid dispersion medium through such membranes is slow it is usual to carry out the filtration process under the influence of either positive pressure or vacuum. Since the sizes of the colloidal particles and the pores in the membrane are so small, the process is referred to as ultrafiltration.

It is possible to manufacture membranes with different degrees of porosity (*see* later). The use of a series of ultrafilters with gradually decreasing pore sizes allows the particle size of a colloid to be determined. Conversely, the pore size of an ultrafilter can be determined by the use of a series of colloidal dispersions of different particle sizes.

The Properties of Colloids

PARTICLE SIZE

The particles in a colloidal dispersion fall within the approximate range of 1×10^{-6} to 1×10^{-3} mm. The actual size may be determined by ultramicroscopic measurements, the use of graded ultrafilters, or the measurement of the rates of sedimentation of the particles in an ultracentrifuge.

ELECTRICAL PROPERTIES

If an electric current is passed through a colloidal dispersion contained in a suitable cell, the particles migrate towards one of the electrodes. This process, which is known as electrophoresis, demonstrates that all the particles in a given system carry a similar charge. In addition, the sign of this charge is indicated by the direction of movement of the particles. Thus, the particles of negatively charged colloids, such as kaolin, sulphur, arsenious sulphide, and metals, will move towards the anode, while the particles of positively charged colloids, such as ferric hydroxide and other metal hydroxides, will move towards the cathode.

Different colloidal materials may possess different electrophoretic mobilities and the process may therefore be used as a method of separating mixtures of colloidal materials. For example, the proteins of blood plasma may be separated by electrophoresis.

The charge carried by colloidal protein molecules, and therefore the direction of movement of these molecules under the influence of an electric current, will depend on the pH of the dispersion medium. This variation in the sign of the charge is explained by the fact that in alkaline solution the carboxylic acid groups of the protein molecules will exist as carboxylate anions, whereas in acid solutions the amino groups of the molecules will be protonated. Thus, proteins are negatively charged in alkaline solutions and positively charged in acid solutions. At an intermediate pH, known as the iso-electric point, the protein exists as a zwitterion, which is electrically neutral, although both groups are ionised. The solubility of the protein is at a minimum at its iso-electric point and therefore precipitation is facilitated at this pH.

OPTICAL PROPERTIES

Colloidal particles are too small to be seen under an ordinary microscope and a colloidal dispersion appears to be perfectly transparent when viewed from the front with a light behind it. However, the colour appears to change if the system is viewed against a dark background at right angles to the incident light, when a turbidity can be seen. Furthermore, if the incident light is restricted to a beam, then a cone of light, known as the Tyndall cone, is observed. These effects arise because the incident light is scattered by the colloidal particles. The progressively greater scattering effect produced as the light passes through the system is responsible for the formation of the Tyndall cone. In addition, the intensity of the scattered light increases with decreasing wavelength. The scattered light therefore has a greater proportion of smaller wavelengths than the incident light, while the transmitted light is deficient in these smaller wavelengths. These changes are responsible for the apparent difference in colour of the system when viewed by scattered or transmitted light. For example, a colloidal dispersion of a white material illuminated by white light will produce a bluish-white scattered light and a yellow-brown transmitted light.

The light-scattering effect is made use of in the design of the ultramicroscope, in which a cell containing the colloid is viewed against a dark background at right angles to an intense beam of incident light. The particles, which exhibit an erratic movement, appear as bright spots against the dark background. The motion of the particles is known as Brownian movement and arises from the bombardment of the particles by the molecules of the dispersion medium. This effect is useful in helping to counteract the effects of sedimentation and in maintaining the dispersion of particles in the continuous medium.

OSMOTIC PROPERTIES

Osmotic pressure is a colligative property (*see* p. 15) and therefore depends only on the number of particles present in a dispersion. The particles in a colloidal system may consist of aggregates of several molecules, but each of these aggregates acts as a single unit for colligative purposes. The osmotic pressure exerted by such a system will therefore be small. If the colloidal particles consist of single large molecules then the osmotic pressure may be used to calculate the molecular weight of the colloidal material. This method is often used for long-chain polymers, but a correction may be necessary to allow for the Donnan membrane effect, which is discussed later (*see* p. 73)

Differences in Properties of Hydrophilic and Hydrophobic Colloids

Some of the differences in the properties of the two types of colloid have already been mentioned in the previous sections. These and other differences are summarised in Table 5.3.

It will be observed from Table 5.3 that hydrophilic colloids are much more stable to the presence of added electrolytes than hydrophobic colloids. The instability of the latter type has already been explained in terms of the compression of the diffuse part of the electrical double layer in the presence of additional counter-ions. A similar effect is insufficient to cause the rapid coalescence of hydrophilic colloids since these are also stabilised by the affinity that exists between the colloidal particles and the

Table 5.3
A Comparison of the Properties of Hydrophilic and Hydrophobic Colloids

Property	*Hydrophilic colloids*	*Hydrophobic colloids*
1. Ease of dispersion of material in dispersion medium.	Usually occurs spontaneously.	Special treatment necessary.
2. Stability towards electrolytes.	High concentrations of very soluble electrolytes are necessary to cause precipitation.	Relatively low concentrations of electrolytes will cause precipitation.
3. Stability towards prolonged dialysis.	Stable.	Unstable because ions necessary for stability are removed.
4 Reversibility after precipitation (i.e. colloidal dispersion reforms when precipitated material is added to fresh dispersion medium).	Reversible.	Irreversible.
5. Tyndall effect.	Weak. Particles are not easily detected by ultramicroscope.	Strong. Particles are readily detected by ultramicroscope.
6. Viscosity.*	Usually higher than that of dispersion medium.	Similar to that of dispersion medium.
7. Protective ability (*see* later).	Capable of acting as protective colloids.	Incapable of acting as protective colloids and often require addition of such a material for stability.

* Further information on the viscosity of colloidal dispersions is given in Chapter 7.

dispersion medium. Precipitation of hydrophilic colloids therefore involves a reduction in this affinity as well as the removal of any energy barrier arising from a repulsive potential between the particles. Very soluble electrolytes such as ammonium sulphate will compress the double layer and so remove an energy barrier. In addition, since the ions of such electrolytes are strongly hydrated, high concentrations of such ions will compete so successfully for the molecules of the dispersion medium that the affinity of the latter for the colloidal particles is reduced. These combined effects lead to the precipitation of hydrophilic colloids, and the process is known as 'salting out'. Variation in the stabilities of different hydrophilic colloids affects the concentration of soluble electrolyte required to produce their precipitation. The components of a mixture of hydrophilic colloids can therefore be separated by a process of fractional precipitation, which involves the 'salting out' of the various components at different concentrations of electrolyte. This technique is used in the purification of antitoxins (*see* p. 403).

Effects of Mixing Different Colloids

1. MUTUAL PRECIPITATION

This may occur when two hydrophobic colloids with oppositely charged particles are mixed, due to the attractive forces exerted between the particles.

2. COACERVATE FORMATION

When oppositely charged hydrophilic colloids are mixed, a layer rich in colloidal material may separate. This layer is known as a coacervate.

3. SENSITISATION

In the presence of very small amounts of hydrophilic colloid a hydrophobic colloid may become more susceptible to precipitation by electrolytes. This effect, known as sensitisation, is thought to arise from the adsorption of various parts of each chain-like hydrophilic colloid onto several hydrophobic particles.

4. PROTECTION

The stability of hydrophobic colloids towards the precipitating effects of electrolytes is increased by the presence of hydrophilic colloids in concentrations several times those required for sensitisation. Hydrophilic colloids used in this way are termed protective colloids. Their protective action is considered to arise from their adsorption over the entire surface of each hydrophobic particle, so that the latter acquires some hydrophilic properties and is therefore able to withstand the effects of electrolytes.

The protection afforded to a specific hydrophobic colloid is often used as an indication of the relative abilities of a series of hydrophilic colloids to act as protective colloids. For example, the 'Gold Number', introduced by Zsigmondy, is the number of milligrammes of protective colloid that must be added to 10 cm^3 of a standard colloidal suspension of gold to prevent flocculation on the addition of 1 cm^3 of a 10 per cent NaCl solution. In the same way the 'Congo Red' number depends on the flocculation of a standard dispersion of congo red by a given amount of KCl. Such methods of comparing the protective abilities of hydrophilic colloids are influenced by the nature of the hydrophobic colloid. The relative abilities as indicated by Gold or Congo Red numbers may therefore be different for other systems.

Pharmaceutical Applications of Colloids

It has been suggested that the efficiency of certain substances used in pharmaceutical preparations may be increased if colloidal forms are used, since these have large surface areas; for example, the adsorption of toxins from the gastro-intestinal tract by kaolin, and the rate of neutralisation of excess acidity in the stomach by aluminium hydroxide may be increased if these compounds are used in colloidal forms.

The use of colloidal iron and colloidal iodine is said to alter the effects of these compounds, the former being less astringent than crystalloidal iron and the latter less toxic than iodine in aqueous solution.

In the purification of proteins, use is made of the changes in solubility of colloidal materials caused by the presence of electrolytes or changes in pH.

The protective ability of hydrophilic colloids is used to prevent the coagulation of hydrophobic particles in the presence of electrolytes. The increased viscosity of a liquid that results from dispersion of a hydrophilic colloid is used to retard the sedimentation of particles in pharmaceutical suspensions; i.e. the hydrophilic colloids are used as suspending agents.

Blood plasma substitutes are colloidal dispersions in which the particle size is such that they are retained in the blood vessels for an adequate time (e.g. Dextran Injection BP 1968).

Association Colloids

It has already been mentioned (Chapter 4) that surface active agents behave as normal compounds in dilute solutions but at certain reasonably well-defined concentrations relatively sharp changes occur in the physical properties of these solutions. These changes are attributed to the association of the amphipathic molecules or ions into aggregates of colloidal dimensions, that are known as micelles.

Ionic and non-ionic substances that exhibit this type of behaviour are referred to collectively as association colloids. Although the older term 'colloidal electrolyte' is strictly applicable to all ionised colloidal materials it is usually reserved for ionic association colloids. Since the early work in this field was carried out solely on ionic association colloids the term 'colloidal electrolyte' is still sometimes used erroneously as a synonym for 'association colloids'.

Critical Micelle Concentration and Micellar Structure

The minimum concentration at which physical properties of solutions of association colloids show marked changes is known as the critical micelle concentration, which is often written in an abbreviated form as CMC. The formation of micelles was originally suggested by McBain to explain the apparently anomalous changes in osmotic properties and electrical conductivity with concentration in solutions of ionic associated colloids. The conductivity indicated that a considerable degree of electrolytic dissociation was occurring in solution, whereas the osmotic properties indicated that considerable aggregation of ions into single colligative units was also occurring above the CMC. McBain's original suggestion allowed for the existence of two types of micelle. These types were—

(*a*) a small, approximately spherical, charged micelle, which existed in all concentrations, i.e. above and below the CMC, and which was largely responsible for the appreciable electrical conductivity, and

(*b*) a large undissociated lamellar micelle, which only existed above the CMC and was responsible for the low osmotic properties at such concentrations.

Hartley later suggested that the experimental facts could be explained on the basis of a single type of micelle. Hartley's model, shown in Fig. 5.4, consists of a spherical charged micelle with a radius approximately equal to the chain length of the amphipathic ion.

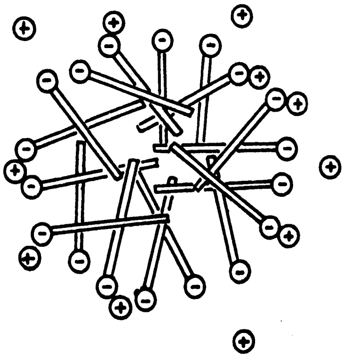


***Fig.* 5.4** **The Hartley spherical micelle**

The spherical type of micelle is now accepted as existing in all solutions of association colloids at and just above the CMC. However, in more concentrated solutions physical measurements, e.g. X-ray diffraction, viscosity, light scattering, indicate the existence of large asymmetric micelles. The rearrangement from spherical to larger and more widely separated asymmetric micelles has been ascribed to a reaction of the system in an effort to reduce the intermicellar repulsive forces that arise from the closer and closer approach of spherical micelles as the concentration of amphipathic material increases. Present evidence suggests that different micellar shapes, e.g. rods, lamellae, are formed in different systems.

STABILITY AND SIZE OF SPHERICAL MICELLES

The cohesive force between water molecules is much stronger than either the attraction between the lipophilic parts of the surface active agents or the attraction between water and the lipophilic chains. Therefore, the surface active agent tends to be squeezed out of solution in order to reduce the large degree of separation of water molecules that would be caused by the presence of many monomeric amphipathic molecules. This effect, which tends to cause a phase separation, is counterbalanced to some extent by the hydrophilic nature of the polar groups. In addition, the attractive forces between water molecules decay very rapidly with distance of separation since they are inversely proportional to somewhere between the fourth and seventh power of the distance (*see* Appendix). Thus, the work of separating water molecules by a relatively large distance on the formation of a micelle is little different from that involved in the introduction of an amphipathic monomer.

The electrical repulsion between adjacent similarly charged ions tends to disrupt the micelles of an ionic surface active agent. In such a case, micelle formation is therefore dependent on the balance between this disruptive effect and the constructive 'squeezing out of solution' effect. Since the electrical repulsive effect is absent in non-ionic micelles, various suggestions have been made regarding the existence of a factor that would tend to oppose micelle formation; e.g. cross-sectional area and solvation of the hydrophilic group. However, the precise nature of such a factor is still in doubt. The association of ionic and non-ionic surface active agents is also aided by the 'hydrophobic bonding' between the hydrocarbon chains. This type of bonding involves weak van der Waals forces of attraction, the effect of which is therefore of less significance than those mentioned previously. In addition, an increase in temperature will have a disruptive effect on the formation of micelles since their rate of deaggregation will be increased.

The size of a spherical micelle depends on the structure of the surface active agent. In the Hartley model of a micelle the radius is approximately equal to the length of the hydrocarbon chain. If the diameter were to increase beyond this point then the unlikely structure would either include a space in the centre into which the hydrocarbon chains could not reach, or the presence of some of the ionic groups between the hydrocarbon chains.

Physical Properties of Solutions of Association Colloids

Since micellar structure and shape influence the physical properties of solutions it is convenient to consider separately the properties of dilute solutions (i.e. up to concentrations just above the CMC, where the micelles are spherical and concentrated solutions containing asymmetric micelles.

DILUTE SOLUTIONS; PHYSICAL PROPERTIES AND METHODS OF DETERMINING CMC.

Colligative Properties

The association of monomeric ions or molecules into micelles at the CMC causes a marked decrease in the colligative properties of solutions of association colloids. A typical graph of osmotic pressure against concentration is shown in Fig. 5.5.

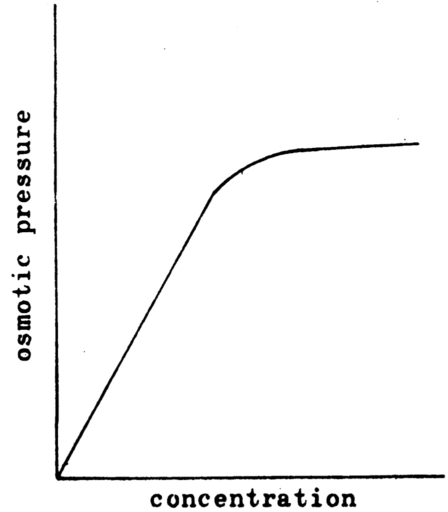


Fig. 5.5 Variation of the osmotic pressure with concentration of a solution of an association colloid

Osmotic pressure, freezing point and vapour pressure measurements have been used to determine CMCs. However, such determinations are not common since these methods are more tedious and more difficult than others, and their accuracy is dependent on the measurement of relatively small changes in physical properties. In addition, the depression of freezing point method is limited by the fact that many surface active agents are too insoluble at the freezing point of the solvent, and even measurements that are possible at these temperatures have little relationship to practical applications.

Surface Properties

The positive adsorption of surface active agents at the surface of water causes a progressive decrease in surface tension with increase in concentration until the CMC is reached. The micelles formed at and beyond this concentration provide an alternative means of removing the lipophilic groups from the aqueous environment, thus causing a decrease in free energy of the system. In fact, in such systems orientation of further surface active material in the form of micelles is energetically preferable to its orientation in the surface. Thus, additional material is used in the production of more micelles, and the concentrations in the surface layer and in true solution (i.e. as monomeric units) remain approximately constant beyond the CMC; a graph of surface tension against concentration levels out at this point, as shown in Fig. 5.6.

A minimum is frequently observed in the surface tension curves, shown by the dotted line in Fig. 5.6. Such minima are usually caused by the presence of surface active impurities in the system. The difficult and/or tedious purification procedures are usually responsible for the presence of such impurities. The initial adsorption into the surface layers of the surface active agent and impurity results in the lowering of the surface tension to a greater degree than that given by the pure surface active agent. At the CMC the impurity is leached out of the surface and taken up by the micelles so that the surface tension rises to that of the pure surface active agent beyond its CMC.

Surface tension measurements are useful in the determination of CMCs and the accuracy of the determination is little affected by the actual value of the CMC (cf. with other methods where the accuracy decreases as the CMC becomes smaller). Because of the slow attainment of adsorption equilibrium that occurs in some systems, static methods of surface tension measurement (i.e. methods in which the surface is not disturbed during measurement) are

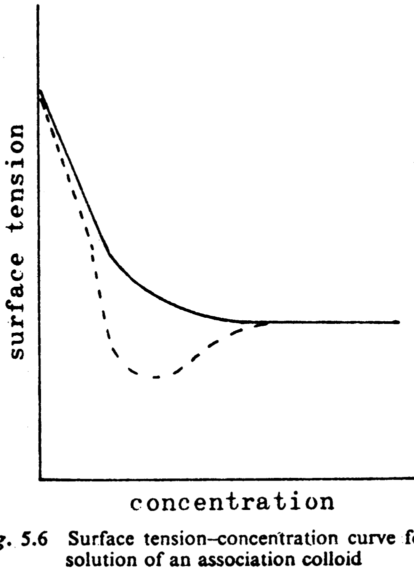


Fig. 5.6 Surface tension–concentration curve for a solution of an association colloid often more useful, since changes in surface tension with time can be followed.

Electrical Conductivity

The effects of micelle formation on the electrical conductivity of solutions of ionic association colloids are shown in Fig. 5.7(i) and (ii). According to Hartley three factors contribute to these changes.

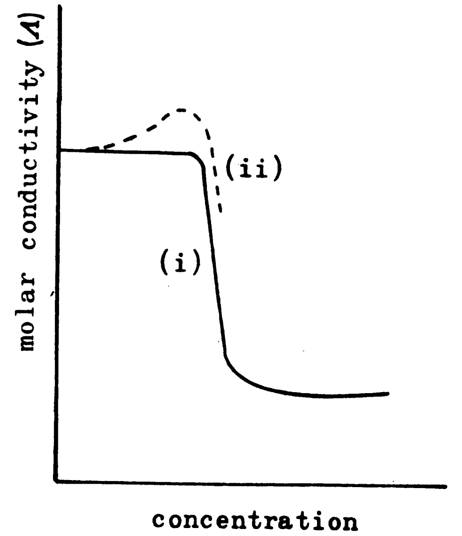


Fig. 5.7 The effect of micelle formation on the molar conductivity (Λ) of a solution at (i) low, and (ii) high field strengths

(*a*) *Change in Viscous Drag Effects.* The movement of ions towards an electrode is retarded by the viscous drag exerted on the charged particle by the solvent. For a spherical particle this resistance is proportional to the particle radius. Consider, therefore, the change that occurs in viscous drag when *n* spherical ions of radius *r* associate to form a single spherical micelle. The radius of the latter is given by $rn^{1/3}$. Thus, the ratio of the viscous drag on a single micelle to the total drag on *n* ions is $rn^{1/3}/rn = n^{-2/3}$. However, the total charge carried by the micelle is the same as that carried by all the separate ions, assuming that all the micellar ions remain dissociated. Therefore, the conducting power of the micelle will be greater than the total conducting power of the ions by a factor of $n^{2/3}$.

(*b*) *Changes in 'Braking Effect' of Gegenions.* The retarding effect of the oppositely charged atmosphere of gegenions surrounding the micelle is much more appreciable than that experienced by simple ions.

(*c*) *Reduction of Net Charge on Micelle.* Since the micellar surface has a high charge density some gegenions adhere to the micelle, thus lowering its net charge.

Under normal conditions the change in electrical conductivity with concentration is represented in Fig. 5.7 (i), since effects (*b*) and (*c*) above, which cause a decrease in the molar conductivity (Λ) on micelle formation, outweigh effect (*a*). However, at high field strengths the micelles move so rapidly that the ionic atmosphere of gegenions cannot reform quickly enough (Wein effect) and the 'braking effect' is therefore reduced. In addition, fewer gegenions are attached to the micelle and the net charge on the latter is increased. Under these conditions, effect (*a*) predominates and the molar conductivity (Λ) shows an increase on micelle formation as shown at (ii) in Fig. 5.7.

Solubility; the Krafft Point

At low temperatures the solubilities of surface active agents show a small increase with rising temperature until a particular temperature is reached when sufficient material is present in solution to allow the formation of micelles. Beyond this temperature, which is known as the Krafft Point, there is a rapid increase in solubility. This behaviour is explained by the fact that the micelles are more soluble than the monomers. The concentration in solution at the Krafft Point is the CMC at the Krafft temperature. The usefulness of this behaviour in the determination of CMCs is limited by the fact that it is difficult to calculate from the results the CMC at other temperatures.

Light Scattering

The scattering of light by solutions of surface active agents is increased by the aggregation of molecules into micelles. The slopes of graphs of amount of light scattered versus concentration therefore show an abrupt increase at the CMC. In addition to the CMC determination this method allows the aggregation number, i.e. the number of monomers per micelle, to be calculated, and provides information on the shape of the micelle.

Solubilisation

The property of surface active agents to cause an increase in the solubility of organic compounds in aqueous systems is called solubilisation. This property is observed only at and above the CMC, thus indicating that the micelles are involved in the phenomenon. In general, the increased solubility of the solubilised material (solubilisate) can be explained in terms of partition between the aqueous

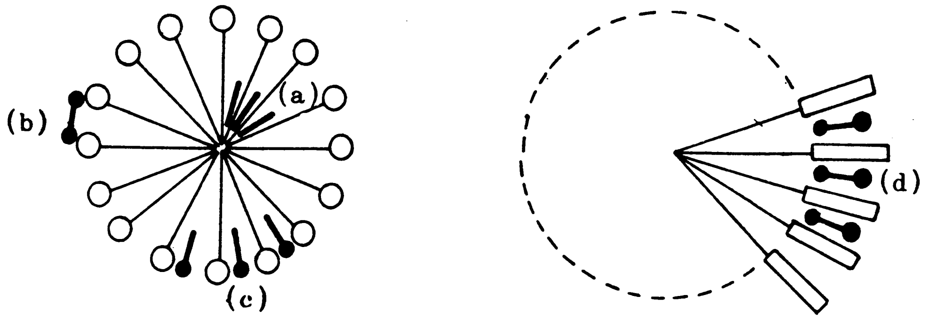


Fig. 5.8 Solubilisation within micelles: (*a*) non-polar hydrocarbon (*b*) polar molecule adsorbed at micelle-water interface (*c*) amphipathic molecules orientated in palisade layer (*d*) polar molecules orientated between polyoxyethylene chains of a non-ionic surface active agent

phase and the hydrocarbon interior of the micelles or by adsorption at the micellar surface. Thus, it is believed that non-polar hydrocarbons are taken into solution in the interior of the micelles, and polar water soluble compounds such as sugar and glycerol are adsorbed at the micelle/water interface, as shown at (*a*) and (*b*) respectively in Fig. 5.8. Compounds with amphipathic character, such as octanol and phenols, are believed to become orientated in the palisade layer in a similar manner to the surface active agent, i.e. with their polar groups directed towards the aqueous phase and their lipophilic groups inside the micelle, as shown at (*c*) in Fig. 5.8. A possible additional mechanism for the effect of non-ionic micelles on the solubility of certain compounds, e.g. phenol, involves inclusion of these compounds between the long hydrophilic polyoxyethylene chains, as shown at (*d*) in Fig. 5.8.

CMCs may be determined by solubilisation measurements preferably using a solid dyestuff that is virtually insoluble in water. The amount of dyestuff in solution remains reasonably constant until the CMC of the surface active agent is reached, and then increases rapidly. Dyes are commonly used in such determinations because of the ease of analysis of the solutions by optical methods. The disadvantages of this method include the long periods of stirring necessary to ensure that equilibrium conditions have been reached, and the fact that the CMCs are affected by the presence of the solubilisate so that the observed values are consequently lower than those obtained by other methods.

Other Methods of Determining CMCs

Refractive index, diffusion rate, viscosity, density, u.v. and i.r. spectrophotometric measurements may also be used in the determination of CMCs. In addition, a method based on the changes in absorption spectra of various compounds on their adsorption by micelles is frequently used (dye absorption method). Pinacoyl chloride and rhodamine G may be used for anionic surface active agents, eosin and fluorescein for cationics, and pinacoyl chloride or iodine for non-ionics. This method suffers from the disadvantage that the CMC may be affected by the presence of the additive.

FACTORS AFFECTING CRITICAL MICELLE CONCENTRATIONS

A. Molecular Structure of the Surface Active Agent

1. *The hydrocarbon chain*

(*a*) *Chain Length.* An increase in the hydrocarbon chain length causes a logarithmic decrease in the CMC at constant temperature as shown by Eqn (5.1)

$$\log C = A - Bm \qquad (5.1)$$

where C is the CMC, m is the number of carbon atoms in the chain, and A and B are constants for a homologous series of compounds.

(*b*) *Branched Hydrocarbon Chains.* Branching of a hydrocarbon chain causes an increase in CMC since the decrease in free energy arising from the aggregation of branched chain molecules is less than that obtained with linear molecules with the same number of carbon atoms.

(*c*) *Unsaturation.* The CMC is increased by about three to four times by the presence of one double bond when compared with the value for the analogous saturated compound.

2. *The hydrophilic group*

(*a*) *Type of Hydrophilic Group.* The value of the constant A in Eqn (5.1) varies with the type of hydrophilic group common to each homologous series. However, the effect of different ionic groups on the CMC is small, provided complete ionic dissociation occurs, because the amount of work necessary to overcome the electrical repulsion between ions of the same charge is similar.

Since the effects of electrical repulsion are absent in non-ionic surface active agents aggregation is facilitated and the CMCs are much lower than those of ionic surface active agents.

(*b*) *Number of Hydrophilic Groups.* The electrical repulsive force between adjacent ions in a micelle increases as the number of ionic groups increases. In addition, an increase in the number of any type of hydrophilic group increases the solubility of the surface active agent. Both these effects will lead to an increase in the CMC.

(*c*) *Position of Hydrophilic Group.* The CMC tends to increase as the polar group is moved from the terminal position towards the middle of the hydrocarbon chain.

B. *Effect of Additives*

1. *Simple electrolytes*

The CMC decreases on addition of salts. The most important factors concerned in the overall effect are the concentration and number of charges on the ions of opposite charge (gegenions) to that carried by the micelle. The effect of these factors is given by Eqn (5.2), where C and C_i are the CMC and

$$\ln C = -\mathrm{K} \, . \ln C_i + \text{constant} \qquad (5.2)$$

total gegenion concentration, respectively, and K is a constant with a value of approximately 0·4 to 0·6. For surface active agents with two ionic groups, Eqn (5.2) becomes

$$\ln C = -2\mathrm{K} \, . \ln C_i + \text{constant} \qquad (5.3)$$

2. *Other Surface Active Agents*

The CMCs of mixtures of surface active agents appear to vary between the limiting values of the highest and lowest CMCs of the individual components.

3. *Alcohols*

The CMCs are decreased by the addition of alcohols. The marked effect of alcohols probably arises from their high selective adsorptivity at the micellar surface and penetration into the palisade layer. The overall effect is dependent on the concentration of the alcohol, the length of its hydrocarbon chain, and the length of the hydrocarbon chain of the surface active agent. In alcohols with shorter chain lengths than the surface active agent the effect is given quantitatively by Eqn (5.4)

$$\ln\left(-\mathrm{d}C/\mathrm{d}C_a\right) = -0{\cdot}69m + 1{\cdot}1m_a + \text{constant} \qquad (5.4)$$

where, C and C_a are the CMC and the alcohol concentration, respectively, and m and m_a are the number of carbon atoms in the surface active agent and alcohol respectively.

4. *Hydrocarbons*

Solubilisation of hydrocarbons causes an increase in micellar size, which results in an increase in the radius of curvature of the micellar surface. This increase in curvature may cause a slight separation of adjacent ions and, therefore, a decrease in the repulsive forces. In addition, the presence of the solubilised hydrocarbon may allow a greater decrease in the surface free energy of the lipophilic part of the surface active agent on micellisation. Consequently, micelle formation is facilitated and the CMC decreases.

Pharmaceutical Applications and Medical Importance of Surface Active Agents

Some of the more general applications of surface active agents, e.g. as emulsifying agents, detergents, wetting agents, etc., have already been mentioned in Chapter 4. The type of application of a particular compound is usually indicated by the balance between its hydrophilic and lipophilic properties. A numerical value (HLB value) can be assigned to this balance from a knowledge of the structural formula of the compound. Thus a detergent and a w/o emulsifying agent would normally possess HLB values of 13–16 and 3–6 respectively. The following list is intended to give more specific examples of the applications and implications of these compounds in pharmacy and related fields.

A. PHYSIOLOGICAL EFFECTS OF SURFACE ACTIVE AGENTS

1. *On Micro-organisms*

Many surface active agents, especially the quaternary ammonium compounds, have useful antibacterial properties, and preparations containing these compounds are widely used; e.g. disinfectants for instruments and skin, antibacterial creams, and throat lozenges.

This activity of surface active agents is thought to arise from their adsorption at the cell surface, with consequent change in the permeability of the cell membrane and eventual death through loss of essential substances from the cell.

2. *On Removal of Bronchial Mucus from the Respiratory Tract*

In various acute and chronic infections of the respiratory tract such as bronchitis, asthma and tuberculosis, there is an increase in the amount and viscosity of bronchial mucus. Under certain conditions the mucus tends to dry out and form hard patches which prevent ciliary movement. The inhalation of surface active agents as sprays or mists by aerosol therapy has been shown to be useful in the treatment of these conditions. The action of the surface active agents is probably connected with their ability to promote wetting and, therefore, softening of the hardened mucus, thus facilitating its removal.

3. *On Human Skin*

It is well known that repeated contact between the skin and certain detergent solutions may cause disorders ranging from mild irritation and 'dry skin' to blisters and pustules. These disorders usually arise from a combination of various effects such as defatting of the skin, change in pH of the skin surface, increased swelling of the skin, and adsorption onto the skin of the surface active agent. These effects may facilitate the onset of infection.

B. THE USE OF SURFACE ACTIVE AGENTS IN PHARMACEUTICAL FORMULATION

4. *As Solubilising Agents*

(*a*) *Disinfectant Solutions.* The apparent water solubility of phenolic compounds such as Cresol BP and Chloroxylenol BP is increased in the presence of an alkali soap. This fact is made use of in the formulation of Lysol BP and Solution of Chloroxylenol BP.

The disinfectant properties of the phenolic compounds are also increased in the presence of these soaps, probably because the surface active agent alters the permeability of the cell membranes of the micro-organisms. However, it must be borne in mind that the presence of an excessive amount of surface active agent may reduce the activity of a given concentration of phenol since there will be a competition for the latter between the micelles and the surface of the micro-organism (*see* p. 338).

(*b*) *Vitamin Preparations.* Many people find that oil-soluble vitamins are unpleasant to take in the form of fish liver oils or as concentrated solutions in oils. This problem may be overcome by the use of oil-in-water emulsions or aqueous solutions of solubilised vitamins. The solubilised systems do not suffer from the instabilities of emulsified preparations. In addition, it has been found that vitamins, such as vitamin A, are more resistant to oxidation in solubilised systems than in emulsions or oily solutions.

(*c*) *Examples of Other Solubilised Systems.* Surface active agents have been used to increase the water solubility of phenobarbitone, volatile oils, chloroform, iodine, hormones, dyes, and sulphonamides.

2. *As Wetting Agents and Deflocculating Agents*

The dispersion of hydrophobic powders in aqueous vehicles is difficult since such powders tend either to float on the water surface or to form large floccules. In the latter case the apparent viscosity of the preparation is increased and homogeneity, even after prolonged shaking, may not be attained. Such properties are often troublesome and may, for example, hinder the pourability of a suspension, the spreading of a lotion, or the withdrawal of the correct dose from an injectable suspension. These problems associated with dispersion may be overcome by the use of a surface active agent which is adsorbed at the solid/liquid interface in such a manner as to increase the affinity of the particles for the surrounding medium and reduce the interparticle attractive forces. These effects are discussed more fully in Chapter 7.

3. *As Flocculating Agents*

A controlled amount of flocculation is often desirable in the formulation of suspensions in order to obtain the required rheological properties and optimum stability. The aggregation of dispersed particles into floccules may be brought about by the use of a suitable surface active agent. This effect is also discussed more fully in Chapter 7.

4. *As Emulsifying Agents*

Synthetic and naturally occurring surface active materials are widely used as emulsifying agents. An account of the compounds commonly used, together with the mechanism of their action, is given later in this chapter.

5. *As Additives to Ointment and Suppository Bases*

The inclusion of surface active agents into the fatty bases used for ointments and suppositories may cause the following effects.

(*a*) The rate of release of active medicament from these preparations may be accelerated or retarded. The exact mechanisms responsible for these effects have not been clarified. However, contributory effects to an accelerated release rate are thought to include the absorption of water by the base from its aqueous environment and the formation of an emulsion at the base/environment interface. Such emulsification would increase the interfacial area and therefore allow a more rapid exchange of medicament between the two phases. A retardation in a release rate may occur if an increase in the hydrophilic character of a base, brought about by the inclusion of a surface active agent, has an adverse effect on the rate of partition of the medicament between the base and its aqueous surroundings.

(*b*) The capacity of the base to take up aqueous liquids may be improved. There are three official ointments that will take up appreciable amounts of water to form oil-in-water emulsions; i.e. Emulsifying Ointment BP, Cetomacrogol Emulsifying Ointment BPC, and Cetrimide Emulsifying Ointment BPC. Such water-miscible ointments are useful where either miscibility with skin secretions or subsequent removal of the ointment by washing is required.

Many of the commercially available suppository bases contain varying amounts of surface active agents such as glycerol monostearate plus traces of soaps, so that a range of bases having different emulsifying properties and release rates is available.

Gels

In certain concentrations dispersions of lyophilic colloids form semi-solid masses, particularly when the solubility of the colloidal material is reduced by change in temperature. These semi-solids are known as gels. The setting of solutions of gelatin and agar on cooling are well-known examples of gel formation. Other methods for the production of gels are available and these include: (*a*) flocculation of lyophilic colloids by salts or precipitating liquids, (*b*) evaporation of certain colloidal solutions (e.g. collodion, which is a solution of nitrocellulose in alcohol and ether, forms a nitrocellulose gel when the solvents are removed by evaporation), (*c*) chemical reactions that lead to a change in shape of the lyophilic molecules (e.g. the denaturation of albumen on heating involves some uncoiling of the protein molecules and a gel structure results), and (*d*) swelling of a dry colloid (xerogel) when placed in contact with a suitable liquid (e.g. starch granules added to water).

The term gel represents a physical state with properties intermediate between those of solids and liquids. However, it is often wrongly used to describe any fluid system that exhibits some degree of rigidity. It is therefore recommended that the term should be restricted to those systems that satisfy the following criteria, which are similar to those suggested by Hermans (1949):

1. they are coherent colloidal systems of at least two components (the gelling agent and a fluid component);
2. they exhibit mechanical properties characteristic of the solid state;
3. each component is continuous throughout the system.

These criteria are fulfilled by gelatin, agar, and bentonite gels, in which the fluid component is aqueous, and by soft paraffin, in which the fluid component is an organic liquid.

Xerogels

Many gels shrink if the fluid component is removed; e.g. by evaporation or freeze-drying. The remaining solid, which will swell and reform the gel on contact with fresh fluid component, is termed a xerogel. Gelatin sheets, acacia tears, tragacanth strips, starch grains, and leather are all xerogels.

The Structure of Gels

The rigidity of a gel arises from the presence of a network formed by the interlinking of particles of the gelling agent. The nature of the particles and the type of force that is responsible for the linkages determines the structure of the network and the properties of the gel.

The individual particles of a hydrophilic colloid may consist of either spherical or anisometric aggregates of small molecules, or single macromolecules. Possible arrangements of such particles in a gel network are shown in Fig. 5.9. In linear macromolecules the network is comprised of entangled molecules, the points of contact between which may either be relatively small or consist of several molecules aligned in a crystalline order, as shown in Fig. 5.9(*c*) and (*d*), respectively.

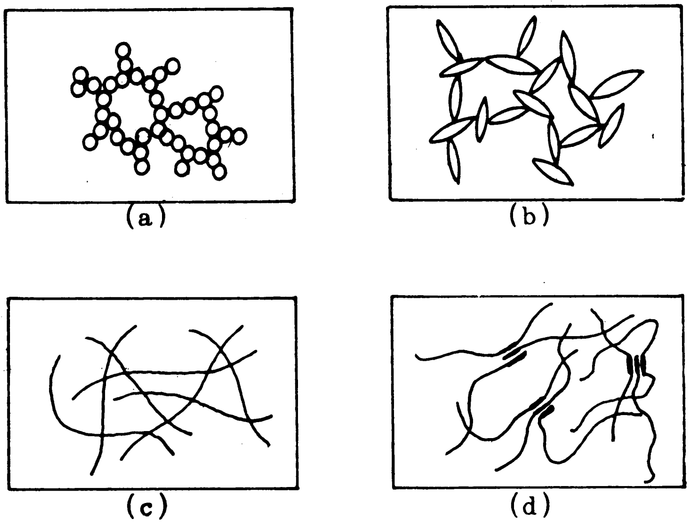


Fig. 5.9 Arrangements of particles in gel structure

The cross-linking of macromolecules by primary valency bonds provides a further mechanism for the formation of a gel network. This behaviour is exhibited by silicic acid gel, which consists of a three-dimensional network of Si—O bonds.

The forces of attraction responsible for the linkages between gelling agent particles may range from strong primary valencies, as in silicic acid gels, to weaker hydrogen bonds and van der Waals forces. The weaker nature of these latter forces is indicated by the fact that a slight increase in temperature often causes liquefaction of a gel. Systems that exhibit this type of transition, such as agar and gelatin gels, are termed thermal gels. In addition, the transition from gel state to a colloidal dispersion may in some cases be brought about by mechanical agitation. Systems, such as bentonite and aluminium hydroxide gels, that exhibit this type of transition are termed thixotropic gels.

The Properties of Gels

1. SWELLING

If a xerogel is placed in contact with a liquid that solvates it then an appreciable amount of the liquid is often taken up and the volume of the xerogel increases. This is referred to as swelling, and the pressure that develops if the xerogel is confined is known as the swelling pressure.

The degree of swelling depends on the number of linkages between individual molecules of the gelling agent and on the strength of these linkages. For example, in a protein xerogel, swelling is at a minimum at the iso-electric point of the protein because the intermolecular linkages are strongest at this point and, therefore, offer more resistance to swelling. In addition, the presence of ions in the swelling liquid influences the degree of swelling of protein xerogels. Sulphate ions, for example, increase the resistance to swelling by forming additional linkages, known as salt bridges, between the protein molecules, whereas the peptising effects of iodide ions produce an increased swelling.

If the gel network consists of primary valency bonds, as in silicic acid gel, the strength of these bonds is sufficient to prevent the occurrence of swelling and only relatively small amounts of liquids are taken up by such systems.

2. SYNERESIS

Gels will often contract spontaneously and exude some of the fluid medium. This effect is known as syneresis and the degree to which it occurs usually decreases as the concentration of gelling agent increases. The occurrence of syneresis indicates that the original gel was thermodynamically unstable. Although syneresis occurs spontaneously it may take place extremely slowly so that the absence of separation of fluid component should not be taken as an indication that the gel is thermodynamically stable.

Blood clot is a common example of a system that exhibits syneresis.

3. AGEING

Colloidal systems usually exhibit slow spontaneous aggregation. This is referred to as ageing. In gels, ageing results in the gradual formation of a denser network of gelling agent. Theimer (1960) suggests that this process is similar to the original gelling process and continues after the initial gelation, since fluid medium is lost from the newly formed gel. He observed the formation of additional thin fibrils in 10-day-old gelatin gels when compared with the fresh gels but found little further change in structure after 21 days since the dehydration processes were slowing down.

4. ADSORPTION OF VAPOURS BY XEROGELS

The porous nature of xerogels provides a large surface area for the adsorption of vapours. In addition, the porosity provides the possibility of uptake of vapour by capillary condensation (*see* p. 46). Xerogels, especially silica gel, are therefore used as drying agents.

The isotherms for adsorption and desorption in a particular system of this type often do not coincide at higher vapour pressures. A hysteresis loop is therefore observed between the curves that represent adsorption and desorption in graphs showing the variation in the amount of vapour taken up at different equilibrium pressures. This behaviour is attributable to the differences that arise when the pores in the xerogel are filled with liquid during adsorption and when they are subsequently emptied during desorption.

5. RHEOLOGICAL PROPERTIES

Gels exhibit the mechanical properties of rigidity, tensile strength, and elasticity that are characteristic of solids. In thixotropic gels these effects are only apparent below the yield value, above which the systems exhibit the flow properties of suspensions.

6. CHEMICAL REACTIONS IN GELS

If the components of a gel are inert towards substances involved in a chemical reaction then the gel forms a suitable medium in which the reaction can be studied since the network provides some protection against disturbances caused by mechanical influences or convection currents. The precipitates produced from such reactions often have a periodic structure; i.e. the precipitate is formed in a rhythmic pattern throughout the gel. Such patterns are often referred to as Leisegang rings. The precipitation of silver chromate that results from the reaction of potassium chromate and silver nitrate in gelatin gel is a common example of this type of behaviour and it is suggested that the layered structure of gallstones offers a further example (Jirgensons, 1958).

7. DIFFUSION IN GELS

Since a gel may be regarded as a random network permeated by pores that are filled with a liquid component, substances that are soluble in the liquid component will tend to permeate through the gel by diffusion in solution through the spaces in the network. The rate of diffusion of substances through gels by this means will therefore be affected by those factors that normally affect simple diffusion in solution and by additional factors that are associated with the presence of the gel network.

(*a*) *Diffusion in Solution*

This process may be defined as the spontaneous transference of solute from regions in the solution where the concentration is high to other regions where the concentration is lower, until there is a uniform distribution throughout. The rate of diffusion of a solute is expressed by Fick's first law (Eqn 5.5),

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -DA\frac{\mathrm{d}c}{\mathrm{d}x} \tag{5.5}$$

where dm is the amount of solute diffusing in time dt across an area A under the influence of a concentration gradient dc/dx. D is known as the diffusion coefficient and it has the dimensions of area per unit time. It is not strictly constant but varies with concentration at constant temperature. The value of D obtained from any measurements should therefore be regarded as a mean value for the concentration range that is considered. The negative sign in Eqn (5.5) is necessary because diffusion occurs in the opposite direction to that of increasing concentration (i.e. dc/dx is negative).

The diffusion coefficient for spherical particles that are of colloidal dimensions (i.e. considerably larger than the solvent molecules) is given by Eqn (5.6), which is known as the Sutherland-Einstein equation.

$$D = \frac{RT}{6\eta r N_{\mathrm{A}}} \tag{5.6}$$

where r is the radius of the spherical particle, η is the viscosity of the liquid medium, R is the gas constant, T is the thermodynamic temperature, and N_{A} is Avogadro's constant.

Equation (5.6) does not apply to non-spherical particles because the frictional force that opposes their movement will vary with the orientation of the particles.

The rates of diffusion of small molecules and ions are greater than those of colloidal particles. Although the diffusion of the former is impeded by the viscosity of the medium, Eqn (5.6) is not applicable to these systems.

The difference in the rates of diffusion of small and large particles tends to cause the separation of charged colloidal particles from their surrounding atmospheres of gegenions. This effect gives rise to a potential difference that resists further separation and may cause an acceleration or a retardation in the diffusion rate of the colloidal particle depending on the conditions of the experiment. The effect can be suppressed by the addition of sufficient simple electrolyte (e.g. KCl), which provides an excess of gegenions throughout the liquid medium and so prevents the development of potential gradients. However, it is necessary to ensure that the diffusion of this electrolyte does not affect the diffusion of the colloidal material. This is usually achieved by following the diffusion of colloid across a boundary on both sides of which the concentration of simple electrolyte is the same.

(*b*) *The Effects Associated with the Presence of a Gel Network*

Sieve Effects. The rates of diffusion of small molecules and ions through gels are virtually the same as their rates of diffusion in simple solution. However, as the size of the diffusing particles becomes comparable with the pore diameter of the gel network diffusion is retarded considerably and ceases when the particles are larger than the widest pores in the gel. This is known as the sieve action of the gel and it may be regarded as a normal filtration process.

The sieve effect of a gel is determined by the average pore diameter. This, in turn, is affected by: (*a*) the concentration of gelling agent (an increase in this concentration usually results in a decrease in the pore diameter), and (*b*) by the age of the gel and the conditions under which it has been stored, because an aged gel is usually denser than a freshly prepared one, and the ageing process may be accelerated by conditions that cause loss of the liquid component.

A further consequence of the porous nature of the medium through which diffusion is occurring arises from the fact that the area over which diffusion can take place is not the total surface area of the gel but only the combined surface area of the pores.

Electrical Effects. An electrical double layer, similar to that which surrounds a colloidal particle, will exist at the interface between the walls of the pores in a gel and the liquid that fills these pores. If the capillary walls possess fixed negative charges then the diffuse part of the electrical double layer that is associated with these walls will contain an excess of positive gegenions. However, the diffuse layers extend into the liquid that fills the pores, and flow of liquid through the gel will tend to displace the gegenions; i.e. the latter are mobile. The effect of gegenion mobility is particularly noticeable when diffusion of an electrolyte through gel occurs. The electrolyte ions that possess a similar charge to the mobile gegenions of the gel are able to enter the pores easily and the mobile gegenions are displaced along the pores. However, the charged framework of the gel will tend to prevent the entry of the other electrolyte ions since they both carry charges with the same sign. This effect gives rise to a selective permeability of the gel for either positive or negative ions. For example, gelatin gel is selectively cation permeable in alkaline media since the gel network then possesses fixed negative charges. Conversely, when the network becomes positively charged as in acid media, the gel is selectively anion permeable.

These effects also apply to the permeability of membranes; further information is given on page 73.

Other Effects. Adsorption of a diffusing solute onto the walls of the pores will retard the rate of diffusion. If a mixture of solutes is passing through the gel then any difference in their affinities for the adsorption sites on the walls may lead to a type of chromatographic separation.

A solute may become distributed between the network phase of a gel and the liquid phase that fills the pores. Diffusion through a gel may therefore proceed by diffusion in solution through the network itself.

Although the viscosity of the medium in which diffusion is occurring will influence the rate of diffusion it should be remembered that in diffusion in gels it is usually the viscosity of the liquid component in the pores that is important in this respect, not the overall viscosity of the gel.

Application of Gels

1. DOSAGE FORMS

Glycogelatin gels are frequently used as a basis for medicated pastilles. They are also used in the formulation of some suppositories, e.g. Glycerin suppositories BP (1968).

Gelatin gels are employed in the preparation of hard and soft capsules that may be used to mask the unpleasant tastes of solids and liquids.

2. MICROBIOLOGICAL MEDIA

Agar and gelatin gels are used as solid media for the culture of micro-organisms (p. 317). The diffusion of antibiotics, antiseptics, vitamins, and enzymes through solid culture media is used in microbiological assays of these materials. Such diffusion produces zones of either retarded or enhanced growth on seeded agar plates depending on the activity of the diffusing substances. The factors involved in this type of assay for antiseptics have been investigated by Cooper and Woodman (1946). These workers determined the diffusion coefficients of crystal violet and penicillin through agar gels by employing an equation, Eqn (5.7) that had been given previously by Eversole and Doughty (1935) for the diffusion of neutral molecules through pure water. If the initial concentration of a solution is m_0 and the concentration at a distance x from the junction of the gel and solution after a time t is m_1 then the diffusion coefficient of the solute is given by

$$4D\,2{\cdot}303(\log m_0 - \log m_1) = x^2/t \qquad (5.7)$$

A graph of x^2 against t is usually plotted, and the diffusion coefficient is then calculated from Eqn (5.7) by inserting the value of x^2/t obtained from the slope of such a graph. This equation applies to the diffusion of neutral molecules only and must be amended for the diffusion of charged particles to allow for the influence of potential gradients. However, the presence of sufficient simple electrolyte prevents the development of such gradients, and Eqn (5.7) may then be applied. In addition, Eqn (5.7) was derived for the case of linear diffusion and is not strictly applicable to the radial diffusion that occurs in most cup plate assay techniques. However, the error involved is small if the cups are of large diameter. A consideration of the theoretical aspects of diffusion assays is given by Cooper (1963). Further details of this type of assay are given later in this book (p. 349).

3. GEL FILTRATION

This is a technique for the separation of solutes of different molecular sizes. It is based on the fact that small molecules are able to penetrate into the networks of gels whereas larger molecules are unable to do so. Gels consisting of cross-linked polysaccharide chains are commonly used, e.g. Sephadex (Pharmacia), and chromatographic techniques are employed in the separation process. For example, a solution containing the mixed solutes is passed through a column packed with gel particles and solvent. The largest solute molecules pass between the particles and through the column at a relatively rapid rate, whereas the rate of passage of the smaller solute molecules is reduced by their penetration into the pores of the gel particles.

Dialysing Membranes

The membranes that are used in dialysis and ultrafiltration are usually derivatives of cellulose (e.g. cellulose acetate, nitrocellulose) and may be classed as xerogels. The structure of these materials may be regarded as a microporous network of macromolecules, as indicated by Fig. 5.10, which shows that the molecules are usually aggregated into bundles or micelles by secondary forces such as hydrogen bonding.

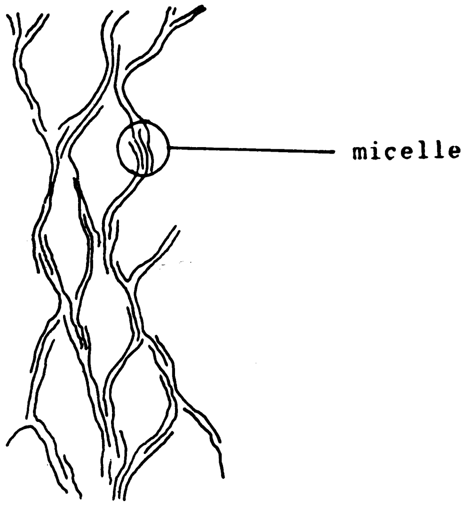


Fig. 5.10 The structure of dialysing membranes

Preparation of Membranes

Suitable membranes should prevent the escape of colloidal material but allow rapid diffusion of solvent and small molecules. They are usually prepared by evaporation of layers of a solution of the membrane material in a volatile solvent (e.g. ether, acetone). The solution is cast on to either a rotating thimble-shaped mould or a flat surface. Several layers are applied and each is allowed to evaporate under controlled conditions of humidity and temperature before the next one is added. The membrane is finally hardened by immersion in a liquid in which it is insoluble.

If the original solvent is non-volatile the membrane material may be deposited as a film in a support such as filter paper.

Permeability of Membranes

The permeability or permeability coefficient (P) of a membrane is given by Eqn (5.8)

$$\frac{Q}{t} = \frac{PA\,\Delta c}{d} \qquad (5.8)$$

where Q is the amount of material that passes through a membrane of area A and thickness d under the influence of a concentration difference Δc in time t.

FACTORS AFFECTING PERMEABILITY

Since the structure of a membrane resembles that of a gel the factors that affect the permeability of membranes are similar to those that affect the rate of diffusion of substances through gels (p. 70).

Pore Size

This is important in determining the contribution of the sieve effect to the permeability of a membrane. The average pore size may be determined by the rate of penetration of water through the membrane, or by the filtration of standardised colloidal particles of varying sizes. For example, colloidal gold particles, the size of which had been previously determined in the electron microscope, were used by Zsigmondy. It is also useful to know the degree of uniformity of pore size. This is often expressed as the ratio of the maximum pore size, which is determined by measuring the minimum pressure required to produce a stream of air bubbles through the membrane, to the average pore diameter, which is determined from the rate of flow of water as indicated previously.

Satisfactory membranes should possess pores that fall within a narrow range of sizes. Preparative methods therefore attempt to produce membranes of different grades with reproducible characteristics. This involves careful control of all factors that may influence pore size during the preparation of membranes. For example, the temperature and humidity under which evaporation of the volatile solvent is carried out are controlled since the final pore size depends on the extent of drying; the shorter the drying time the greater the pore size. In addition, solvent mixtures containing gelifying liquids as well as normal solvents are used to preserve the regularity of a gel during the hardening process. These liquids cause swelling of the gel material as opposed to complete solution. They are not removed by the evaporation but dissolve out of the membrane into the hardening liquid, thus preserving the porosity of the initial gel in the final membrane. If only solvent liquids are used the resulting membranes are very compact and relatively impermeable.

The pore size may also be affected by various treatments after the membranes are formed. For example, stretching by means of hydrostatic pressure, partial digestion by cellulose enzymes or treatment with zinc chloride solution will all increase the permeability of cellophane membranes, while longitudinal stretching or acetylation cause a decrease in permeability (Craig and Konisberg, 1961).

Electrical Effects

(*a*) *Charges on Membranes.* These will influence the permeability of membranes to anions and cations in the same way as they affect diffusion through gels. Thus, a membrane such as collodion, which possesses fixed anions and mobile cations, will be selectively permeable to cations, since these are able to displace the original cations through the pores, anions being repelled by the negatively charged membrane material. Conversely, a membrane with fixed cations and mobile anions will be selectively permeable to anions. The separation of charge that results from such selectivity produces a difference in electrical potential between the opposite sides of the membrane.

(*b*) *The Donnan Membrane Effect.* The diffusion of small ions through a membrane will be affected by the presence of a charged macromolecule that is unable to penetrate the membrane because of its size. This is known as the Donnan membrane effect and it may be explained by considering a system consisting of a solution of sodium chloride separated by a membrane from another solution containing a charged macromolecule (R^-) and its gegenion (Na^+). If the initial concentration of sodium chloride is c_1 and that of the macromolecule is c_2 the system may be represented diagrammatically as

$$\begin{array}{lcc:cc} & Na^+ & + \; Cl^- & Na^+ & + \; R^- \\ \text{Initial concn.} & c_1 & c_1 & c_2 & c_2 \end{array}$$

In dilute solutions the product of the concentrations of the diffusible ions (i.e. Na^+ and Cl^-) will be equal on both sides of the membrane when equilibrium is established. The condition of electroneutrality must also apply; i.e. the concentrations of positively and negatively charged ions on each side of the membrane must balance. If the amount of sodium chloride that has diffused through the membrane when equilibrium has been established is x then the system may be represented as

$$\begin{array}{lcc|ccc} & \mathrm{Na^+} \; + & \mathrm{Cl^-} & \mathrm{Na^+} & + \, \mathrm{R^-} & + \, \mathrm{Cl^-} \\ \text{Equilibrium} & & & & & \\ \text{concn.} & (c_1 - x) & (c_1 - x) & (c_2 + x) & c_2 & x \end{array}$$

$$\begin{aligned} \therefore \quad \text{at equilibrium } (c_1 - x)(c_1 - x) &= (c_2 + x)x \\ \therefore \quad c_1^2 - 2c_1x + x^2 &= c_2x + x^2 \\ \therefore \quad c_1^2 &= c_2x + 2c_1x \\ &= x(c_2 + 2c_1) \\ \therefore \quad x &= \frac{c_1^2}{(c_2 + 2c_1)} \\ \therefore \quad \frac{x}{c_1} &= \frac{c_1}{(c_2 + 2c_1)} \end{aligned} \tag{5.9}$$

In Eqn (5.9) x/c_1 represents the fraction of sodium chloride that has diffused through the membrane when equilibrium has been established. It can be seen that this fraction is influenced by the concentration of the non-diffusible ion R^-. If the unequal distribution of diffusible electrolyte that occurs in the presence of a charged macromolecule is not taken into account then determination of the osmotic pressure exerted by the macromolecule will be incorrect. However, if the concentration of diffusible electrolyte is relatively high then the denominator on the right-hand side of Eqn (5.9) is approximately equal to $2c_1$ (i.e. c_2 may be ignored). The equation then reduces to $x/c_1 = \frac{1}{2}$, indicating that the simple electrolyte is equally distributed on both sides of the membrane and that the Donnan effect is negligible in such cases.

Solubility Effects. Substances may pass through membranes by diffusion in solution through the membrane material itself. This process involves an initial partition of the substance between the original solvent and the membrane material. It may be of importance when an aqueous solution of an organic solute is in contact with a membrane with marked lipoidal properties (*see* Chapter 6).

Aerosols

Aerosols are colloidal dispersions of liquids or solids in gases. In general, mists and fogs possess liquid disperse phases, while smokes, fumes, and dusts are dispersions of solid particles in gases.

Stability

Fogs and smokes appear to be very stable when compared with the aqueous dispersions of colloidal particles discussed in the preceding section. However, the concentration of particles in the latter is usually of the order of 10^{10} mm^{-3}, whereas the concentration of particles in fogs and smokes is much lower (about 10^3 mm^{-3}). Aqueous dispersions of similar low concentrations would be even more stable than the fogs and smokes, since the particles in the latter are not stabilised to any appreciable extent by electrical charges, which may arise from collisions between the particles and naturally occurring ions in air.

Preparation

In common with other colloidal dispersions aerosols may be prepared by either dispersion or condensation methods. The latter type involves the initial production of supersaturated vapour of the material that is to be dispersed. This may be achieved by supercooling the vapour. The supersaturation eventually leads to the formulation of nuclei, which grow into particles of colloidal dimensions.

The preparation of aerosols by dispersion methods is of greater interest in pharmacy and may be achieved by the use of pressurised containers. For example, the pressure inside these containers may be produced by a liquefied gas, which is known as the propellant. If a solution or a suspension of active ingredients is contained in the liquid propellant or in a mixture of this liquid and an additional solvent then when the valve on the container is opened the vapour pressure of the propellant forces the mixture out of the container. The large expansion of the propellant at room temperature and atmospheric pressure produces a dispersion of the active ingredients in air. Although the particles in such dispersions are usually larger than those in colloidal systems, the term aerosols is still generally applied to them.

Various types of product can be obtained by the use of different propellants, valves, and solvents. These products range from aerosols to sprays, which contain larger liquid particles, and to foams and semi-solid products such as toothpaste. A review of the various aspects of pressurised containers used for these different products is given by Sciarra (1970).

Applications and Advantages in Pharmacy

The advantages obtained from the use of pressurised containers, such as those used for the production of aerosols, are common to all the types of product that may be presented in this way; e.g. foams, sprays, and semi-solids. These advantages include—

(*a*) ease of use,
(*b*) protection from contamination with foreign materials, since the product is sealed inside the container (this is particularly important in the preservation of sterility),
(*c*) protection from the effects of air and moisture,
(*d*) regulation of dosage by the use of a metered valve (although such metering is no more accurate than oral dosage forms it is an advantage when used for topical preparations),
(*e*) economic usage of dosage form, when, for example, metered valves are used for expensive topical preparations which may otherwise be formulated as ointments, creams or lotions. The usage of these latter dosage forms will depend on the user's attitude to economy and often involves some loss of the product on an applicator.

The use of aerosols as a dosage form is particularly important in the administration of drugs via the respiratory system or nasal passages. In addition to local effects, systemic effects may be obtained if the drug is absorbed into the bloodstream from the lungs.

Topical preparations are also well suited for presentation as aerosols or sprays. The irritation of a sore wound caused by the rubbing in of an ointment or cream is avoided, and the cooling effect of aerosols containing liquefied gases may be advantageous.

Suspensions

Stability

As previously stated, coarse suspensions are less stable than colloidal dispersions since the larger particles in the former settle more rapidly under the influence of gravity. Nevertheless, it is possible to obtain suspensions of relatively coarse particles that will remain as dispersions for considerable periods.

The factors that determine the rate of sedimentation are indicated by Stokes's law (Eqn 5.10)

$$u = \frac{d^2(\rho_s - \rho_l)g}{18\eta} \qquad (5.10)$$

where, u is the average velocity of sedimentation of the particles, d is the mean diameter of the particles, ρ_s is the density of the solid particles, ρ_l is the density of the fluid phase, η is the viscosity of the fluid phase, and g is the acceleration due to gravity.

Stokes's law was derived for dilute suspensions of rigid uniform spheres settling at a velocity which produced no turbulence in the fluid medium. Under such conditions the particles do not interfere with each other; i.e. free settling occurs.

Equation (5.10) is not applicable to the more concentrated suspensions usually encountered in pharmaceutical practice, since the particles in these suspensions are rarely spherical and their concentrations are high enough to interfere with the sedimentation of individual particles; i.e. hindered settling occurs. However, the factors indicated by Eqn (5.10) are still important, and an increase in the mean particle size or in the difference between the densities of the solid and liquid phases will produce a faster rate of sedimentation, while an increase in the viscosity of the liquid medium will decrease the sedimentation rate.

FLOCCULATED AND DEFLOCCULATED SUSPENSIONS

Aggregation of particles will also lead to a more rapid rate of sedimentation because the size of the sedimenting units is increased. As pointed out previously, the existence of a positive free energy associated with the interface between solid particles and liquid medium will mean that the dispersion is unstable from a thermodynamic point of view. Aggregation of the particles will tend to occur spontaneously in order to reduce this interfacial free energy. However, the rate of coalescence of particles will depend on the combined effects of repulsive and attractive forces that exist between approaching particles in a similar manner to the effects that were discussed in relation to the stability of colloidal dispersions. Thus, the overlapping of electrical double layers that occurs when two particles approach each other gives rise to a repulsive force and the attractive force arises from the van der Waals interactions between all the atoms in one particle and all those of the other.

In coarse suspensions of lyophobic materials the size of the dispersed particles introduces an effect that is not apparent in colloidal dispersions. The van der Waals attractive force falls off with distance more slowly than the repulsive force so that although the latter is greater than the attractive force at some distances of separation, and thus gives rise to a repulsion barrier, the van der Waals forces may again predominate at larger distances and produce a secondary minimum in the net interaction energy curve for the approach of two particles. Since the total attractive force is approximately proportional to the size of the dispersed particles, it follows that the secondary minimum is more pronounced for

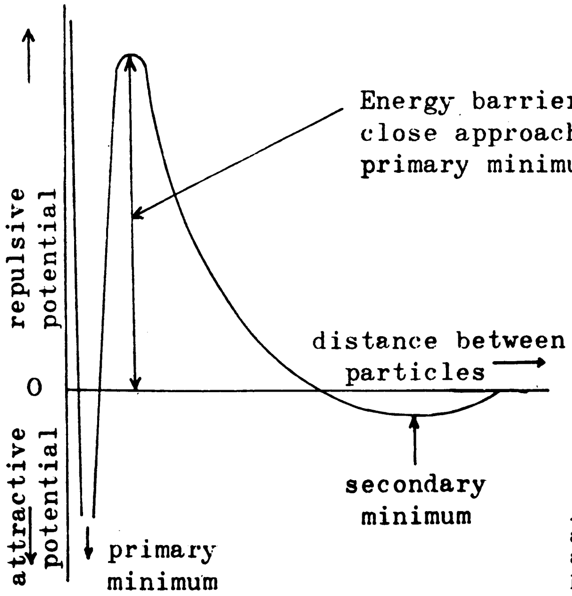


Fig. 5.11 Variation in net potential of interaction arising from the combined effects of repulsive and attractive forces between two approaching coarse particles

coarse particles than for those of colloidal dimensions. This difference between coarse and colloidal dispersions is illustrated by a comparison of Fig. 5.11, which represents the variation in net potential for the approach of coarse particles, with Fig. 5.1, which relates to colloidal systems, when it will be seen that no secondary minimum is evident in the latter figure.

The existence of two minima in the net potential energy curve for the interaction between coarse particles allows two types of coagulation to occur in these systems. If approaching particles possess sufficient energy to overcome the repulsion barrier they will come into very close contact in the deep primary minimum. In order to reseparate they will need to acquire even more energy since the barrier to redispersion is represented by the difference in levels of the primary minimum and the repulsion barrier maximum. Since this barrier to redispersion is relatively great, separation of particles coagulated in this manner does not occur easily.

The second type of coagulation may arise from the trapping of an approaching particle in the secondary minimum. Since this minimum is shallow and the particles remain separated by a thin film of liquid, the resulting coagulation is more reversible and the aggregated particles are easily redispersed by mechanical agitation. Although the DLVO theory (p. 54) predicts that a secondary minimum will exist for all lyophobic dispersions it may be too shallow to have any significant effect, since the influence of gravity on coarse particles may be greater than the attractive force exerted between particles. In addition, the shape of the particles may be important; for example, as the particles become more elongated the effect of the secondary minimum becomes significant in relation to gravitational forces.

If aggregation of the particles in a suspension does occur then the system is said to be flocculated. The nature of the flocs will be determined by the closeness of contact of the individual particles. Aggregation in the primary minimum will produce rather compact flocs, and an increase in the proportion of aggregations in the secondary minimum will increase the fluffy nature or porosity of the flocs.

If the barrier to repulsion is sufficient to prevent aggregation in the primary minimum and the effect of the secondary minimum is insignificant then the particles will remain as individual units and the system is said to be deflocculated. Solvation effects will also help to maintain the dispersion of individual particles and any lyophilic properties of the surface of the particles, e.g. caused by the adsorption of surface active agents, will tend to produce a deflocculated suspension.

The Sedimentation Behaviour of Flocculated and Deflocculated Suspensions

It has already been pointed out that the formation of aggregates will cause an increase in the rate of sedimentation of particles since the size of the sedimenting units is increased. Thus, a flocculated suspension will show a more rapid sedimentation rate than a deflocculated system. The rate of sedimentation in a flocculated system is often referred to as subsidence and it depends not only on the size of the aggregates or flocs but also on their porosity, since the liquid medium flows through, as well as around, them as they fall.

In addition to their different sedimentation rates there is a difference in the nature of the sediment that is formed by the two types of suspension. In a flocculated suspension the loose structure of the rapidly sedimenting flocs tends to be preserved in the sediment, which therefore contains an appreciable amount of entrapped liquid. The volume of the final sediment is therefore relatively large and the sediment is easily redispersed by agitation. It should be noted that the sedimentation volume of a flocculated suspension may show small decreases on storage due to the compaction of underlying flocs by the material resting upon them in the sediment.

In a deflocculated suspension the repulsive forces between individual particles allow the particles to slip past each other in a sediment. This property, together with the slow rate of sedimentation, which prevents the entrapping of liquid medium, allows the formation of a compact sediment; i.e. one with a small volume. This type of sediment is usually difficult to redisperse by agitation and is often referred to as a cake.

Pharmaceutical suspensions usually contain a reasonable distribution of particle sizes. In such systems the largest particles will sediment first and the smaller ones will fall slowly through the supernatant liquid. This is particularly noticeable in deflocculated suspensions where the particles retain their individual identities and the supernatant liquid often remains cloudy because of the presence of the smallest particles when the majority of the sediment has already formed. However, in a completely flocculated suspension even the smallest particles are involved in floc formation so that they do not remain in the supernatant, which is therefore not cloudy (Michaels and Bolger, 1964).

Figure 5.12 illustrates the difference in appearance of flocculated and deflocculated suspensions after various periods of standing in an undisturbed condition.

Rheological Properties of Suspensions

These properties are discussed in more detail in Chapter 7, where it will be seen that flocculated suspensions tend to exhibit plastic or pseudoplastic behaviour while deflocculated systems tend to be dilatant. This means that the apparent viscosity of flocculated suspensions is relatively high when the applied shearing stress is low but it decreases as the applied stress increases. In fact, if plastic behaviour is exhibited then the system behaves like a solid up to a particular shearing stress, which is known as the yield value, and no flow occurs in the system until this value is exceeded. Conversely, the apparent viscosity of deflocculated suspensions is low at low shearing stresses and increases as the applied stress increases. One of the important consequences of this difference in behaviour is that deflocculated suspensions are easy to pour, whereas flocculated ones are not, since the shearing stresses involved in pouring are relatively low. However, dilatant suspensions are often troublesome, and attention to this type of system should be paid in milling operations because the high speed involved in some of these operations may cause such an increase in viscosity that the mill may sieze up. It should be reiterated that this paragraph gives only a brief indication of the differences in the rheological properties of flocculated and deflocculated suspensions, and Chapter 7 should be consulted for explanations of these and other properties.

Pharmaceutical Applications of Suspensions

A pharmaceutical suspension consists of a dispersion of finely divided insoluble material suspended in a liquid medium. Most suspensions are prepared and stored in their final form until required for use. In other cases, where unwanted effects may be caused by prolonged contact between the solid particles and the liquid dispersion medium, the suspension is prepared immediately before use and is acceptable for only a limited period thereafter. For example, Ampicillin Mixture BPC (1968) is a suspension of ampicillin or ampicillin trihydrate in a suitable flavoured aqueous vehicle. However, the antibiotic is susceptible to hydrolysis and the activity of an aqueous suspension gradually decreases. The BPC therefore directs that the mixture should be freshly prepared by dispersing the dry mixed ingredients in Purified Water, stored in a cool place, and used within one week of preparation.

The major applications of suspensions in pharmacy can be conveniently divided into the following groups.

(*a*) *Preparations for Oral Administration*

A suspension provides a convenient means of administering an insoluble nrug, especially where difficulty may be encountered in swallowing solid dosage forms such as tablets or capsules. Solids such as kaolin, magnesium carbonate, calcium

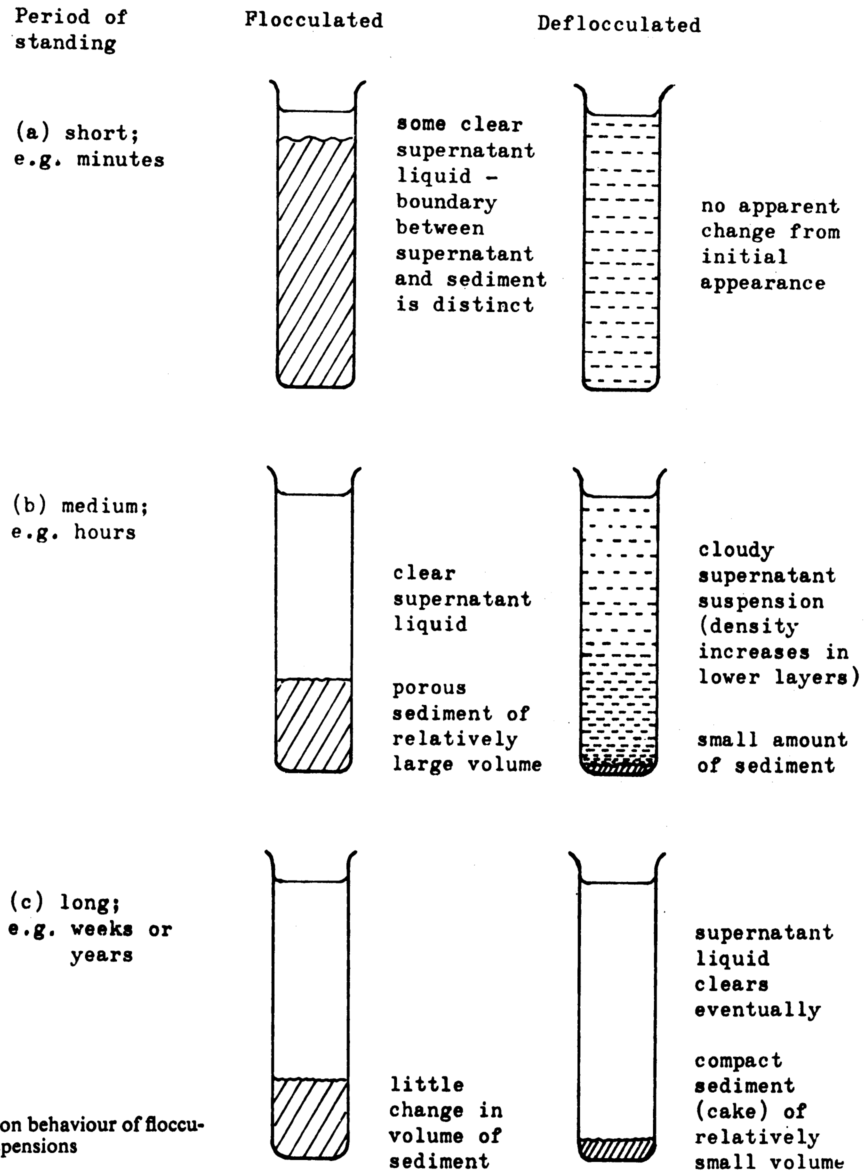


Fig. 5.12 The sedimentation behaviour of flocculated and deflocculated suspensions

carbonate, and magnesium trisilicate are often administered for their adsorptive or antacid properties in the treatment of gastric disorders. The effectiveness of these substances is influenced by the surface areas of the solids and administration as fine powders in the form of suspensions allows a large surface area to be available immediately for contact with the gastric contents.

Insoluble derivatives are often used to reduce the unpleasant taste of the active form of drugs and the formulation of these insoluble compounds in suspensions is common. For example, insoluble chloramphenicol palmitate is used in Chloramphenicol Mixture BPC (1968) to reduce the bitter taste of the free base.

(*b*) *Suspensions for Injection*

Such preparations are of particular importance in the field of depot therapy. The various types of product and the problems associated with their formulation are discussed by Gunn and Carter (1965).

(*c*) *Suspensions for External Use*

Preparations of this type that are pourable are usually referred to as lotions (e.g. calamine lotion) and those that possess semi-solid characteristics are termed pastes (e.g. zinc and salicylic acid paste and magnesium sulphate paste).

The physico-chemical problems associated with the formulation of pharmaceutical suspensions are concerned with the control of sedimentation behaviour and rheological properties, both of which are influenced by the degree of flocculation in a particular system. In addition, the sedimentation behaviour is influenced by the rheological properties of the liquid continuous phase. Because of this interdependence of properties relevant to the formulation of pharmaceutical suspensions, the problems associated with this type of formulation are discussed in the section concerned with the rheology of suspensions (Chapter 7).

Emulsions

An emulsion is a dispersion of a liquid as globules in another liquid that is immiscible with the first. The diameters of the globules usually vary between about $0{\cdot}1 \times 10^{-3}$ to 100×10^{-3} mm. One of the liquids is usually water and the other is an oil. Two types of emulsion are possible; i.e. oil-in-water (o/w) and water-in-oil (w/o) and examples of both types are frequently encountered in pharmaceutical and cosmetic preparations. A third component is required to stabilise the emulsion because a simple dispersion of oil in water (or water in oil) obtained by vigorous shaking of the two liquids is unstable and coalescence of the globules and eventual separation into two phases occurs rapidly. This additional component is termed the emulsifying agent, emulgent, or emulsifier.

Emulsifying Agents

Various types of agent are used as indicated by the following list.

1. SYNTHETIC SURFACE ACTIVE AGENTS

A classification of these has already been given in Chapter 4. This type of agent is the most common in modern emulsion technology and many commercial products are available. Surface active agents may be used singly or in combination with other surface active agents for the purpose of stabilising an emulsion. For w/o emulsions the balance between the hydrophilic and lipophilic properties of the emulsifier should be such as to give an HLB value in the range 3 to 6 (*see* p. 66) and for an o/w emulsion the HLB value of the emulsifier should be in the range 8 to 13. For a particular emulsion the required HLB value of the emulsifier may be calculated more precisely and many surface active agents have been synthesised to satisfy the requirements of specific systems.

2. MACROMOLECULAR EMULSIFYING AGENTS

The majority of these agents produce o/w emulsions. In addition, many of them are obtained from natural sources and therefore may vary in quality.

(*a*) *Gums*

Acacia is the most common gum used in pharmaceutical preparations. It is particularly useful for emulsions prepared with a mortar and pestle, provided that the correct proportions of oil, water, and gum are employed in the preparation of a so-called primary emulsion that may then be diluted with more aqueous phase. These proportions depend on the nature of the oil phase and more information on the preparation of emulsions by this means is given by Gunn and Carter (1965).

Tragacanth gum is often included in emulsion formulations but is not a very efficient agent when used alone. It is usually used as an additional agent in emulsions containing acacia as the main emulsifier.

(*b*) *Carbohydrates and Derivatives*

Starch is a poor emulsifying agent but is sometimes used in the preparation of enemas containing oils. Cellulose derivatives such as methylcellulose and sodium carboxymethylcellulose, and soluble salts of alginic acid are often used.

(*c*) *Proteins*

Certain proteins such as gelatin and casein are occasionally used in pharmaceutical emulsions.

3. FINELY DIVIDED SOLIDS

Solid particles that are wetted by oils and water may be adsorbed around the globules in an emulsion, and provide stability against coalescence. Examples of solid emulsifying agents used in pharmacy include aluminium hydroxide and bentonite, both of which yield o/w emulsions.

Determination of Emulsion Type

Several tests are available to distinguish between o/w and w/o emulsions. The most common of these involve: (*a*) miscibility tests with oil or water—the emulsion will only be miscible with liquids that are miscible with its continuous phase; (*b*) conductivity measurements—systems with aqueous continuous phases will readily conduct electricity whereas systems with oily continuous phases will not; and (*c*) staining tests—the oil is stained by the incorporation of a dye that does not partition into the aqueous phase and facilitates microscopic examination of the emulsion and determination of its type. Details of these tests are given by Gunn and Carter (1965).

Types of Instability in Emulsions

A stable emulsion may be defined as a system in which the globules retain their initial character and remain uniformly distributed throughout the continuous phase. Various types of deviation from this ideal behaviour are common.

1. BREAKING OR CRACKING

This involves coalescence of the dispersed globules and produces eventual separation of the emulsion into two phases. It will occur if the barriers to coalescence are inefficient or if they are weakened by some means. Some of the factors that cause breaking are—

(*a*) The addition of a substance that is incompatible with the emulsifier may destroy its emulsifying ability, e.g.

(i) The effect of large anions on cationic emulsifying agents.
(ii) The effect of magnesium and calcium ions in hard water on the alkali soaps.
(iii) The effect of phenolic substances on cetomacrogol.

It should be noted that a stable emulsion must not be taken as conclusive evidence of the absence of incompatibility between an emulsifying agent and other ingredients. Satisfactory tests on the particular activities of these other ingredients should be carried out on new formulations.

Incompatibilities between an emulsifying agent and another ingredient may produce a compound that stabilises the opposite type of emulsion to the initial emulsifier so that phase reversal or phase inversion occurs. For example, the addition of sufficient calcium chloride to an o/w emulsion stabilised by a sodium soap causes phase inversion, and the resulting w/o emulsion is stabilised by the calcium soap that is formed by reaction between the original emulsifier and the added calcium chloride.

(*b*) An increase in temperature will increase the number of collisions between globules that are effective in overcoming the barriers to coalescence. This effect is made use of in accelerated tests on the stability of emulsions.

(*c*) An increase in temperature may coagulate certain types of macromolecular emulsifying agent (e.g. proteins). Such adverse effects of high temperature must be taken into account if temperature is used in accelerated testing as a means of increasing the rate of coalescence.

(*d*) Freezing of the aqueous phase will produce ice crystals that may exert unusual pressures on the oil globules. In addition, dissolved salts will concentrate in the remaining unfrozen water and may affect electrical barriers to coalescence.

(*e*) Attempts to incorporate excessive amounts of disperse phase may cause breaking of an emulsion or phase inversion. It can be shown that uniform spheres arranged in the closest packing will occupy 74 per cent of the total volume irrespective of their size. Although it is possible to obtain more concentrated emulsions than this, because of the non-uniform sizes of the globules, there is a critical point at which the emulsion will break.

2. CREAMING

This involves the concentration of dispersed globules in either the upper or lower layers of the emulsion. It is caused by the influence of gravity on the globules, since they usually have a different density from that of the continuous phase. If the globules have a lower density they will tend to concentrate in the upper layers of the system. This is common in dilute o/w emulsions; the cream on milk is a familiar example. Globules will settle to the lower layers if they have a greater density than the continuous phase. This is common in dilute w/o emulsions.

In creaming, the dispersed globules retain their identities and do not coalesce as they do in breaking. Furthermore, a uniform dispersion of globules can be re-obtained by shaking the system. Creaming is not therefore such a serious instability as breaking, since redispersion by shaking cannot be brought about in breaking. However, creaming is undesirable from a pharmaceutical point of view because a creamed emulsion is inelegant in appearance, provides the possibility of inaccurate dosage, and increases the likelihood of coalescence, since the globules are closer together in the cream.

To prevent or delay creaming it is necessary to understand the factors that influence the rate at which it occurs. These factors are similar to those involved in the sedimentation rates of coarse suspensions and are indicated by Stokes' law (Eqn 5.10). A consideration of this equation will show that the rate of creaming will be decreased by (*a*) a reduction in the mean globule diameter, (*b*) a decrease in the difference in the densities of the two phases, and (*c*) an increase in the viscosity of the continuous phase. Conversely, it will be increased by centrifugation which increases the value of g in the equation, and by an increase in temperature, which will decrease the viscosity of the continuous phase.

The main factors utilised in reducing the rate of creaming in pharmaceutical emulsions are—

(*a*) reduction in mean globule diameter—this is usually achieved by using an efficient homogeniser;

(*b*) increase in viscosity of the continuous phase by using thickening agents; for example tragacanth and methylcellulose for o/w emulsions and soft paraffin for w/o emulsions;

(*c*) storage at low temperatures—freezing the aqueous phase should be avoided since this may cause breaking of the emulsion. This method is less attractive to manufacturers as a sole means of ensuring adequate stability because it introduces a stability factor that passes out of their control when the product is sold.

Although creaming is prevented altogether by equalisation of the densities of the two phases this method is little used in pharmaceutical practice since it usually involves the addition of substances that are unacceptable in medicinal preparations.

3. FLOCCULATION OR COAGULATION

This involves globules aggregating into loose masses within the emulsion. The globules do not coalesce and may be redispersed by shaking. Thus, flocculation is less serious than breaking. However, as is emphasised later, flocculation must precede coalescence in the overall process of breaking, and factors that prevent or retard flocculation will therefore increase the stability of emulsions.

Assessment of Stability

Approximate assessments of the relative stabilities of a series of emulsions may be obtained from estimations of the degree of separation of the disperse phase as a distinct layer, or from the degree of creaming. These methods are suitable only for unstable emulsions that exhibit rapid breaking or creaming, and difficulties may arise in measuring the degree of separation accurately.

More precise assessments are obtained from changes in the distribution of globule sizes in an emulsion with time. Microscopy provides the simplest way of determining a globule size distribution but this method is very tedious since large numbers of globules must be measured. Several other methods, which have been reviewed by Sherman (1968), are available.

The single determination of a globule size distribution, such as that shown in Fig. 5.13, is not sufficient for the purposes of assessing stability because stability is not related to the initial globule size in an emulsion. The changes in the shapes of such curves on ageing of the emulsion must be considered. Instability arising from coalescence results in a decrease in the total number of globules and an increase in the size of the remainder, hence the maximum in the distribution curves moves towards the right in Fig. 5.13 and, also, decreases in height. Stability can therefore be assessed in terms of the rate at which such changes occur.

It is possible to observe changes that arise from flocculation, i.e. decrease in the number of single globules, and from coalescence, i.e. the decrease in total number of globules. The kinetics of flocculation and coalescence have been investigated by van den Tempel (1953) and reviews of these and other studies are given by Becher (1965) and Kitchener and Mussellwhite (1968).

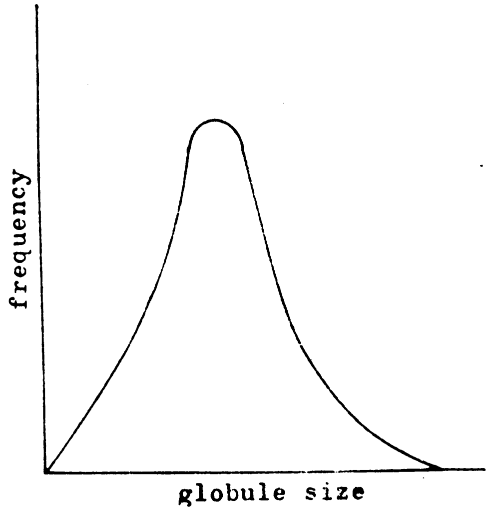


Fig. 5.13 Globule size distribution in an emulsion

Theories of Stability of Emulsions

Several theories attempt to explain the stability of emulsions. Some of the earliest suggestions considered that the reduction in interfacial tension between the oil and water, caused by adsorption of the emulsifying agent at the interface, was the main stabilising factor. However, although a reduction in this interfacial tension will facilitate the formation of a disperse system, if the tension still possesses a finite value, coalescence and breaking will lead to a decrease in the total free energy associated with the oil-water interface. The Bibliography should be consulted for reviews of these and other previous suggestions since the present discussion will be limited to modern theories.

It has already been pointed out that breaking is the most serious instability of emulsions since it is not reversed by shaking. However, it should be noted that the actual coalescence of globules must be preceded by their aggregation or flocculation. In other words, breaking is a two-stage process that involves flocculation followed by coalescence. In any process that involves two consecutive stages, the overall reaction rate is determined by the slower of the two. In breaking emulsions the rate determining stage may therefore be flocculation or coalescence, and whichever is responsible is determined by the properties of the particular emulsion. For example, in a very dilute emulsion the slow rate of flocculation between the low number of globules will tend to be the rate determining stage. In such a case the actual stability of the emulsion will be markedly influenced by those factors that determine the rate of flocculation. In more concentrated emulsions the rate of flocculation is increased, when the coalescence stage may become the rate determining process for breaking of an emulsion. The kinetics of flocculation and coalescence have been investigated by van den Tempel (1953) for systems in which flocculation or coalescence were the rate determining stages and, also, for systems where flocculation and coalescence occurred at similar rates.

FACTORS INFLUENCING THE FLOCCULATION OF GLOBULES

1. *Electrical Double Layer Effects*

The DLVO theory of the stability of lyophobic colloids (p. 54) may be used to explain the stabilisation of emulsions against flocculation.

The amphipathic nature of surface active agents causes them to become adsorbed at an oil-water interface in a particular manner; i.e. with their hydrophilic groups projecting into the aqueous phase and their lipophilic groups projecting into the oily phase. Such adsorption often increases the density of charges around dispersed globules. This is particularly obvious in the case of ionic surface active agents adsorbed at the interface between an oil globule and a surrounding aqueous phase, as shown by Fig. 5.14.

O/w Emulsions. The presence of charges on the surface of dispersed oil globules will create an electrical double layer around each globule. Overlapping of these double layers gives rise to a repulsion, which opposes the van der Waals force of attraction between approaching globules. The variation in net energy of interaction with distance of separation of the globules is shown in Fig. 5.15. Thus, if the potential energy barrier is high enough, colliding globules will not make contact but will bounce apart, and the emulsion will be stable except for the creaming effects. If the approaching globules are able to overcome the energy barrier they will pass into the primary minimum, where they are held by strong attractive forces. In this position the globules will be separated only by very small distances and their stability will be determined by the resistance to rupture of the interfacial films of adsorbed emulsifying agent.

In o/w emulsions, flocculation may also occur in the secondary minimum shown in Fig. 5.15. The attractive forces in this position are relatively weak and the globules are separated by a layer of continuous phase. These aggregates are easily separated by mechanical agitation.

For the repulsive energy barrier to be efficient in preventing globules from passing into the primary minimum it is necessary for the charge density on each globule to be appreciable. This fact implies that an ionised emulsifying agent should form a relatively close packed film around the globules. Such a film is referred to as a condensed film. However, close packing of similarly charged amphipathic ions in an interfacial film will be hindered by the repulsive forces between the similar charges. Thus, a lipophilic hydrocarbon chain of sufficient length must be present in the molecules of the emulsifying agent to promote their adsorption and produce a condensed film. This effect is illustrated by a comparison of the emulsion stabilising abilities of potassium stearate ($C_{17}H_{35}$ COOK) and potassium laurate ($C_{11}H_{23}$ COOK). The longer hydrocarbon chain in the former soap promotes adsorption at the oil-water interface because of its greater hydrophobic character, and the electrostatic barrier is sufficient to provide a good degree of stability to an emulsion. The adsorption of potassium laurate at the oil-water interface is less than that of the stearate since the hydrocarbon chain is shorter. The amount of

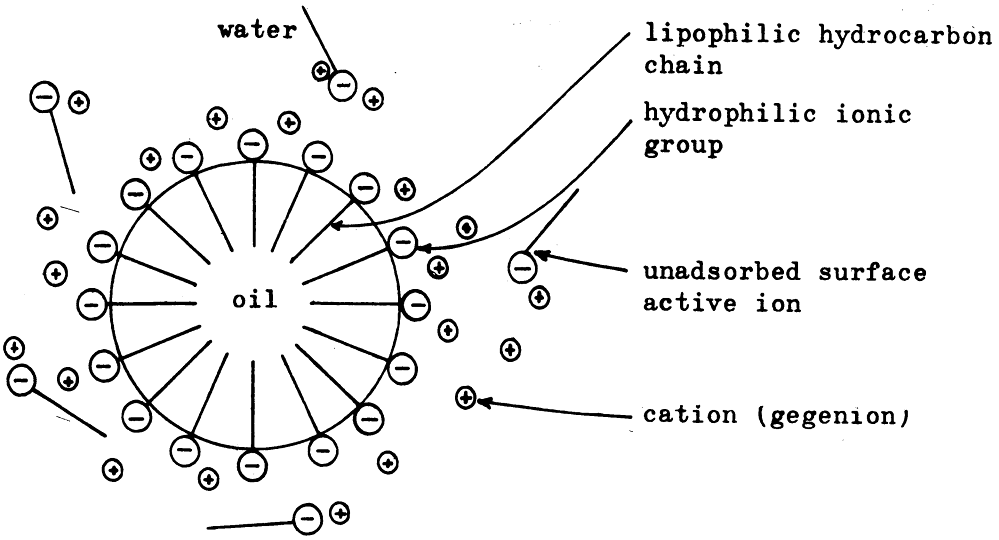


Fig. 5.14 Adsorption of surface active ions on to an oil globule in an o/w emulsion

adsorption is insufficient to provide a satisfactory electrostatic barrier, and poor emulsion stability is obtained with this compound.

The adsorption of a water-soluble amphipathic agent at an oil-water interface is also promoted by the presence of an oil-soluble amphipathic substance in the oil phase. In fact, the effect is mutual and the adsorption of the oil-soluble substance is also increased by the presence of the water-soluble compound. This mutual effect leads to the formation of closely packed films that are referred to as complex condensed films. If the water-soluble agent is ionised, then the electrostatic barrier to flocculation in the primary minimum is increased by complex film formation. Mixtures of water-soluble and oil-soluble amphipathic agents therefore act as efficient emulsifiers although the individual components are only poor emulsifying agents (Schulman and Cockbain, 1940). Such mixtures form the basis of emulsifying waxes such as

(*a*) Emulsifying Wax BP (1968)

Sodium lauryl sulphate (anionic and water-soluble)	1 part
Cetostearyl alcohol (non-ionic and oil-soluble)	9 parts.

(*b*) Cetrimide Emulsifying Wax BPC (1968)

Cetrimide (cationic and water soluble)	1 part
Cetostearyl alcohol (non-ionic and oil-soluble)	9 parts

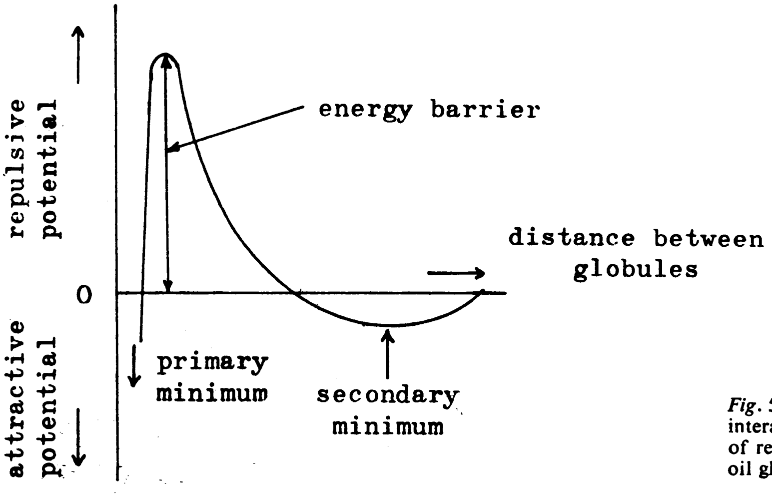


Fig. 5.15 The variation in net potential of interaction arising from combined effects of repulsive and attractive forces between oil globules in an o/w emulsion

The term 'complex film' suggests that the components interact to form some type of molecular complex, which is strongly adsorbed at the oil-water interface. However, it has been suggested (Dervichian, 1958) that the marked changes in the properties of the interface that occur when so-called complex formation takes place may be explained on the basis of the various arrangements of the packed molecules in the adsorbed layer.

W/o Emulsions. Although the ionisation of substances in oils is weak because of the low dielectric constant of these liquids, a small degree of ionisation may occur. Electrical double layers may therefore be associated with water globules dispersed in an oily continuous phase. However, the diffuse part of such double layers is thick and the repulsion energy therefore decreases much more slowly with increasing distance between globules than in o/w emulsions. It has been shown that the potential energy barriers to flocculation that arise from overlapping of electrical double layers are lower in w/o than in o/w emulsions and, furthermore, no secondary minimum is apparent in the net potential energy curves (Sherman, 1963). Flocculation therefore occurs more easily in w/o systems, and flocculated globules are separated only by the small distances associated with primary minima. In addition, the attraction between flocculated globules is strong.

It has been suggested by Schulman and Cockbain (1940) that if the ionic character of an interfacial film is low and its rigidity is high, e.g. unionised solid condensed films, then w/o emulsions will tend to be formed.

2. *Solvation Effects*

It has often been suggested that the effects of hydrophilic groups of adsorbed emulsifier on water molecules in the surrounding aqueous phase give rise to thick solvation sheaths which aid in preventing flocculation. However, although Derjaguin (1964) still advocates this hypothesis, it is more generally believed that solvation effects are of short range and do not contribute to the stability of emulsions against flocculation (Kitchener and Mussellwhite, 1968).

3. *Steric Effects*

The hydrophilic groups of many non-ionic emulsifying agents consist of extended polyethylene oxide chains. Oil globules carrying adsorbed molecules of this type behave as though they are coated with a thick layer of concentrated solution of a hydrophilic colloid, which may prevent close approach of the oil globules. In addition, the very long hydrocarbon chains of some emulsifying agents may prevent the close approach of water globules dispersed in an oil phase so that the distances of separation are too great for the attractive van der Waals forces to be effective.

FACTORS INFLUENCING THE COALESCENCE OF GLOBULES

It is considered that the process of coalescence of stabilised globules can itself be divided into two stages. The first stage, which involves the drainage of the films of liquid from between the adjacent globules, is followed by rupture of the interfacial films around the globules.

1. *Surface Active Agents*

The close approach of two globules causes mutual distortion of their spherical shapes. In globules stabilised by surface active agents this distortion produces a flattening of the approaching surfaces as the liquid layer that separates them becomes thinner. Since a change from a sphere to any other shape must involve an increase in surface area, it follows that thinning of the liquid lamella between the globules is resisted by surface tension forces. The presence of surface active agent provides additional effects that oppose unequal thinning of the liquid layer. These surface effects delay the drainage time of the liquid from between the globules. When the lamella between the globules reaches a critical dimension, rupture may occur and the globules will coalesce. Mechanical disturbance of the system will increase the likelihood of rupture. If the interfacial films are reasonably stable, drainage may proceed further until the globules are separated by a bimolecular layer of surface active agent containing a small amount of liquid. The resistance to coalescence then depends on the stability of such a layer.

2. *Macromolecular Emulsifying Agents*

These possess many hydrophilic and lipophilic groups and are, therefore, strongly adsorbed at the oil-water interface. In fact, the adsorption is often irreversible. High degrees of emulsion stability can be achieved with these agents since the interfacial films possess mechanical properties that are able to withstand collisions and prevent coalescence. The majority of these agents stabilise o/w emulsions.

3. *Finely Divided Solids*

The particles of such solids will be adsorbed at an oil-water interface if they possess a suitable balance

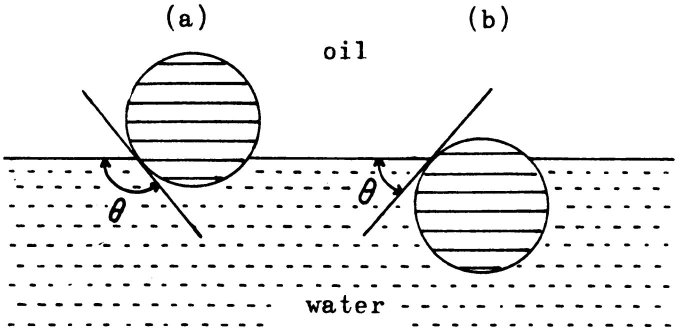


Fig. 5.16 The effect of angle of contact (θ) between solid particles and water on adsorption of particles at the oil-water interface
(*a*) $\theta > \pi/2$ radians; particle is mainly in the oil phase and will tend to stabilise w/o emulsions
(*b*) $\theta < \pi/2$ radians; particle is mainly in the aqueous phase and will tend to stabilise o/w emulsions

between their hydrophilic and lipophilic properties. The relative wettabilities of the particles by oil and water will determine their position at the interface, as shown by Fig. 5.16. Since the interfacial area is at a minimum when the bulk of each particle is in the continuous phase, then the type of emulsion that is stabilised is also influenced by the relative hydrophilic and lipophilic properties of the particles.

Although spherical particles are illustrated in Fig. 5.16, anisometric particles are the most efficient emulsion stabilisers. The mechanism of stability is considered to be concerned with the prevention of thinning of liquid lamellae between globules (Kitchener and Mussellwhite, 1968).

The Preparation of Emulsions

The small-scale methods of preparing emulsions are considered by Gunn and Carter (1965). Various types of machine are used for large scale production. These machines may be classified into the following groups.

1. MIXERS, WHISKS OR CHURNS

These involve agitating the ingredients of an emulsion by beaters. The beaters and containing vessel should be designed to provide a continuous movement in all parts of the liquid. In some cases the container may be fitted with a jacket so that heating or cooling effects may be applied.

2. COLLOID MILLS

As the name suggests these mills (*see* Fig. 5.3) are capable of producing colloidal dispersions. These machines are suitable for preparing emulsions and, unlike the mixers in the previous group, they are capable of continuous production.

3. HOMOGENISERS

These may be operated continuously and are based on the principle that the large globules in a coarse emulsion are broken into smaller globules by passage under pressure through a narrow orifice. Figure 5.17 illustrates the operation of a homogenising valve.

4. ULTRASONIC EMULSIFIERS

If a liquid is subjected to ultrasonic vibrations, alternate regions of compression and rarefaction are produced in it. Cavities are formed in the regions of rarefaction and, later, these collapse with great force which produces emulsification. The required frequency of vibration may be produced electrically, but simple mechanical methods may be used, in which the vibration is produced in a thin metal blade when a jet of the liquid mixture impinges on it. This latter effect is employed in the Rapisonic Homogeniser (Ultrasonics Ltd). Circulation of a coarse emulsion from a reservoir past the vibrating blade into a receiver is involved and further circulation may be applied until an emulsion with the desired globule size distribution is obtained. More information on this machine and its method of use is given by Myers and Goodman (1954).

Pharmaceutical Applications of Emulsions

O/w emulsions are convenient preparations for the oral administration of unpalatable oils or oily solutions of drugs with unpleasant tastes. However, their use in this manner has been decreased by introducing preparations in which the oils are solubilised within micelles, since these systems are not subject to the physical instabilities of emulsions. In addition, the stability of solubilised material is often greater than when it is included in the oil globules of emulsions.

O/w emulsions are also used as a dosage form for the intravenous administration of oils and fats with high calorie contents to patients who cannot ingest food by other means. The globules in these emulsions should be similar in size to chylomicrons.

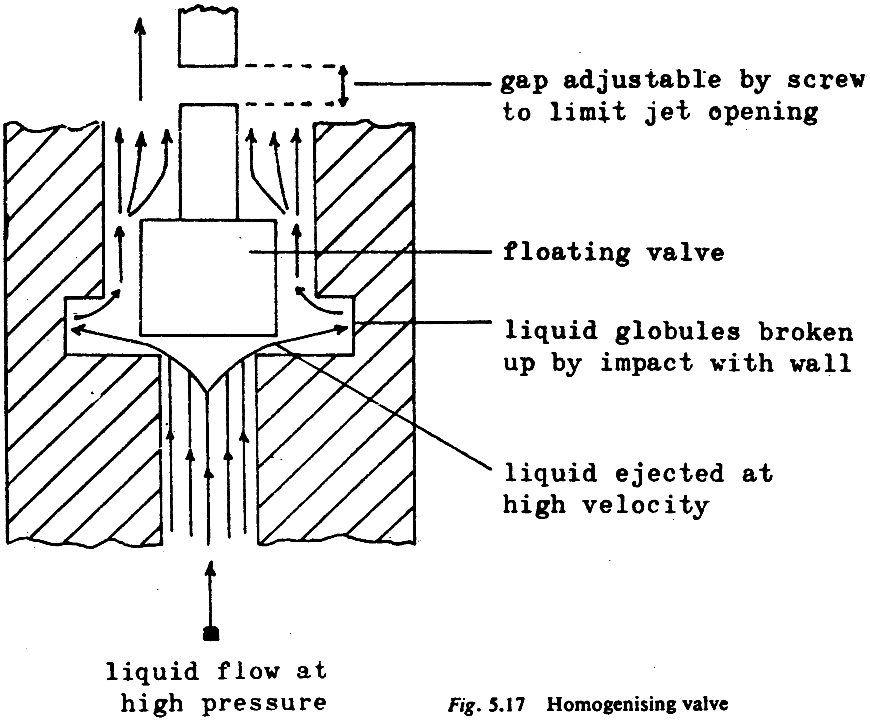


Fig. 5.17 Homogenising valve

Emulsions of both types (i.e. o/w and w/o) are used extensively in pharmaceutical preparations for external use and in cosmetic preparations. The factors that should be taken into account in formulating these products are considered by Gunn and Carter (1965), Pernarowski (1970), and Becher (1965). The rheological properties of emulsions are a particularly important aspect of the formulation of these systems. The factors that influence these properties are discussed in Chapter 7.

The use of foams in many cosmetic preparations and some pharmaceutical preparations provides an expanding field for the applications of emulsion technology since many of these products are emulsified systems (p. 87).

Foams

A foam is a coarse dispersion of a gas in liquid, which is present as thin sheets or lamellae between the gas bubbles.

Stability of Foams

This is determined by the rate of drainage of the liquid from the foam, which leads to a decrease in the strength of the lamellae. When the lamellae rupture, the liquid contained in them breaks into drops that have a lower total surface than the original film and, consequently, the free energy of the system is decreased. A foam is therefore unstable from a thermodynamic point of view. However, although the lifetimes of foams containing pure liquids and gases are very low, the presence of suitable stabilising agents (foaming agents) can provide almost indefinite stability. Many stabilising agents are available, the most important types of which are the proteins, insoluble powders, and synthetic surface active agents that are adsorbed at the liquid-gas interface.

The drainage rates of liquids from lamellae is reduced by increases in the surface and bulk viscosities of the liquids. In addition, these increases also protect the liquid films against disturbances arising from mechanical and thermal effects. Such changes are particularly important in foam stabilisation, since foaming agents that produce marked increases in the surface viscosity yield foams with high stabilities, whereas those that do not, produce rapidly draining foams of low stability.

The thinning of lamellae is also opposed by the adsorption of ionic foaming agents at the opposite faces of the liquid films. Excessive drainage from such lamellae would bring the similarly charged surfaces closer together. In addition, the increasing concentration of gegenions that builds up in the liquid films tends to oppose drainage by exerting a high osmotic pressure.

The presence of surface active agents also opposes the unequal thinning of liquid films that may, for example, be caused by mechanical shock. This effect arises from the transient change in distribution of the surface active agent that occurs when a film is stretched. It causes the adsorbed monolayer to spread into the extended region of the interface and so prevents an increase in the interfacial free energy.

Foams are often troublesome, and knowledge of the action of substances that cause their destruction (foam breakers or antifoams) is therefore useful. These substances (e.g. ether, n-octanol, silicones) are usually adsorbed at the liquid-gas interface and produce regions of lower surface tension. Extension of the interface in these regions allows further thinning of the lamellae to occur until they eventually rupture.

Preparation of Foams

Foams are usually prepared by shaking the liquid and gas together in the presence of the foaming agent, or by bubbling the gas through a solution of the foaming agent.

The use of pressurised containers for the presentation of foams is of particular interest in pharmacy. The general advantages of this type of container have already been listed (p. 74). Systems used in this type of preparation include:

(*a*) emulsions containing the liquefied propellant, water, and active ingredients. When the valve on the container is opened the vapour pressure of the propellant forces the emulsion out of the container through a nozzle. At atmospheric pressure and room temperature the emulsified propellant reverts to a vapour and so produces a foam;

(*b*) emulsified products used in conjunction with soluble compressed gases such as carbon dioxide. This type of system is usually shaken before use in order to disperse some of the gas throughout the liquid concentrate. When the valve is opened, emulsion plus entrapped gas is forced out of the container by the pressure of the compressed gas inside. The liberation of entrapped gas whips the concentrated emulsion into a foam.

Pharmaceutical Applications of Foams

Variation of the ingredients used in the emulsified systems that are presented in pressurised containers allows the formulation of aqueous or non-aqueous foams that are either stable or rapidly breaking. It has been suggested that various types of product would be useful for topical preparations, rectal and vaginal medication and for burn dressings.

A further application of foams is encountered in the preparation of the absorbable haemostats Human Fibrin Foam and Gelatin Sponge (p. 425).

BIBLIOGRAPHY

ADAMSON, A. W. (1967) *Physical Chemistry of Surfaces.* 2nd ed. Interscience, New York.

BECHER, P. (1965) *Emulsions: Theory and Practice.* 2nd ed. Reinhold, New York.

BIKERMAN, J. J. (1958) *Surface Chemistry.* 2nd ed. Academic Press, New York.

DAVIES, J. T. and RIDEAL, E. K. (1963) *Interfacial Phenomena.* 2nd ed. Academic Press, New York.

JIRGENSONS, B. (1958) *Organic Colloids.* Elsevier, Amsterdam.

KRUYT, H. R. (1952 and 1949) *Colloid Science. Vols. I and II.* Elsevier, Amsterdam.

MOILLIET, J. L., COLLIE, B., and BLACK, W. (1961) *Surface Activity.* 2nd ed. Spon, London.

MULLEY, B. A. (1964) *Advances in Pharmaceutical Sciences. Vol. 1.* (H. S. Bean, A. H. Beckett, and J. E. Carless, Eds). Academic Press, London. Chapter 2.

SHERMAN, P. (1968) *Emulsion Science.* Academic Press, London.

SHINODA, K., NAKAGURA, T., TAMAMUSHI, B., and ISEMURA T. (1963) *Colloidal Surfactants.* Academic Press, London.

TROTMAN-DICKENSON, A. F. and PARFITT, G. D. (1966) *Chemical Kinetics and Surface and Colloid Chemistry.* Pergamon, London.

WILLIAMS, V. R. and WILLIAMS, H. B. (1967) *Basic Physical Chemistry for the Life Sciences.* Freeman and Co., San Francisco.

REFERENCES

BECHER, P. (1965) *Emulsions: Theory and Practice.* 2nd ed. Reinhold, New York.

COOPER, K. E. (1963) *Analytical Microbiology.* (F. Kavanagh Ed.) Academic Press, London. Chapter 1.

COOPER, K. E. and WOODMAN, D. (1946) The diffusion of antiseptics through agar gels, with special reference to the agar cup assay method of estimating the activity of penicillin. *J. Path. Bact.* **58,** 75–84.

CRAIG, L. C. and KONISBERG, W. (1961) Dialysis studies. III. Modification of pore size and shape in cellophane membranes. *J. phys. Chem., Ithaca* **65**, 166–172.

DERJAGUIN, B. V. and LANDAU, L. D. (1941) Theory of the stability of strongly charged lyophobic sols and the adhesion of strongly charged particles in solutions of electrolytes. *Acta phys.-chim. URSS*, **14**, 633.

DERJAGUIN, B. V. (1964) *Nineteenth Symposium of the Society of Experimental Biology*, Swansea, 55.

DERVICHIAN, D. G. (1958) *Surface Phenomena in Chemistry and Biology.* (J. F. Danielli, K. G. A. Pankhurst, and A. C. Riddiford, Eds) Pergamon, London. Chapter 5.

EVERSOLE, W. G. and DOUGHTY, E. W. (1935) Diffusion coefficients of molecules and ions. *J. phys. Chem. Ithaca*, **34**, 289.

GUNN, C. and CARTER, S. J. (1965) *Dispensing for Pharmaceutical Students.* 11th ed. Pitman Medical, London.

HERMANS, P. H. (1949) *Colloid Science. Vol. II.* (H. R. Kruyt, Ed.) Elsevier, Amsterdam. Chapter 12.

JIRGENSONS, B. (1958) *Organic Colloids.* Elsevier, Amsterdam.

KITCHENER, J. A. and MUSSELLWHITE, P. R. (1968) *Emulsion Science.* (P. Sherman, Ed.). Academic Press, London. Chapter 2.

MICHAELS, A. S. and BOLGER, J. C. (1964) Particle interactions in aqueous kaolinite dispersions. *Ind. Engng. Chem. Fundamentals*, **3**, 14–20.

MYERS, J. A. and GOODMAN, J. E. (1954) Pharmaceutical applications of ultrasonics. *Pharm. J.* **173**, 422-424.

OVERBEEK, J. TH. G. (1952) *Colloid Science. Vol. I.* (H. R. Kruyt, Ed.) Elsevier, Amsterdam. Chapter 2.

PERNAROWSKI, M. (1970) *Remington's Pharmaceutical Sciences.* 14th ed. Mack Publishing Co., Pennsylvania. Chapter 80.

SCHULMAN, J. H. and COCKBAIN, E. G. (1940) Molecular interactions at oil–water interfaces. (I) Molecular complex formation and the stability of oil-in-water emulsions. (II) Phase inversion and the stability of water-in-oil emulsions. *Trans. Faraday Soc.* **36**, 651–660 and 661–668.

SCIARRA, J. J. (1970) *Remington's Pharmaceutical Sciences.* 14th ed. Mack Publishing Co., Pennsylvania. Chapter 90.

SHELUDKO, A. (1966) *Colloid Chemistry.* Elsevier, Amsterdam.

SHERMAN, P. (1963) Changes in the rheological properties of emulsions on ageing, and their dependence on the kinetics of globule coagulation. *J. phys. Chem. Ithaca* **67**, 2531-2537.

SHERMAN, P., Ed. (1968) *Emulsion Science.* Academic Press, London. Chapter 4.

THEIMER, W. (1960) Structure and ageing of gelatin gels. *Z. Naturf.* Pt. 15b, 346–350.

VAN DEN TEMPEL, M. (1953) Stability of oil-in-water emulsions. I The electrical double layer at the oil-water interface. II Mechanism of the coagulation of an emulsion. III Measurement of the rate of coagulation of an oil-in-water emulsion. *Recl. Trav. chim. Pays-Bas Belg.* **72**, 419–432, 433–441 and 442–461.

VAN OLPHEN, H. (1963) *An introduction to Clay Colloid Chemistry.* Interscience, London.

VERWEY, E. J. W. and OVERBEEK, J. TH. G. (1948) *Theory of the Stability of Lyophobic Colloids.* Elsevier, Amsterdam.

6

Kinetics

INVESTIGATIONS on the rates at which changes occur in a particular system and the factors that influence such rates are known as kinetic studies. They are useful in providing information that (*a*) gives an insight into the mechanism of the change involved, and (*b*) allows prediction of the degree of change that will occur after a given time has elapsed.

In general, the theories and laws of chemical kinetics are well proven and provide a sound basis for the application of such studies to pharmaceutical problems that involve chemical reactions; e.g. the decomposition of medicinal compounds. However, kinetic studies of the processes involved in the absorption of drugs into, their distribution within, and removal from living organisms are more recent. The theoretical basis of drug kinetics is therefore often less well-founded than that of chemical kinetics. Nevertheless, the application of knowledge derived from these studies allows a more rational approach to be made towards the synthesis of medicinal compounds with particular activities and to the effects of formulation on these activities. In addition, the increasing amount of information that is becoming available will increase the application of these studies in pharmacy.

This chapter attempts to indicate the more important applications of kinetic studies in pharmacy. Those concerned with chemical reactions will be discussed initially and followed by those concerned with the behaviour of drugs in living organisms.

CHEMICAL KINETICS

EXTENT OF REACTION

Consider the formation of a compound B from the reaction between two molecules of a compound A; e.g. a dimerisation. This reaction may be described by Eqn (6.1).

$$2A = B \tag{6.1}$$

The extent to which this reaction has proceeded may be defined in terms of the change in either the amount of product B or the amount of reactant A.

The change in the amount of B is the difference between the amount of B produced (n_B) and the amount of B present at the starting point ($n_{B.0}$). (The starting point is usually the beginning of the reaction where $n_{B.0} = 0$.) Thus, the extent of reaction (ξ) defined in terms of the change in amount of substance B is given by Eqn (6.2).

$$\xi = n_B - n_{B.0} \tag{6.2}$$

The change in the amount of A is the difference between the amount present at the starting point ($n_{A.0}$) and the amount that remains (n_A). However, two molecules of A are involved in the balanced Eqn (6.1) and the extent of reaction (ξ) defined in terms of the change in the amount of reactant A is given by Eqn (6.3).

$$\xi = \frac{n_{A.0} - n_A}{2} \tag{6.3}$$

It is obvious from Eqn (6.1) that Eqns (6.2) and (6.3) will give the same value of ξ. It is therefore convenient to write a general equation for the definition of extent of reaction. This is represented by Eqn (6.4),

$$\xi = \frac{n_X - n_{X.0}}{\nu_X} \tag{6.4}$$

where n_X and $n_{X.0}$ are the actual and initial amounts of a substance X that is involved in a reaction as a reactant or a product. ν_X is called the stoichiometric number of X and it indicates the number of molecules (or atoms, or radicals, or ions) of X that are involved in the balanced equation, which describes the reaction under consideration. When X is a product the difference ($n_X - n_{X.0}$) is positive because the amount of product increases as the reaction proceeds; (i.e. $n_X > n_{X.0}$). However, if X is a reactant, ($n_X - n_{X.0}$) is negative because the amount of X will decrease as reaction proceeds; (i.e. $n_X < n_{X.0}$). This variation in the sign of ($n_X - n_{X.0}$) is overcome by using the convention that the stoichiometric number ν_X is positive for a product and negative for a reactant, so that the right-hand side of Eqn (6.4) is always positive.

It can be seen from Eqn (6.4) that the dimension of extent of reaction, ξ, will be amount of substance and, therefore, the units of ξ will be the same as those of n_X, for which the basic SI unit is the mole.

The change in extent of reaction ($\mathrm{d}\xi$) with increasing change in amount of substance X ($\mathrm{d}n_X$) may be expressed by Eqn (6.5), where ν_X is defined as before.

$$\mathrm{d}n_X = \nu_X\,\mathrm{d}\xi \qquad (6.5)$$

Rate of Reaction ($\dot{\xi}$)

This may be defined as the rate of increase of the extent of reaction (i.e. $\mathrm{d}\xi/\mathrm{d}t$) as is shown by Eqn (6.6).

$$\dot{\xi} = \mathrm{d}\xi/\mathrm{d}t \qquad (6.6)$$

The dimensions of $\dot{\xi}$ will be amount of substance per unit time and its basic SI unit will be moles per second (mol s^{-1}). From Eqns (6.5) and (6.6) it can be seen that

$$\dot{\xi} = \frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{1}{\nu_X}\frac{\mathrm{d}n_X}{\mathrm{d}t} \qquad (6.7)$$

where $\mathrm{d}n_X/\mathrm{d}t$ may be called the rate of formation of X if X is a product, or the rate of disappearance of X if X is a reactant.

Equations (6.6) and (6.7) are general ones for the definition of rate of reaction and they are independent of the conditions under which a reaction is carried out. For example, they are valid for reactions in which the volume varies with time or where the reaction occurs in more than one phase.

When discussing chemical kinetics it is convenient to divide reactions into two classes; homogeneous and heterogeneous reactions. The former occur in one phase, e.g. in solution, while the latter occur at the interface between two phases, e.g. at solid–gas or solid–liquid interfaces.

HOMOGENEOUS REACTIONS

These reactions are usually classified on the basis of the effect of the concentration of reactants on the rates of increase in concentration of products or decrease in concentration of reactants. To convert Eqn (6.7) to concentration terms it is necessary to divide by the volume, V, of the homogeneous system, when Eqn (6.8) is obtained.

$$\frac{1}{V}\frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{1}{V\nu_X}\frac{\mathrm{d}n_X}{\mathrm{d}t} \qquad (6.8)$$

If the volume, V, of the phase does not vary with time during the reaction then Eqn (6.8) may be written in the form of Eqn (6.9),

$$\frac{1}{V}\frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{1}{\nu_X}\frac{\mathrm{d}(n_X/V)}{\mathrm{d}t} \qquad (6.9)$$

However, the term n_X/V represents the concentration of substance X and may be replaced by a single term; e.g. c_X or $[X]$, and Eqn (6.9) may be rewritten as,

$$\frac{1}{V}\frac{\mathrm{d}\xi}{\mathrm{d}t} = \frac{1}{\nu_X}\frac{\mathrm{d}c_X}{\mathrm{d}t} = \frac{1}{\nu_X}\frac{\mathrm{d}[X]}{\mathrm{d}t} \qquad (6.10)$$

If X is a product the rate at which its concentration increases is represented by the symbol v_X and is equal to the term $\mathrm{d}c_X/\mathrm{d}t$ in Eqn (6.10). If X is a reactant, then v_X represents its rate of disappearance, given by $v_X = -\mathrm{d}c_X/\mathrm{d}t$. The minus sign is necessary in this latter case because the concentration of a reactant decreases with time. Thus, the term $-\mathrm{d}c_X/\mathrm{d}t$ is positive because $\mathrm{d}c_X$ is negative.

Order of Reaction

Experiments show that the rates of change in the concentrations of products or reactants usually depend on the concentrations of the reactants. This dependence is indicated by Eqn (6.11),

$$v \propto [A]^a[B]^b \quad \text{or} \quad v = k[A]^a[B]^b \qquad (6.11)$$

where $[A]$ and $[B]$ are the concentrations of reactants A and B respectively, a and b are some powers of these concentrations, and k is a proportionality constant known as the rate constant of the reaction at a particular temperature.

Classification of reactions may be achieved by division into various orders, where the order of a reaction is given by the sum of the powers of the concentration terms involved in equations, such as that represented by Eqn (6.11). For example, in this particular case the order of reaction is given by $(a + b)$. In fact, this represents the overall order of the reaction. It is also possible to refer to the order of a reaction with respect to a particular reactant. Thus, the above reaction is of order a with respect to A, and of order b with respect to B.

Homogeneous reactions are usually interpreted in terms of simple orders of reaction, which are discussed below.

(a) First Order Reactions

The rate of change in the concentrations of products and reactants in this type of reaction is proportional to the first power of the concentration (c_X) of a single reactant (X) and is independent of the concentration of any other substance that may be present.

The rate of formation of a product Y is therefore given by Eqn (6.12) and the rate of disappearance of reactant X is given by Eqn (6.13)

$$v_Y = \frac{\mathrm{d}c_Y}{\mathrm{d}t} = kc_X \qquad (6.12)$$

$$v_X = -\frac{\mathrm{d}c_X}{\mathrm{d}t} = kc_X \qquad (6.13)$$

where k is the rate constant.

If the concentration of reactant X at the beginning of a reaction when time $t = 0$ is denoted by a, and the amount that has reacted after time t is denoted by x, then the amount of X that remains at this time is given by $(a - x)$. Equation (6.13) may therefore be rewritten in the form shown by Eqn (6.14)

$$-\frac{\mathrm{d}c_X}{\mathrm{d}t} = k(a - x) \quad \text{or} \quad \frac{\mathrm{d}c_X}{a - x} = -k\,\mathrm{d}t \qquad (6.14)$$

where $-\mathrm{d}c_X/\mathrm{d}t$ represents the rate of decrease in the concentration of X. Integration of Eqn (6.14) between the time limits of 0 and t gives

$$\int_a^{(a-x)} \frac{\mathrm{d}c_X}{a - x} = -k\int_0^t \mathrm{d}t$$

$$\therefore \quad \ln(a - x) - \ln a = -kt$$

Converting from natural logarithms gives

$$\log(a - x) - \log a = -\frac{kt}{2{\cdot}303}$$

or

$$\log(a - x) = \log a - \frac{kt}{2{\cdot}303} \qquad (6.15)$$

It can be seen that Eqn (6.15) is representative of a linear relation (i.e. $y = c + mx$) and the variables are $(a - x)$ and t. If the first order law is obeyed, then a graph of $\log(a - x)$ versus t will give a straight line with a slope of $-k/2{\cdot}303$ and an intercept at $t = 0$ of $\log a$.

The rate constant k may therefore be calculated from the slope of such a graph. It may also be obtained by substitution of experimental values into Eqn (6.16) which is a rearranged form of Eqn (6.15).

$$k = \frac{2{\cdot}303}{t}\log\frac{a}{a - x} \qquad (6.16)$$

It can be seen that the dimension of k for a first order reaction is reciprocal time and the basic SI unit is the reciprocal second (s^{-1}).

(*b*) *Second Order Reactions*

The rate of change in the concentrations of products and reactants in this type of reaction is proportional either to the second power of the concentration of a single reactant, or to the first powers of the concentrations of two reactants. These two possibilities are illustrated by the following Eqns (6.17) and (6.18), which refer to the rate of decrease in the concentration of reactant X—

(i) $$v_X = -\mathrm{d}c_X/\mathrm{d}t = k[X][Y] \qquad (6.17)$$

(ii) $$v_X = -\mathrm{d}c_X/\mathrm{d}t = k[X]^2 \qquad (6.18)$$

where the rates of concentration change are dependent on the concentrations of reactants X and Y, or X only. In the first type of second order reaction given above the rate of decrease in concentration of reactant Y will be equal to that of X (i.e. $v_Y = v_X$). If the concentrations of reactants X and Y at time $t = 0$ are given by a and b respectively, and the concentration of each substance that has reacted after time t is equal to x, then the concentrations of X and Y that remain after this time are given by $(a - x)$ and $(b - x)$ respectively. Equation (6.17) may therefore be rewritten as

$$-\frac{\mathrm{d}x}{\mathrm{d}t} = k(a - x)(b - x) \qquad (6.19)$$

where $-\mathrm{d}x/\mathrm{d}t$ represents the rate of decrease in the concentration of X (or Y). Integration of Eqn (6.19) yields Eqn (6.20)

$$kt = \frac{2{\cdot}303}{(a - b)}\log\frac{b(a - x)}{a(b - x)} \qquad (6.20)$$

where k is the rate constant.

Rearrangement of Eqn (6.20) yields

$$\log\frac{a - x}{b - x} = \frac{(a - b)kt}{2{\cdot}303} + \log\frac{a}{b} \qquad (6.21)$$

and if the second order law is obeyed, then a graph of the left-hand side of Eqn (6.21) against time t should produce a straight line with a slope equal to $(a - b)k/2{\cdot}303$, and an intercept at $t = 0$ of $\log(a/b)$.

When the initial concentrations of reactants X and Y are equal (i.e. $a = b$) or when the second power of a single reactant determines the rate of concentration change, then the integrated form of the rate equation is given by Eqn (6.22)

$$kt = \frac{x}{a(a - x)} \qquad (6.22)$$

Rearrangement of Eqn (6.22) yields

$$\frac{1}{(a - x)} - \frac{1}{a} = kt \qquad (6.23)$$

and if the second law is obeyed then a graph of $1/(a - x)$ against time should produce a straight line with a slope of k and an intercept of $1/a$ at $t = 0$.

It can be shown from the previous equations that the rate constant for a second order reaction has the dimensions of time^{-1} concentration^{-1} and the SI unit is m^3 mol^{-1} s^{-1}.

(*c*) *Pseudo First Order Reactions*

If a large excess of one of the reactants in a second order reaction is present throughout the reaction, then its concentration remains virtually constant and the rate of concentration change follows the first order law. Hydrolysis reactions in dilute aqueous solution are common examples of this type of reaction, which is known as a pseudo first order reaction.

However, it should be borne in mind that a decrease in the concentration of water, e.g. caused by a change in the composition of the solvent, may lead to a reaction that follows second order kinetics.

(*d*) *Reactions of Third and Higher Orders*

The rates of change in concentrations in this type of reaction are proportional to three concentration terms. However, such reactions are rare and their analysis is complex. Reactions of even higher orders are unlikely to occur.

(*e*) *Zero Order Reactions*

The rates of change in the concentrations of reactants and products in reactions of this type are dependent on some factor other than the concentration of a reactant. They include photochemical reactions that depend on the absorption of light and heterogeneous reactions that depend on the area of the interface at which reaction occurs (p. 94, 96).

The rate of change in concentration of a reactant X is constant in a zero order reaction as indicated by Eqn (6.24), where k is the rate constant, which has the same dimensions as v_X; i.e. mol s^{-1}.

$$v_X = -\frac{dc_X}{dt} = k \qquad (6.24)$$

If the amount of X that reacts in time t is denoted by x then the integrated form of the equation for a zero order reaction is given by

$$x = kt \qquad (6.25)$$

Half-life ($t_{\frac{1}{2}}$)

The half-life of a reaction is the time required for the concentration of a reactant to decrease to half its original value; i.e. when $x = a/2$ in the previous equations. The expressions of half-life in terms of the rate constant (k) for the disappearance of reactant are shown in Table 6.1 for simple orders of reaction. It can be seen from these expressions that half-life can be defined in terms of Eqn (6.26) for the simple orders given in Table 6.1.

$$t_{\frac{1}{2}} \propto \frac{1}{a^{(n-1)}} \qquad (6.26)$$

The half-life of a reaction provides a convenient means of expressing the rate of concentration change of a reactant and is particularly useful because the time taken for complete reaction is theoretically infinite except in zero order reactions.

Methods of Determining the Order of a Reaction

1. *Graphical Method*

A straight line is obtained when the data from kinetic experiments are plotted in the form that is relevant to the order of a particular reaction. The

Table 6.1

Summary of Information on Reactions of Simple Orders

			Linear graph			
Order	*Integrated rate equation*	*Half-life equation*	*Ordinate*	*Abscissa*	*Slope*	*Intercept*
Zero	$x = kt$	$t_{\frac{1}{2}} = \frac{a}{2k}$	x	t	k	0
First	$\log\frac{a}{(a-x)} = \frac{kt}{2{\cdot}303}$	$t_{\frac{1}{2}} = \frac{0{\cdot}693}{k}$	$\log(a-x)$	t	$-\frac{k}{2{\cdot}303}$	$\log a$
Second ($a = b$)	$\frac{x}{a(a-x)} = kt$	$t_{\frac{1}{2}} = \frac{1}{ak}$	$\frac{1}{a-x}$	t	k	$\frac{1}{a}$

ordinates and abscissae of graphs for simple orders are summarised in Table 6.1, together with the values of the slopes and intercepts of these graphs.

2. *Substitution Method*

The order of a reaction is indicated by the particular integrated rate equation that gives a constant value of k for data obtained from a kinetic experiment. The integrated equations are given in Table 6.1 for simple orders of reaction.

3. *Half-life Method*

Consider a reaction in which the initial concentration of the reactant is a_1 and the concentration at a later time is a_2. These concentrations may be regarded as the starting points of two separate reactions, and the corresponding half-lives may be represented by $t_{\frac{1}{2}(1)}$ and $t_{\frac{1}{2}(2)}$, respectively. From Eqn (6.26) it is possible to write that

$$t_{\frac{1}{2}(1)} \propto \frac{1}{a_1^{(n-1)}}$$

and

$$t_{\frac{1}{2}(2)} \propto \frac{1}{a_2^{(n-1)}}$$

$$\therefore \quad \frac{t_{\frac{1}{2}(1)}}{t_{\frac{1}{2}(2)}} = \frac{a_2^{(n-1)}}{a_1^{(n-1)}}$$

$$\therefore \quad \log \frac{t_{\frac{1}{2}(1)}}{t_{\frac{1}{2}(2)}} = (n-1) \log \frac{a_2}{a_1}$$

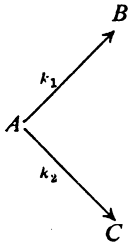
$$\therefore \quad n = \frac{\log (t_{\frac{1}{2}(1)}/t_{\frac{1}{2}(2)})}{\log (a_2/a_1)} + 1 \qquad (6.27)$$

COMPLEX REACTIONS

Many reactions involve several stages and the kinetics of the overall reaction cannot be expressed in terms of simple orders of reaction. However, it is often possible to determine the order of the separate stages. The following types of complex reaction are encountered frequently and examples of each type are given by Martin *et al.* (1969) and Saunders (1966).

1. *Simultaneous Reactions*

These involve the production of two or more products from the same reactant by different reactions, which proceed at different rates. They may be represented diagrammatically as



Substitution reactions into aromatic nuclei are examples of this type of complex reaction; e.g. the simultaneous formation of 1- and 4-nitrophenols during the nitration of phenol.

2. *Consecutive Reactions*

These may be represented diagrammatically as

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

and each stage proceeds at a particular rate. The overall rate will be determined by the rate of the slowest stage.

3. *Chain Reactions*

These reactions usually proceed at slow rates under normal conditions but these rates are increased, often very markedly, by the presence of substances that are able to produce free radicals. (A free radical is an electrically uncharged atom or group of atoms with an unpaired electron and is highly reactive.) Polymerisation and oxidation of unsaturated organic compounds are important examples of reactions that involve a chain mechanism.

4. *Reversible Reactions*

These are reactions in which the reaction products react to produce the original reactants. They may be represented diagrammatically as

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B$$

where the arrows indicate that the opposing reactions are proceeding simultaneously.

The rate of increase in concentration of $A(v_A)$ is given by Eqn (6.28)

$$v_A = \frac{dc_A}{dt} = -k_1 c_A + k_2 c_B \qquad (6.28)$$

where c_A and c_B are the concentrations of substances A and B, and k_1 and k_2 are the rate constants for the forward and reverse reactions, respectively. In the initial stages of reaction the concentration of A will be large while that of B will be small. The product k_1c_A will therefore be greater than k_2c_B and the concentration of A will decrease and that of B will increase. Eventually k_1c_A will equal k_2c_B, and an equilibrium will be established where $\mathrm{d}c_A/\mathrm{d}t = \mathrm{d}c_B/\mathrm{d}t = 0$. If the initial concentration of A is denoted by $c_{A,0}$, and the concentrations of A and B at equilibrium are $c_{A,e}$ and $c_{B,e}$ respectively, then

$$c_{B,e} = c_{A,0} - c_{A,e}$$

and

$$\frac{c_{B,e}}{c_{A,e}} = \frac{k_1}{k_2}$$

$$\therefore \quad \frac{c_{A,0} - c_{A,e}}{c_{A,e}} = \frac{k_1}{k_2} = K \qquad (6.29)$$

where K, which is known as the equilibrium constant, is the ratio of the forward reaction rate constant to the reverse reaction rate constant. Its dimensions will depend on the number of concentration terms in the numerator and denominator respectively of Eqn (6.29), and if it is not a dimensionless number, i.e. if the concentration terms do not cancel completely, its units will depend on those in which the remaining terms are expressed.

All chemical reactions probably come to some equilibrium if sufficient time is allowed, provided the products are not removed from the system. However, in many cases reaction may be regarded as being irreversible for practical purposes, since the rate constant for the reverse reaction is very small. The higher the value of K the closer the forward reaction approaches completion.

HETEROGENEOUS REACTIONS

As previously stated these reactions occur at the interface between two phases. Although the rates of these reactions may be defined by the general Eqns (6.6) and (6.7) the interpretation of their kinetics in terms of simple models such as first and second order laws is prevented by the influence of other factors on the rates of concentration change of products and reactants. For example, the area of the interface and the porosity of the solid phase are particularly important in controlling these rates.

Ion Exchange

The exchange of ions between a solid phase and a solution is one of the most important examples of heterogeneous reactions. These reactions involve an exchange between mobile ions associated with a solid and similarly charged ions in a solution and they may be illustrated by the following equations—

$$RSO_3H + NaCl \rightleftharpoons RSO_3Na + HCl \qquad (6.30)$$

In this equation, RSO_3H represents an insoluble material which produces fixed anions (RSO_3^-) and mobile cations (H^+) on ionisation. The mobile cations are exchangeable with other cations (e.g. Na^+) as shown in the above equation. The insoluble RSO_3H is therefore termed a cation exchanger.

$$RN(CH_3)_3OH + NaCl \rightleftharpoons RN(CH_3)_3Cl + NaOH \qquad (6.31)$$

This equation represents the exchange of mobile OH^- ions by Cl^- and the insoluble material $RN(CH_3)_3OH$, which contains fixed cations, is termed an anion exchanger.

The materials (RSO_3H and $RN(CH_3)_3OH$) involved in the above reactions are typical examples of organic ion exchange resins, where R represents a resin, which usually consists of polymeric chains of styrene that are cross-linked by divinyl benzene, as shown in Fig. 6.1. An increase in the proportion of divinyl benzene increases the degree of cross-linking and therefore decreases the porosity and increases the rigidity of these resins.

The ion exchange groups (e.g. —SO_3H and —$N(CH_3)_3OH$) are introduced by chemical reaction between the hydrocarbon polymer and suitable reagents. The nature of the ion exchange group that is introduced will determine the type of ions that can be exchanged and provides a means of classifying ion exchange resins into the following groups.

(*a*) *Strong Cation Exchangers*

These contain strongly acidic groups such as —SO_3H, and the mobile H^+ ions are capable of exchanging with the cations of salts of strong acids. For example, Eqn (6.30) illustrates the exchange of sodium ions from a solution of sodium chloride. The opposing arrows in this equation indicate that the exchange process is reversible and an equilibrium is therefore obtained. The reversibility also indicates that the original acid form of the ion exchange resin (i.e. RSO_3H) can be regenerated from the sodium form by washing with dilute hydrochloric or sulphuric acid.

(*b*) *Strong Anion Exchangers*

These contain quaternary ammonium groups attached to the hydrocarbon resin, and the mobile OH^- ions of the hydroxides of these resins are capable of exchanging with the anions of salts of

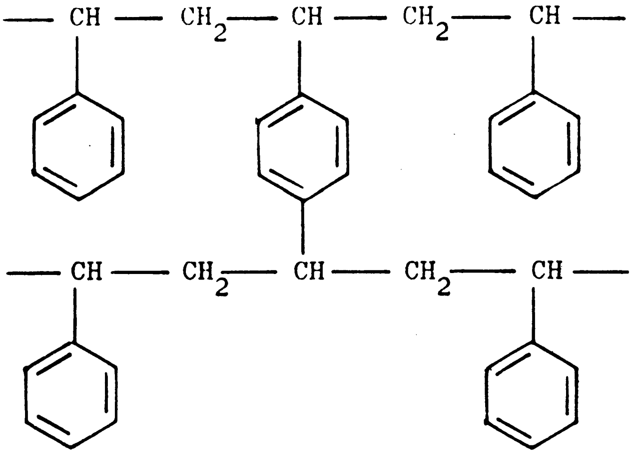


Fig. 6.1 Structure of a styrene-divinyl benzene resin

strong bases. Equation (6.31) illustrates the exchange of chloride ions from a solution of sodium chloride, and the reversibility of this equation indicates that the original ion exchange resin can be regenerated by treatment with dilute NaOH.

(*c*) *Weak Cation Exchangers*

These contain weaker acidic groups than the sulphonated strong cation exchangers. For example, carboxylic acid groups may be introduced into the resin structure. This type of resin is only capable of exchanging hydrogen for the cations of weak acids and its efficiency decreases as the pH of the solution decreases because the ionisation of the weakly acidic —COOH groups is depressed at low pH.

(*d*) *Weak Anion Exchangers*

These usually contain amine groups and the disappearance of anions from the solution in contact with such exchangers may be explained on the basis of an acid adsorption mechanism, i.e.

$$RNH_2 + HX = RNH_3X$$

or by the following exchange mechanism

$$\begin{array}{c} RNH_2 + H_2O \rightleftharpoons RNH_3^+ + OH^- \\ \qquad\qquad\qquad\quad \Big\updownarrow X^- \\ \qquad\qquad\quad RNH_3X + OH^- \end{array}$$

APPLICATIONS OF ION EXCHANGERS

1. *Demineralisation of Water*

This is the most important use of ion exchange. It involves the removal of cations and anions from impurities in water and their replacement by hydrogen and hydroxyl ions respectively. The exchange reactions may be carried out in two separate stages: the first involves passage of the water through a column packed with beads of a cation exchanger; the second involves a similar treatment with an anion exchanger. Each resin can be regenerated easily when necessary. However, it is also possible to carry out the exchange reactions in a single process, which involves the use of a mixed bed of cation and anion exchangers. The different resins must then be separated before regeneration of each can be carried out. Separation is achieved by forcing water into the bottom of the column, when the lighter anion exchanger beads rise above the cation exchanger beads.

This method of water purification is cheaper than distillation and for many purposes the product is superior to distilled water.

Although considerable evidence indicates that certain ion exchange resins will remove pyrogens from water the BP does not allow the use of demineralised water as Water for Injections (Gunn and Carter, 1965).

2. *Other Purification Processes*

(*a*) Ionic impurities can be removed from solutions of non-electrolytes such as glycerol, ethanol, and sugars. Similar impurities can also be removed from solutions of macromolecules such as proteins.

(*b*) The collection and purification of ionic compounds of pharmaceutical importance may be achieved by using ion exchange techniques; for example, streptomycin is a cation and carboxylic acid cation exchange resin may be used in its collection and purification. The efficiency of the process is usually increased by the use of resins that adsorb streptomycin selectively (Nachod and Schubert, 1956). Similar resins are also used for recovering and purifying neomycin, other antibiotics, and some vitamins.

(*c*) Concentration of radioactive salts from waste liquids into ion exchange resins is useful in decontaminating such liquids and aids disposal of the radioactive material.

3. *Analytical Processes*

(*a*) *Ion Exchange Chromatography.* Solutes with different acidic or basic strengths can be separated by chromatographic techniques using columns packed with beads of ion exchangers.

(*b*) *Other Analytical Uses.* Ion exchangers may be used as collectors for ions present in dilutions that are too low for accurate determination. They may also be used to remove interfering ions before analytical procedures are carried out.

4. *Formulation of Prolonged Release Dosage Forms*

The exchange of ionic forms of drugs held by resins with ions present in body fluids is a relatively slow process and the rate of exchange can be controlled by variation in the properties of the resin. Such a mechanism is therefore used in formulating dosage forms that are intended to release the drug into the body over a prolonged period.

5. *Medical Uses*

(*a*) *As Antacids.* The removal of acid by weak anion exchangers is achieved without causing side effects such as constipation or diarrhoea that often result after treatment with conventional antacids.

(*b*) *As Laxatives and Anti-obesity Agents.* The marked swelling of weakly crosslinked ion exchange resins allows them to act as bulk laxatives. Such resins are also used as anti-obesity agents because on swelling they create a feeling of fullness in the patient.

(*c*) *Control of Salt Intake.* A low salt intake by the body is required in the treatment of various oedemas and hypertensive conditions. Cation exchangers allow the use of more salt in the diet of patients suffering from these disorders and the diet can therefore be made more palatable than those that have no or low salt contents

The potassium level in the blood can also be regulated by cation exchange resins, when necessary, in disorders where the kidneys do not function properly.

FACTORS INFLUENCING RATE OF CHEMICAL REACTIONS

1. CATALYSIS

The rate of change in the concentrations of products and reactants in a chemical reaction may be altered by the presence of a catalyst. The latter takes part in the reaction by forming an intermediate complex which subsequently decomposes into the products and regenerated catalyst.

Catalysts may dissolve in the reaction medium and so result in what is termed homogeneous catalysis, or they may remain as solids, when the reactants are adsorbed on to their surfaces. This latter process is referred to as heterogeneous catalysis.

Homogeneous catalysis includes the effect of pH on the hydrolysis of many compounds. For example, the rate of decrease in concentration of sucrose in acid solution that results from hydrolysis is approximately proportional to the concentration of hydrogen ions. This effect of pH is important in the stabilisation of many pharmaceutical preparations since an optimum pH for maximum stability of an active ingredient often exists. For example, Injection of Ergometrine BP (1968) is adjusted to pH 3 by the addition of maleic acid. In addition, the containers used for many official preparations must comply with a test for the limit of alkalinity (*see* Gunn and Carter, 1965).

Chemisorption is now accepted as being necessary in the process of heterogeneous catalysis. The strong adsorption of reactants at active sites on the surfaces of catalysts produces a local concentration of reactants in a particular orientation. In addition, the bonds within the chemisorbed molecules are weakened. All these factors will affect the rate of reaction. Heterogeneous catalysts are often used in hydrogenation reactions and these catalysts include palladium, platinum, and nickel.

Since heterogeneous catalysis occurs at the surface of a solid catalyst it is influenced by the surface area of the solid and by the presence of other compounds, which may also be adsorbed strongly, and so poison the surface.

Enzymes are usually regarded as heterogeneous catalysts although they are often soluble in water. However, adsorption of the substrate on to the enzyme macromolecules is involved in the catalysed reaction. The first stage of the reaction may be considered to be the formation of a complex, ES, between the enzyme, E, and substrate, S, and this is followed by decomposition of the complex into products plus enzyme. This series of stages may be represented by Eqn (6.32), which shows that the first stage is reversible but the second is not.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + \text{products} \tag{6.32}$$

The kinetics of such reactions catalysed by enzymes is interpreted by the Michaelis-Menten theory that is named after the original workers in this field. This theory is based on the fact that the second stage in Eqn (6.32) is the rate determining stage, since it is much slower than the first. The changes in concentrations of substrate and products are therefore dependent on the concentration of the complex *ES*, since this determines the rate of the second stage. However, the concentration of the complex cannot be measured and must therefore be defined in other terms that are measurable. This treatment leads to the Michaelis-Menten Eqn (6.33), which expresses the rate of decomposition of the substrate ($-\mathrm{d}s/\mathrm{d}t$) in terms of the substrate concentration S at any time, a constant, V, which is equal to the maximum rate that is possible (i.e. when all the enzyme is involved in complex formation), and a constant, K_m, that is known as the Michaelis constant. In fact this constant is the ratio of the rate constant for the reverse reaction in the first stage of Eqn (6.32) to the rate constant for the forward reaction in the same stage (i.e. $K_\mathrm{m} = k_{-1}/k_1$).

$$v = -\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{K_\mathrm{m}VS}{1 + K_\mathrm{m}S} \tag{6.33}$$

Equation (6.33) does not permit easy determination of the constants K_m and V, and various forms of it have been suggested to facilitate such determination. For example, the Lineweaver-Burk equation, Eqn (6.34), indicates that a straight line will be obtained if $1/v$ is plotted against $1/S$.

$$\frac{1}{v} = \frac{K_\mathrm{m}}{V}\frac{1}{S} + \frac{1}{V} \tag{6.34}$$

The slope of this line is equal to K_m/V and the intercept at $1/S = 0$ is equal to $1/V$ so that these constants can be determined.

2. TEMPERATURE

The effect of temperature on a rate constant, k, is indicated by the Arrhenius equation, Eqn (6.35)

$$k = Z\,\mathrm{e}^{-E/RT}$$

or

$$\log k = \log Z - \frac{E}{2{\cdot}303R}\frac{1}{T} \tag{6.35}$$

where Z is a constant that is termed the frequency factor, E is the energy of activation, R is the gas constant, and T is the thermodynamic temperature.

Different interpretations of the constant Z are given by the collision theory and the transition state theory of reaction rates. In addition, there is a slight difference in the interpretation of the constant E, and the student is recommended to read accounts of these theories given in the books listed at the end of this chapter.

It can be seen from Eqn (6.35) that a straight line will be obtained from a graph of $\log k$ against $1/T$ as shown in Fig. 6.2. The constants E and Z may be determined from the slope and intercept of this line, which are equal to $-E/2{\cdot}303R$ and $\log Z$, respectively. It should be borne in mind that temperature is decreasing when $1/T$ is increasing so that the higher temperatures lie towards the left-hand side of the horizontal axis of the graph.

3. LIGHT

Light energy may be absorbed by certain molecules, which then become sufficiently activated for participation in a reaction. Only frequencies in the visible and ultra-violet regions can provide sufficient energy to cause photochemical reactions. Since the energy for activation is provided by light in these reactions the rates of the latter are independent of temperature. However, it is often difficult to separate photochemical reactions from thermal reactions,

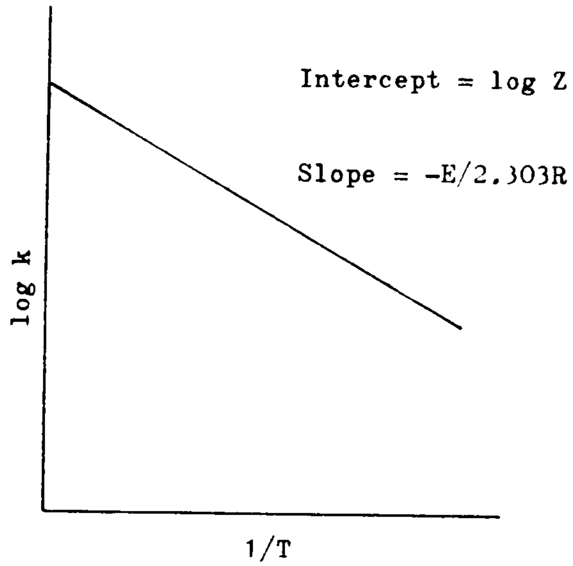


Fig. 6.2 Graph showing plot of the Arrhenius equation since thermal effects may: (*a*) be involved in the subsequent chain of reactions that often follows a photochemical reaction, or (*b*) arise from the heating effect produced by the increase in kinetic energy of molecules that have collided with others that have absorbed some light energy.

4. SOLVENT

The effects of solvents on the rate constants of chemical reactions are complex. Their influence on the ability of solutes to dissociate into ions is particularly important in reactions that involve these ions.

Stability Testing

Pharmaceutical compounds and preparations often exhibit chemical or physical instabilities. The deterioration that results from such instabilities may lead to:

(*a*) a reduced activity of the compound or preparation,
(*b*) the formation of toxic reaction products, or
(*c*) the formation of an inelegant or unusable product, e.g. a broken emulsion.

In addition, microbial contamination may be unacceptable or lead to deterioration.

Stability testing is therefore carried out to ensure that deterioration does not exceed an acceptable level in order to:

(*a*) ensure the safety of the patient,
(*b*) maintain the activity of the product, and
(*c*) maintain sales, because a deteriorated product is either unusable or a poor advertisement.

The common causes of chemical instability in pharmaceutical materials involve hydrolysis or oxidation.

Hydrolysis

This is particularly important in systems containing water, e.g. solutions, suspensions, and emulsions. It is also important in the deterioration of ingredients contained in solid dosage forms, since water may enter as vapour from the atmosphere or as water of crystallisation in other ingredients.

The following list gives some examples of chemical instabilities that result from hydrolysis.

1. *Penicillin and Derivatives*

These are particularly susceptible to hydrolysis and the reaction is often catalysed by the enzyme penicillinase. The antibiotic activity decreases on hydrolysis and contact between penicillin preparations and moisture should therefore be avoided whenever possible. Liquid dosage forms (e.g. syrups and suspensions) are prepared immediately before use by adding water to the dry ingredients, and unused liquid should be discarded after a relatively short period (e.g. 1 to 2 weeks) if stored at room temperature. Dry preparations may be stored for considerably longer periods; e.g. 2 to 3 years at room temperature for crystalline sodium benzylpenicillin if the moisture content is kept below 0·5 per cent (Johnson, 1967).

2. *Aspirin*

Hydrolysis of this acid to a mixture of acetic and salicylic acids occurs readily and the powdered drug should therefore be protected against atmospheric water vapour. In fact, as Johnson (1967) has pointed out, if aspirin were introduced at the present time as a new drug it would probably be regarded as unsuitable because of its marked instability.

3. *Alkaloids*

Solanaceous alkaloids that contain ester linkages are susceptible to hydrolysis. It is usually assumed that if hydrolysis of the alkaloid occurs in aqueous solution then it is likely to occur in the crude drug if the moisture content increases. Drying the crude drugs excessively may produce a hygroscopic material, which is undesirable, so that an optimum level of drying is usually aimed at.

Other examples of instabilities that arise from hydrolysis are given by Macek (1970), Garrett (1967) and Johnson (1967). In addition, Johnson discusses the physical deterioration that may be caused by moisture.

METHODS OF PROTECTION AGAINST HYDROLYSIS

The obvious method of protection is to prevent contact between the material and water. This method is mainly concerned with protecting the dry material against water vapour and therefore involves control of atmospheric humidity during preparation, purification, and packing, and satisfactory design of the final container so that adequate protection against the entry of water vapour from the atmosphere is provided. Extra protection is also achieved by including porous envelopes containing desiccants in the containers.

Prevention of contact is impossible in liquid dosage forms that contain water. The method of protection is therefore concerned with reducing the rate of hydrolysis in the system. Since many hydrolytic reactions are catalysed by acids and bases, one method of achieving a reduction in the rate of decomposition is by adjusting the pH to an optimum level at which the catalytic effect is at a minimum.

It has also been shown that the formation of molecular complexes between the hydrolysable substance and a second component may inhibit hydrolysis. For example, Higuchi and Lachman (1955) have found that benzocaine is protected against hydrolysis by complexation with caffeine.

Since hydrolysis is a solvolytic reaction the rate at which it causes decomposition is affected by the solubility of the hydrolysable substance. A decrease in solubility caused by a change in pH, or the use of insoluble derivatives, will therefore provide means of reducing the decomposition; for example, insoluble chlorothiazide is stable in neutral aqueous suspension but the soluble sodium salt that is formed in alkaline solution is readily hydrolysed.

Oxidation

Oxidation and reduction involve the loss and gain of electrons, respectively. Many oxidation reactions result from the presence of atmospheric oxygen but the required loss of electrons may sometimes occur even when oxygen is absent; e.g. in reactions between oxidising and reducing agents. However, the decomposition of medicinal compounds usually involves molecular oxygen. Such oxidations are usually termed autoxidations because they occur spontaneously under normal conditions, and often involve free radicals. The latter, which contain one or more unpaired electrons, are particularly reactive and the products of their reactions are often free radicals themselves, hence chain reactions are initiated and proceed until the remaining free radicals are destroyed or rendered less active.

AUTOXIDATION OF UNSATURATED FATS AND OILS

This type of decomposition may be used to illustrate the mechanism of autoxidation and the methods of obtaining protection against its effects. The decomposed products usually develop objectionable odours and tastes and are said to be rancid.

The mechanism of a chain reaction may be considered in terms of the following stages.

(a) Chain Initiation

The presence of unsaturated C=C linkages has an activating effect on adjacent —CH_2— groups in a hydrocarbon chain. The reaction between molecular oxygen and an unsaturated hydrocarbon is considered to involve the abstraction of a hydrogen free radical ($\dot{H}$) from an activated methylene group and results in the formation of a hydrocarbon free radical. This reaction is illustrated by the following equation, in which the free radicals are indicated by a dot over their chemical formulae.

$$-CH{=}CH{-}CH_2- \xrightarrow[-\dot{H}]{O_2} -CH{=}CH{-}\dot{C}H-$$

Chain initiation may also occur from the ready dissociation of so-called initiators (e.g. benzoyl peroxide) into free radicals or from the effects of heat, light or ionising radiation on some ingredients of an oil, e.g. (*i*) dissociation of a peroxide (R = hydrocarbon group)

$$R.OOH \rightarrow R\dot{O} + \dot{O}H$$

and (*ii*) photo-oxidation of a ketone (R and R′ = hydrocarbon groups)

$$R.CO.R' \rightarrow R.C\dot{O} + \dot{R}'$$

(b) Chain Propagation

The free radicals produced in the previous stage are involved in further reactions that yield other free radicals. Peroxides are formed as intermediates, but the concentration of these compounds eventually decreases when their rate of decomposition exceeds their rate of production. The use of peroxide values as an indication of the extent of an autoxidation process may therefore be misleading. The reactions involved in the propagation stage can be summarised as follows:

$$\left.\begin{array}{r} \dot{R} + O_2 \rightarrow R\dot{O}_2 \\ R\dot{O}_2 + RH \rightarrow ROOH + \dot{R} \\ \text{hydroperoxide} \end{array}\right\} \begin{array}{l} \text{These reactions} \\ \text{increase the} \\ \text{rate of peroxide} \\ \text{formation} \end{array}$$

$$\left.\begin{array}{r} ROOH \rightarrow R\dot{O} + \dot{O}H \\ 2\,ROOH \rightarrow R\dot{O} + R\dot{O}_2 + H_2O \\ ROOH \rightarrow \text{non-radical products} \end{array}\right\} \begin{array}{l} \text{These reactions} \\ \text{increase the} \\ \text{rate of peroxide} \\ \text{disappearance.} \end{array}$$

(c) Chain Termination

(i) *Self Termination.* This involves reaction between two free radicals and the production of inactive (i.e. non-free radical) products.

(ii) *Chain Breaking Termination* This involves reaction between free radicals and compounds that are known as chain inhibitors and results in the formation of stable and comparatively unreactive free radicals, e.g.

$$\dot{R} + \underset{\substack{\text{chain}\\ \text{inhibitor}}}{IH\cdot} \rightarrow RH + \underset{\substack{\text{stable}\\ \text{free radical}}}{\dot{I}}$$

Autoxidations may be followed by measuring the amount of oxygen taken up during a reaction. The results of such a measurement are illustrated in Fig. 6.3 which shows that three separate stages may be defined in the reaction. The first of these stages, *AB*, corresponds to an induction period, in which little oxygen is used and the chain initiation reaction occurs. This is followed by propagation, *BC*, where the oxygen uptake increases rapidly until termination begins to occur and the rate of uptake decreases slowly (beyond *C*).

The kinetics of the overall process are affected by the effects of heat and light, both of which increase the rate of autoxidation, by the nature of the fatty material, and by the presence of certain metals and salts, which probably catalyse the decomposition of hydroperoxides.

METHODS OF PROTECTION AGAINST OXIDATION

Reducing agents are often used to provide protection against atmospheric oxygen and oxidising agents. Sodium metabisulphite is commonly used for this purpose in many injections (e.g. Injection of Adrenaline BP).

Antioxidants that are effective against atmospheric oxygen are often used to provide protection against autoxidation. These compounds may influence the chain initiation or propagation stages, or both, and this provides a basis for their classification.

(*a*) *Chain Initiation Suppressors*

(i) *Ultra-violet Stabilisers.* These absorb ultraviolet radiation and dispose of the energy without forming free radicals. They include phenyl salicylate and 2-hydroxybenzophenone.

(ii) *Hydroperoxide Destroyers.* These react with hydroperoxides to form non-free radical products. They are mainly organic compounds of phosphorus and sulphur and their toxicity precludes any wide use in pharmacy.

(iii) *Metal Deactivators.* These form complexes with metals, which are then unavailable for catalysing peroxide decomposition. They include citric acid, ethylenediamine tetra-acetic acid (EDTA) and 8-hydroxyquinoline. Care must be taken in their use because the metal complex may be even more effective in catalysing peroxide decomposition than the original metal.

(*b*) *Propagation Suppressors*

These combine with free radicals to produce stable products; i.e. they act as chain terminating agents

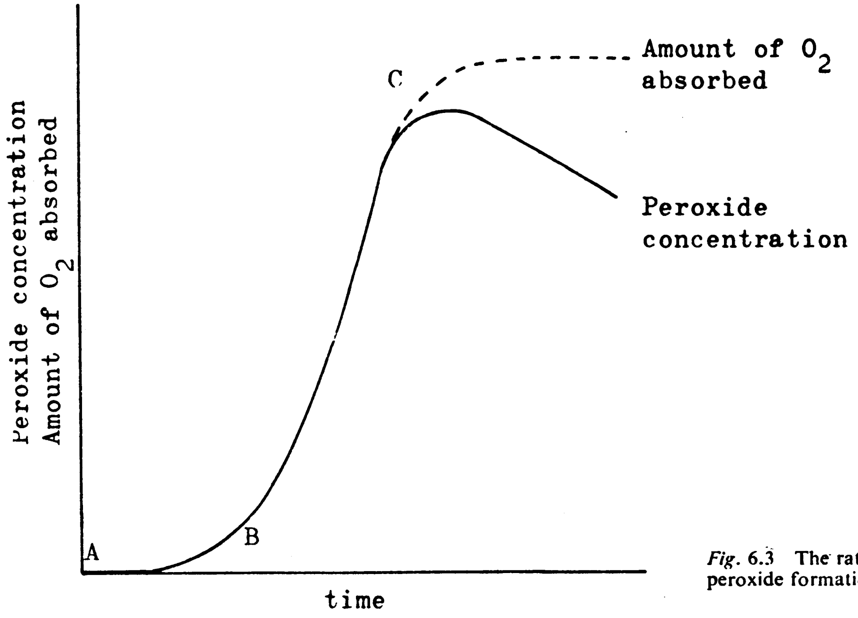


Fig. 6.3 The rates of oxygen absorption and peroxide formation in an autoxidation

They include phenolic and secondary amino compounds, the former type being in common use in pharmaceutical preparations; e.g. butyl hydroxyanisole (BHA), nor-dihydroguaiaretic acid (NDGA), and gallic acid esters such as propyl and n-butyl gallates.

Two or more different antioxidants may be more effective than would be expected on the basis of their individual activities. This is known as synergism, and in general the best mixtures are those that contain an initiation suppressor and a propagation suppressor. However, the characteristics of the system that requires protection, or its intended storage conditions, may indicate more precise combinations; for example, if prolonged exposure to light is expected then an ultra-violet stabiliser plus a propagation suppressor should be used.

The effectiveness of antioxidants may be determined from their effects on the oxygen uptake of the protected substrate or on the rate of peroxide formation. An increase in the induction period should be obtained with an initiation suppressor and a decrease in the subsequent rate of reaction should be obtained with a propagation suppressor.

Other Types of Chemical Decomposition

Reaction between various ingredients of a formulation may be responsible for a loss in activity or other undesirable change. Examples of incompatibilities that may arise in extemporaneous preparations are given by Gunn and Carter (1965). Reactions occurring in new formulations or arising from the leakage of substances from container materials are less easy to predict.

Isomerisation of a compound to a less active structure may occur, especially in optically active compounds, where the optical activity may be lost during the racemisation which is accompanied by a decrease in pharmacological activity.

Determination of the Cause of Decomposition

The route of a particular decomposition and therefore the factors that cause it may be indicated from a consideration of:

(*a*) the chemical structure of the decomposing substance,
(*b*) the chemical structure(s) of the product(s) of decomposition,
(*c*) the properties of other ingredients, possible impurities and container material,
(*d*) the storage conditions,
(*e*) the appearance and odour of the decomposed preparation, and
(*f*) a knowledge of the behaviour of similar compounds.

More information is obtainable from a series of tests such as that outlined below.

Solution of drug sealed in ampoules

	Series A Contains O_2 above solution in ampoules	Series B Contains N_2 above solution in ampoules
1. Expose to light	A1	B1
2. Heat	A2	B2
3. Store in dark	A3	B3

Results

(*a*) Decomposition in all cases indicates hydrolysis.
(*b*) Decomposition in all of series A but not in series B indicates oxidation.
(*c*) Decomposition in A1 but not A3 indicates a photochemical reaction.
(*d*) Decomposition in A2 and B2 but not in A1, A3, B1 and B3 indicates a thermal decomposition.

Similar series of tests may be carried out at different pH's to indicate the effect of hydrogen and hydroxyl ion concentrations on decompositions.

Accelerated Stability Testing

Instabilities in modern formulations are often detectable only after considerable storage periods under normal conditions. To reduce the time required to obtain information, various tests that involve storage of the products under conditions that accelerate decomposition have been introduced. The objectives of such accelerated tests may be defined as

1. the rapid detection of deterioration in different initial formulations of the same product—this is of use in selecting the best formulation from a series of possible choices;
2. the prediction of shelf-life, which is the time a product will remain satisfactory when stored under expected or directed storage conditions; and
3. the provision of a rapid means of quality control, which ensures that no unexpected change has occurred in the stored product.

All these objectives are based on obtaining a more rapid rate of decomposition by applying to the product a storage condition that places a higher stress or challenge to it when compared with normal storage conditions. However, the use of this basic method depends on the particular objective required. For example, in the first objective the best formulation from a series of possible choices is the one that exhibits the least amount of decomposition in a given time under the influence of a reasonably high stress. The results of such a test are illustrated in Fig. 6.4(*a*).

The second objective is achieved by using the results obtained from an accelerated test to predict the amount of decomposition in a product after a longer period of storage under normal conditions. This is illustrated in Fig. 6.4(*b*) where the amount of decomposition X obtained after the short time t_1 is used to predict the value of Y after time t_2.

The use of accelerated tests in achieving the third objective is illustrated by Fig. 6.4(*c*) which shows that a single measurement taken after a given time t should fall below an acceptable limit of decomposition for a product subjected to the challenge involved in the test.

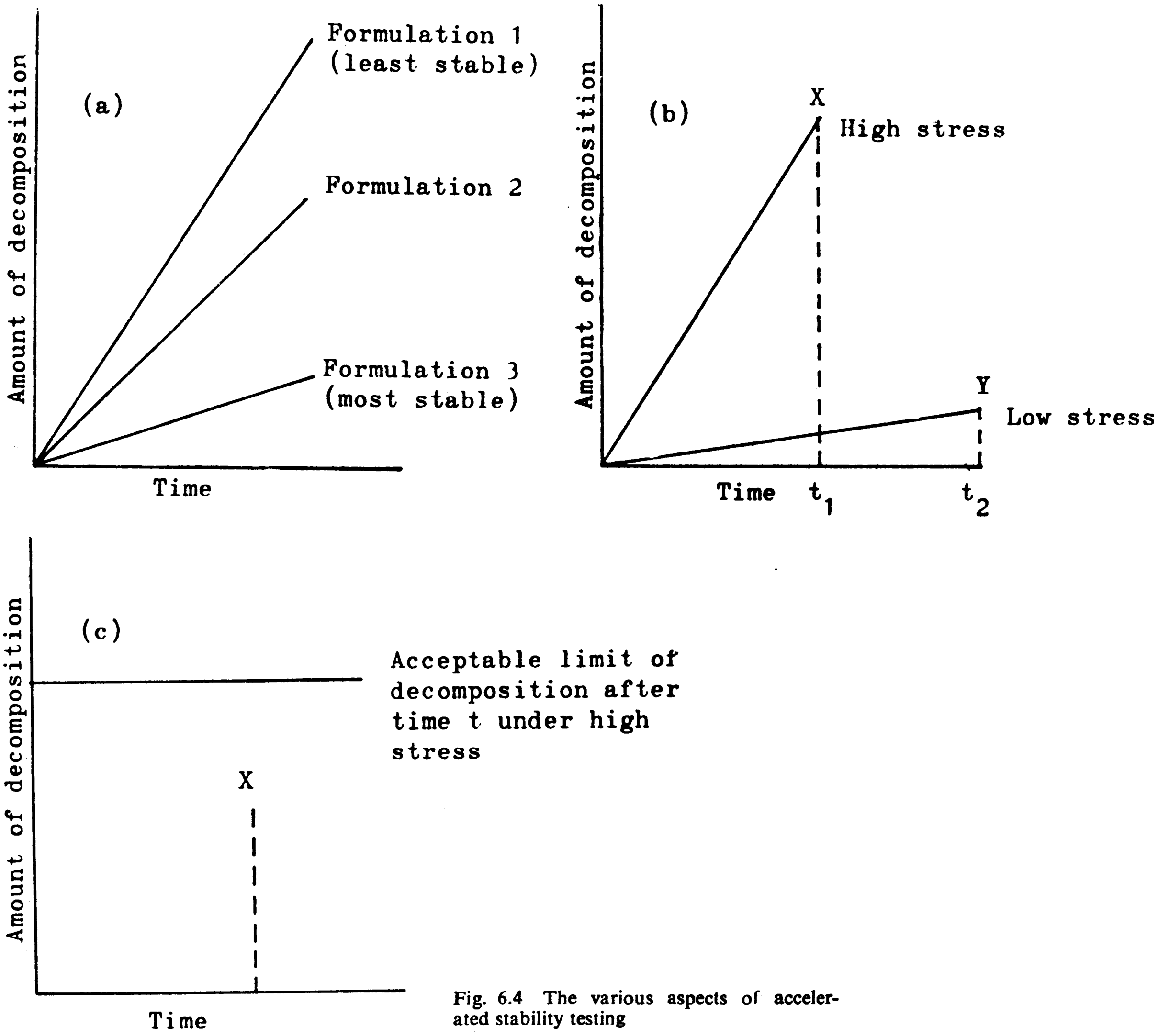


Fig. 6.4 The various aspects of accelerated stability testing

Common High Stresses or Challenges

(*a*) *Temperature*

An increase in temperature causes an increase in the rate of chemical reactions. The products are therefore stored at temperatures greater than room temperature. The nature of the product often determines the range covered in the accelerated test. Samples are removed at various time intervals and the extent of decomposition is determined by analysis. Sensitive analytical methods should be used in all stability tests of this nature since small changes may be detected after very short storage periods.

The effects caused by high temperatures should not be confused with those that arise from the effect of low humidity. Such confusion is possible because the relative humidity inside a high temperature storage cabinet will be lower than that in the room. This low humidity causes loss of moisture, which may lead to apparent increases in the concentration of ingredients. If these concentration changes are not allowed for in subsequent analyses decomposition may be unsuspected.

(*b*) *Humidity*

Storage of the product in atmospheres of high humidity will accelerate decompositions that result from hydrolysis. Marked acceleration will be obtained if the 'naked' product (i.e. not enclosed in a container) is subjected to these tests, which usually indicate the minimum humidity tolerated by the product without undue decomposition, and are therefore useful in determining the degree of protection that should be afforded by a container.

(*c*) *Light*

A source of artificial light is used to accelerate the effects of sunlight or sky light. The source should emit a similar distribution of radiant energy to that in sunlight because photochemical reactions involve the absorption of light of definite wavelengths. Daylight fluorescent lamps provide a satisfactory source, and banks of such lamps may be used to accelerate the effects of light. However, although these lamps do not have a marked heating effect the use of glass plates to reduce such an effect is recommended, otherwise it is difficult to separate the accelerated decomposition caused by light from that caused by increased temperatures.

The Prediction of Shelf-life

This is the second of the objectives that were listed previously and it is based on the application of the Arrhenius equation, Eqn (6.35), which indicates the effect of temperature on the rate constant, k, of a chemical reaction. Figure 6.2 (*see* p. 97) shows that a graph of $\log k$ versus the reciprocal of thermodynamic temperature, $1/T$, is a straight line. If the slope of this line is determined from the results of accelerated tests at high temperatures it is possible to determine the value of the rate constant at other temperatures (e.g. normal room temperature) by extrapolation. Substitution of this value of k into the appropriate order of reaction (i.e. the rate equation that applies to the reaction involved in the particular decomposition) allows the amount of decomposition after a given time to be calculated. As pointed out, this approach involves a knowledge of the order of the reaction involved and preliminary experiments such as those outlined on p. 92 are therefore necessary to determine this order.

Several difficulties and limitations are involved in this aspect of accelerated stability testing. First, as in all accelerated tests, there is the possibility that the application of high stresses may cause reactions that would not take place under the lower stresses associated with normal storage conditions. Secondly, the uncertain conditions defined by the term 'normal storage conditions' introduces a difficulty when attempting to forecast the shelf-life of a product. Unless the storage conditions are defined precisely on the container, then allowance should be made for variations in the conditions likely to be encountered under normal storage. Attempts to allow for such a contingency often involve accepting the shortest shelf-life for the range of conditions likely to be encountered. The climate of the country in which a product is to be marketed is particularly important in defining this range.

Decompositions in formulated products often proceed via a complex reaction series and may involve simultaneous, consecutive, or chain reactions, because the formulated products themselves are complex systems. In addition, the order of a reaction may change after a certain time. Predictions of the extent of decomposition at future times are then impracticable and prolonged tests under normal storage conditions must be carried out.

In spite of these difficulties the application of accelerated testing to pharmaceutical products is often useful, and predicted shelf-lives are sufficiently accurate. Statistical methods of designing such tests have therefore been reported, which allow the selection of the number of replicates, sampling times and other factors involved in the tests to be made on a logical basis for attaining the required degree of accuracy without wasting time on unnecessary experimentation (Tootill, 1961; Jones and Grimshaw, 1963).

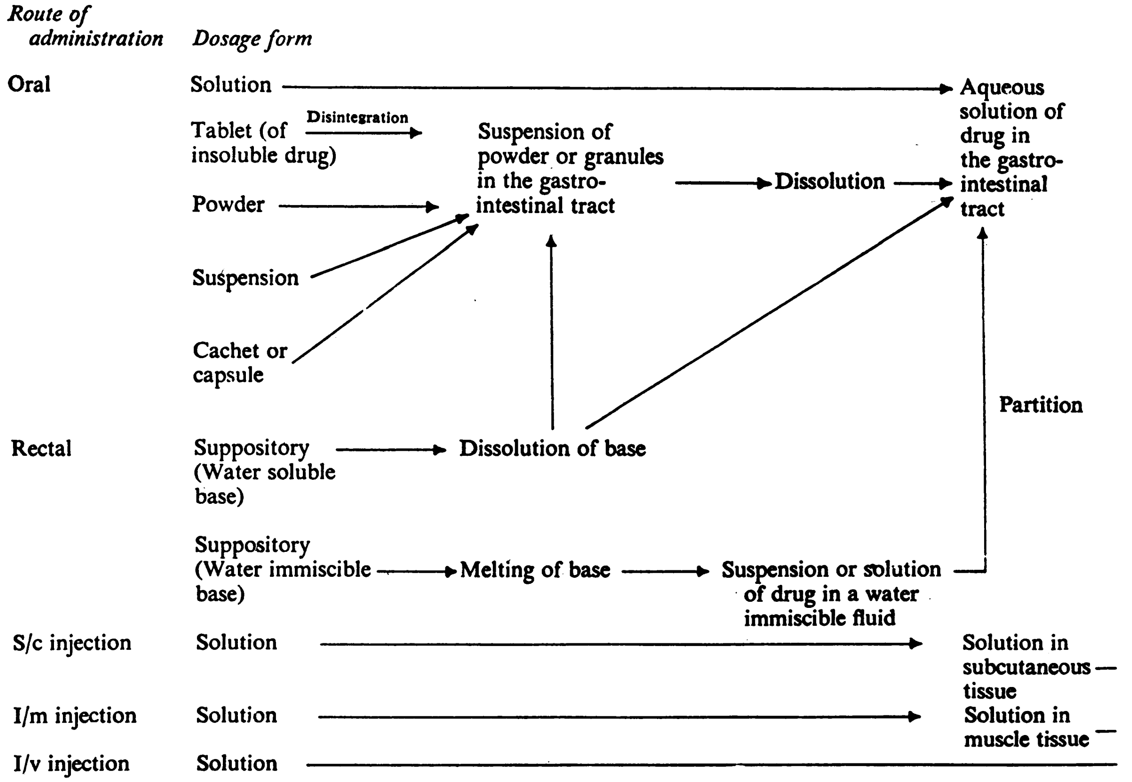


Fig. 6.5 Diagram illustrating some of the factors that influence

KINETICS OF DRUG TRANSPORT *IN VIVO*

The purpose of administering a drug to a patient is to produce a particular response. This response is achieved when enough of the drug is concentrated at a certain site in the body. The number of molecules of the active form of the drug that reach the site where they can exert their effect (locus of action or receptor site) and the rate at which these molecules arrive there depend on several factors. Some of these factors will be generally applicable but others will be determined by the route of administration of the drug and the properties of the dosage form.

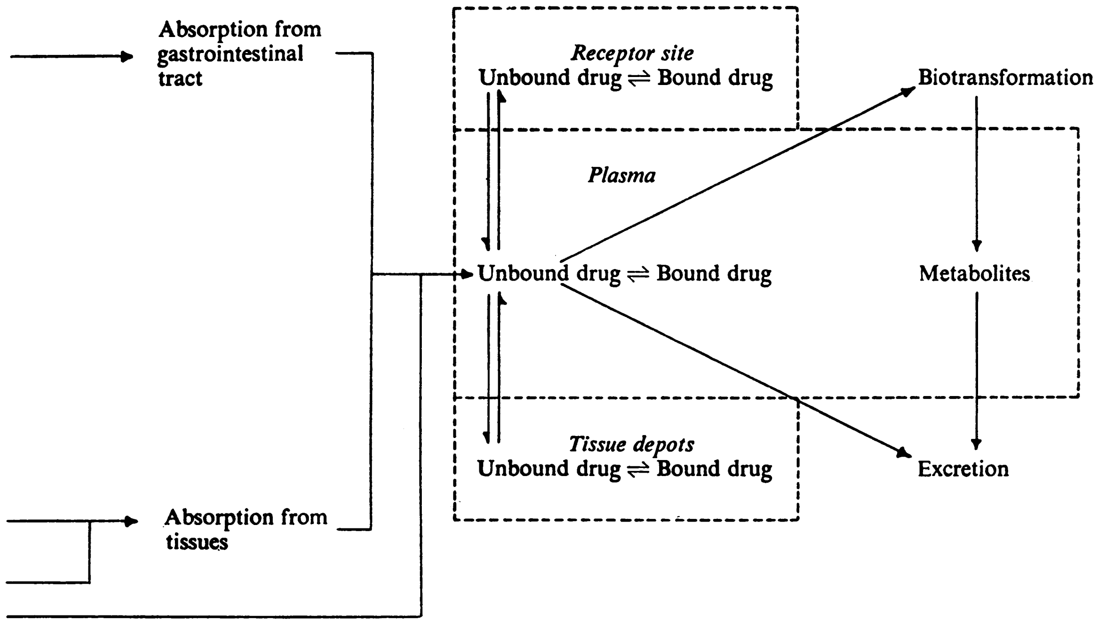
Systemically Active Drugs

These are drugs that after absorption into the body are carried to their site of action by the blood. Figure 6.5 illustrates some of the *in vivo* processes that are involved in the transport of such drugs after administration as various dosage forms. It should be realised that all drugs except those injected intravenously must be transported across some type of body membrane before they can enter the plasma. Transport into other tissues, or excretion, again involves passage through membranes. The various body membranes constitute the major barriers to the movement of drugs *in vivo* and it is therefore necessary to consider the general structure of such membranes and the various mechanisms involved in the transport of substances across them.

STRUCTURE OF THE CELL MEMBRANE

Although a variety of membranes exist in the body (e.g. cytoplasmic membrane surrounding a single cell, sheets of tissue several cells in thickness around certain organs) the same transport mechanisms are involved in all cases. It is therefore useful to consider the structure of a cell membrane and show how this structure is penetrated by different substances.

A cell membrane is regarded as having two layers of lipid molecules orientated with their polar groups



the concentration of a systemically active drug at its receptor site

facing outwards, as shown in Fig. 6.6. These molecules include lecithin, sphingomyelin, cephalin, and cholesterol. The bimolecular layer is stabilised by a layer of unfolded protein molecules, which cover each side. This type of structure therefore constitutes a lipid barrier which separates the aqueous intra- and extra-cellular liquids. The thickness of the membrane is about 75×10^{-7} mm, and the structure is perforated by water-filled pores, which usually have an effective radius of the order of 4×10^{-7} mm. However, in certain specialised membranes this pore size may be about 10 times larger (e.g. in glomeruli of the kidney).

Mechanisms of Drug Transport Across Cell Membranes

1. PASSIVE DIFFUSION THROUGH LIPID BARRIERS

This is a common mechanism that is involved in the transport of drugs *in vivo*. Much of the work in this field has been reviewed by Brodie (1964).

The process of passive diffusion through a lipid barrier initially involves partition of a drug between the aqueous phase on one side of the membrane and the lipid phase that constitutes the membrane. This stage is followed by diffusion across the membrane and a second partition between the lipid phase and the aqueous liquid on the other side of the membrane. The overall rate of transport will therefore be affected by factors that influence these consecutive processes, which are illustrated in Fig. 6.7, where the *k* values refer to the rate constants of the individual processes. In any reaction that involves consecutive stages, the overall rate of transfer is determined by the rate of transfer in the slowest stage. However, the relative rates of the three stages shown in Fig. 6.7 will depend on the nature of the drug molecules; for example, a molecule that is more soluble in water than in the lipid phase will be taken into the latter phase at a slow rate only but will be released rapidly at the other side after it has diffused across the membrane, hence process (1) in Fig. 6.7 will be the rate-determining

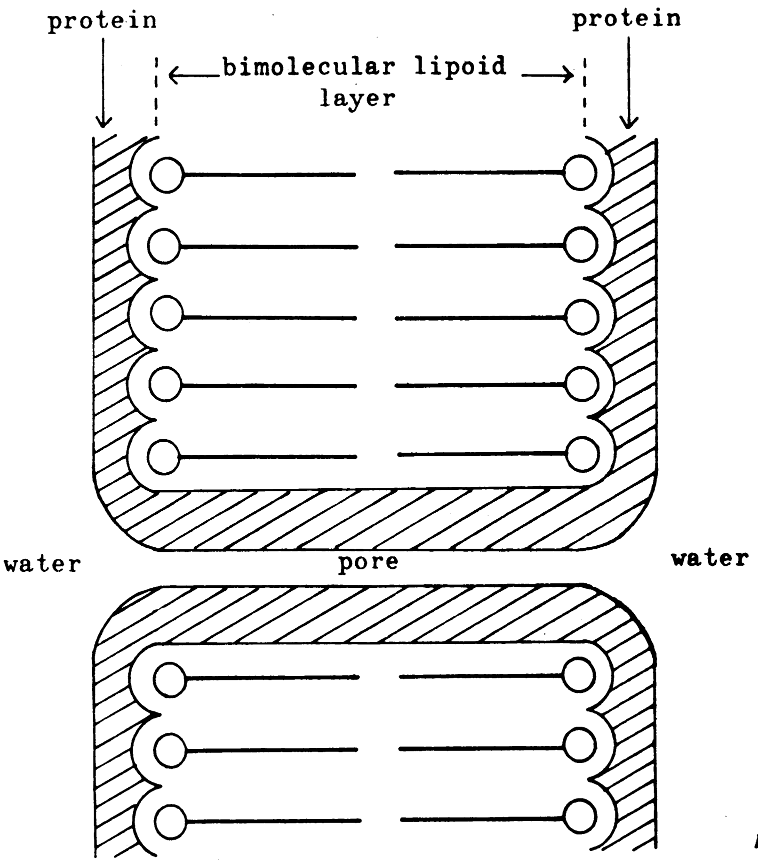


Fig. 6.6 Structure of the cell membrane

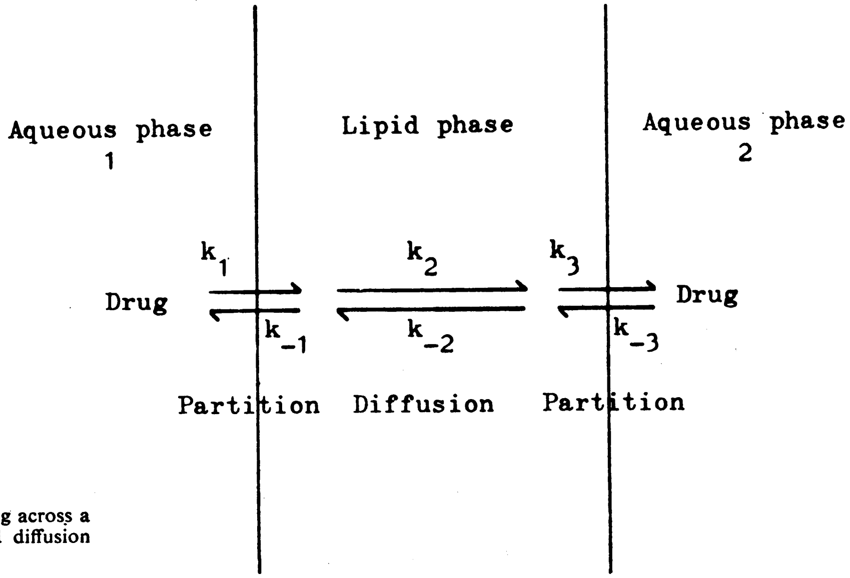


Fig. 6.7 Transport of a drug across a membrane by partition and diffusion processes stage. Conversely, a substance that is more soluble in the lipid phase than in water will enter the lipid stage rapidly but will be released with difficulty at the other side, hence process (3) will now be the rate-determining stage. In compounds with intermediate partition coefficients, the diffusion process (2) may become the limiting stage, when the rate of transfer will be indicated by Fick's law (Eqn 5.5). However, the thinness of the membrane is low enough for this diffusion process to be relatively rapid, when the partition processes may often be the rate-determining stages, as mentioned above. It should be pointed out that the diffusion stage is almost always the rate-determining process in many *in vitro* models that have been used to represent drug transport by this mechanism because the thickness of the organic solvent layer that is used in place of a lipid membrane creates a long diffusion pathway.

It has been assumed in the previous discussion that the drug exists in aqueous solution as a single species with a definite partition coefficient for distribution between water and the lipid phase of the membrane. However, most drugs are weak electrolytes and exist in aqueous solution as a mixture of ionised and unionised forms. The latter are much more soluble in the lipid phase than the water soluble ions so that body membranes tend to be preferentially permeable to unionised forms of drug molecules. In other words, the rate constant k_1 in Fig. 6.7 is extremely low for ions but is much more appreciable for the corresponding unionised molecules.

pK$_a$ *Values*

The fraction of drug that is ionised in solution is given by the dissociation constant of the drug. Such dissociation constants are conveniently expressed in terms of pK$_a$ values for both acidic and basic drugs.

The equilibrium between unionised molecules of an acidic drug (HA) and the ions produced by dissociation may be represented by Eqn (6.36)

$$\mathrm{HA} \rightleftharpoons \mathrm{H^+} + \mathrm{A^-} \qquad (6.36)$$

The dissociation constant K$_a$ of this drug is therefore defined by Eqn (6.37), where the square brackets represent concentration terms of the species enclosed within them.

$$K_a = \frac{[\mathrm{H^+}][\mathrm{A^-}]}{[\mathrm{HA}]} \qquad (6.37)$$

Taking logarithms of both sides of Eqn (6.37) yields

$$\log K_a = \log [\mathrm{H^+}] + \log [\mathrm{A^-}] - \log [\mathrm{HA}]$$

and the signs in this equation may be reversed to give Eqn (6.38)

$$-\log K_a = -\log [\mathrm{H^+}] - \log [\mathrm{A^-}] + \log [\mathrm{HA}] \qquad (6.38)$$

The symbol pK$_a$ is used to represent the negative logarithm of the acid dissociation constant K$_a$ in the same way that pH is used to represent the negative logarithm of the hydrogen ion concentration [H$^+$], and Eqn (6.38) may therefore be rewritten as Eqn (6.39)

$$\mathrm{p}K_a = \mathrm{pH} + \log [\mathrm{HA}] - \log [\mathrm{A^-}] \qquad (6.39)$$

or

$$\mathrm{p}K_a = \mathrm{pH} + \log \frac{[\mathrm{HA}]}{[\mathrm{A^-}]} \qquad (6.40)$$

Thus, a general equation, Eqn (6.41), that is applicable to any acidic drug with one ionisable group may be written, where c_u and c_i represent the concentrations of the unionised and ionised species, respectively. This equation is known as the Henderson-Hasselbalch equation.

$$\mathrm{p}K_a = \mathrm{pH} + \log \frac{c_u}{c_i} \qquad (6.41)$$

The protonation of a basic drug (B) may be represented by the equilibrium shown in Eqn (6.42)

$$\mathrm{B} + \mathrm{H^+} \rightleftharpoons \mathrm{BH^+}$$

The acid dissociation constant, K$_a$, of the protonated base is therefore expressed by Eqn (6.43).

$$K_a = \frac{[\mathrm{H^+}][\mathrm{B}]}{[\mathrm{BH^+}]} \qquad (6.43)$$

Taking negative logarithms yields Eqn (6.44)

$$-\log K_a = -\log [\mathrm{H^+}] - \log [\mathrm{B}] + \log [\mathrm{BH^+}] \qquad (6.44)$$

or

$$\mathrm{p}K_a = \mathrm{pH} + \log \frac{[\mathrm{BH^+}]}{[\mathrm{B}]}$$

The Henderson-Hasselbalch equation for any weak base with one ionisable group may therefore be written as shown by Eqn (6.45)

$$\mathrm{p}K_a = \mathrm{pH} + \log \frac{c_i}{c_u} \qquad (6.45)$$

where c_i and c_u refer to the concentrations of the protonated and unionised species, respectively.

The Influence of pK$_a$ *Values on the Transport of Drugs Across Biological Membranes*. Because the ionised forms of acidic and basic drugs have low lipid:water partition coefficients compared to the coefficients for the corresponding unionised molecules, lipid membranes are preferentially permeable to the latter species. Thus, an increase in the fraction of a drug that is unionised will increase the rate of transport of the drug across a lipid membrane. However, the Henderson-Hasselbalch equation indicates that the ratio of unionised:ionised forms ($c_u:c_i$) of a given drug will depend on the pH of the medium and the pK_a value of the drug For example,

(i) the pK_a value of aspirin, which is a weak acid, is about 3·5, and if the pH of the gastric contents is 2·0 then from Eqn (6.41)

$$\log \frac{c_u}{c_i} = pK_a - pH = 3\cdot5 - 2\cdot0 = 1\cdot5$$

so that the ratio of the concentration of unionised acetylsalicyclic acid to acetylsalicylate anion is given by

$$c_u:c_i = \text{antilog } 1\cdot5 = 31\cdot62:1$$

(ii) the pH of plasma is 7·4 so that the ratio of unionised:ionised aspirin in this medium is given by

$$\log \frac{c_u}{c_i} = pK_a - pH = 3\cdot5 - 7\cdot4 = -3\cdot9$$

and

$$c_u:c_i = \text{antilog } -3\cdot9 = \text{antilog } \bar{4}\cdot1 = 1\cdot259 \times 10^{-4}:1$$

(iii) the pK_a of the weakly acidic drug sulphapyridine is about 8·0 and if the pH of the intestinal contents is 5·0 then the ratio of unionised:ionised drug is given by

$$\log \frac{c_u}{c_i} = pK_a - pH = 8\cdot0 - 5\cdot0 = 3\cdot0$$

and

$$c_u:c_i = \text{antilog } 3\cdot0 = 10^3:1$$

(iv) the pK_a of the basic drug amidopyrine is 5·0, and in the stomach the ratio of ionised:unionised drug is shown from Eqn (6.45) to be given by

$$\log \frac{c_i}{c_u} = pK_a - pH = 5\cdot0 - 2\cdot0 = 3\cdot0$$

and

$$c_i:c_u = \text{antilog } 3\cdot0 = 10^3:1$$

while in the intestine the ratio is given by

$$\log \frac{c_i}{c_u} = 5\cdot0 - 5\cdot0 = 0$$

and

$$c_i:c_u = \text{antilog } 0 = 1:1$$

Equilibrium Concentration Ratios. When an unionised drug attains an equilibrium distribution across a membrane its concentration is the same on both sides of the membrane. However, if the pHs of the aqueous phases on either side of the membrane are different (e.g. pH_1 and pH_2) then the total concentrations of a partly ionised drug in these aqueous phases will also differ. The ratio, R, of these concentrations (i.e. $R = c_1/c_2$, where c_1 and c_2 are the total concentrations of the drug in the two aqueous phases, respectively) may be calculated from Eqn (6.46), for an acidic drug, and from Eqn (6.47) for a basic drug.

$$\text{For an acidic drug } R = \frac{c_1}{c_2} = \frac{1 + 10^{(pH_1 - pK_a)}}{1 + 10^{(pH_2 - pK_a)}} \tag{6.46}$$

$$\text{For a basic drug } R = \frac{c_1}{c_2} = \frac{1 + 10^{(pK_a - pH_1)}}{1 + 10^{(pK_a - pH_2)}} \tag{6.47}$$

Equations (6.46) and (6.47) can also be used to calculate an unknown pH if all the other factors are known, or an unknown pK_a if all the other factors are known.

Table 6.2

Equilibrium Concentration Ratios of Various Drugs between Gastric or Intestinal Contents (pH 2 or 5, respectively) and Plasma (pH 7)

Type of drug	$c_{gastric}:c_{plasma}$	$c_{intestinal}:c_{plasma}$
Strongly acidic ($pK_a = 3$)	$1 + 10^{-1}:1 + 10^4$	$1 + 10^2:1 + 10^4$
Weakly acidic ($pK_a = 8$)	$1 + 10^{-6}:1 + 10^{-1}$	$1 + 10^{-3}:1 + 10^{-1}$
Strongly basic ($pK_a = 8$)	$1 + 10^6:1 + 10$	$1 + 10^3:1 + 10$
Weakly basic ($pK_a = 5$)	$1 + 10^3:1 + 10^{-2}$	$1 + 1:1 + 10^{-2}$

Table 6.2 shows the equilibrium concentration ratios of acidic and basic compounds with a range of pK_a values between gastric or intestinal contents and plasma. (The pHs of these fluids have been taken as a whole number for the convenience of calculation.) From this table it can be seen that for an acidic drug the $c_{gastric}:c_{plasma}$ ratio is lower than the $c_{intestinal}:c_{plasma}$ ratio; i.e. a greater proportion of the drug is absorbed into the plasma

from the stomach than from the intestine. The reverse is true for basic drugs. It should be pointed out that the figures given in Table 6.2 are theoretical ones based on the assumption that an equilibrium is established. However, such an equilibrium is rarely obtained *in vivo* because the concentrations of drug on either side of the membrane will be affected by other factors; e.g. flow of blood, dilution of gastric contents.

2. DIFFUSION THROUGH PORES IN LIPID BARRIER

The pores in cell membranes are filled with water, and substances may therefore be transported through the pores by diffusion in aqueous solution. The rate of such transport may be affected by the movement of water under the influence of osmosis. In addition, the movement of ions through these pores will be affected by the presence of static charges on the walls of the pores; e.g. small anions may pass through positively charged pores whereas cations may be hindered by electrostatic repulsion (p. 73). The size of pores in membranes is the main factor that affects transport of substances by this mechanism; i.e. a filtration process is involved. The pore size in cell membranes is usually about 4×10^{-7} mm, and since the radius of most drug molecules is greater than this the entry of drugs into individual cells by this mechanism is of little importance. However, certain body membranes consist of sheets of epithelial cells, and the pores between these cells may be as large as 4×10^{-6} mm; e.g. in the glomeruli of the kidney and in the epithelium of capillaries. The importance of the filtration mechanism of drug transport is therefore greater in such membranes.

3. SPECIAL TRANSPORT SYSTEMS

Some naturally occurring substances are too large to pass through the pores in cell membranes and too lipid insoluble to be transported by diffusion through the lipid barrier; (e.g. large ions and large monosaccharides). It is therefore supposed that the ready movement of such substances across cell membranes is brought about by special mechanisms which involve compounds that act as carriers for the transported substances. It is suggested that an ion or molecule forms a complex with the carrier at one surface of the membrane, and the complex moves across the membrane. After liberation of the substrate (i.e. the transported substance) at the other side, the cycle is completed by the return of the uncomplexed carrier to its initial position.

If the substrate is moved across a membrane in the opposite direction to a concentration gradient, then energy must be expended in the transference, which is then referred to as an active transport process. This type of transport is inhibited by compounds that interfere with the cellular reactions that provide the required energy for the process.

The rates of transference in all special systems that involve carrier molecules will depend on the available number of these molecules, and a maximum rate will be obtained when all the carrier molecules are being used. It is therefore possible to saturate this type of transport mechanism with an excess of substrate molecules. Furthermore, although carrier molecules usually show some degree of specificity for a particular substrate, other compounds that possess a similar structure to the normal substrate may compete for the available carrier molecules and, therefore, prevent the transport of the normal substrate.

4. PINOCYTOSIS

It is considered that small particles such as microorganisms may be engulfed by macrophages that have migrated through the epithelial lining of the gastro-intestinal tract. The engulfed particles are transported across this membrane, when the macrophages migrate back again into the lymphoidal tissue. This process appears to be more significant at sites in the body where lymphoidal tissue and mucosal epithelium are close together; e.g. the tonsil and pharynx.

The first three transport mechanisms that have been discussed above are involved to some extent in the passage of drugs across all biological membranes. Their relative importance depends on the structure of the particular membrane. In certain cases it is usual to refer to specialised membranes or barriers in the body (e.g. the blood-brain barrier). Such barriers consist of membranes with structures that differ from that of the simple cell membrane or contain specific carriers for the transport of particular compounds.

The Binding of Drugs *in vivo*

In Fig. 6.5 it is shown that a drug is bound in a reversible manner in various parts of the body. The nature of the binding material may vary but, in general, the bound drug is unable to pass through cell membranes because of the size and nature of the complex in which it is involved.

PROTEIN BINDING

The majority of drugs show some form of reversible binding with proteins. Several sites on the protein molecules may be involved and the stability of the drug-protein complex will determine its influence on the activity of the drug.

The formation of a complex, DP, by interaction of a drug, D, with a protein, P, may be represented by Eqn (6.48) where k_1 and k_2 are the rate constants for the forward and reverse reactions.

$$D + P \underset{k_2}{\overset{k_1}{\rightleftharpoons}} DP \tag{6.48}$$

The amount of drug held in the complex form DP will depend on the concentrations of the drug and protein ([D] and [P], respectively), on the number of binding sites per protein molecule, *n*, and the affinity of the drug for the binding sites. This latter affinity may be represented by the association constant K for the reaction expressed in Eqn (6.48); (i.e. $K = k_1/k_2$). Goldstein (1949) has shown that if equilibrium is established then the fraction of total drug that is bound to a protein is given by Eqn (6.49)

$$\frac{[\text{Bound drug}]}{[\text{Total drug}]} = \frac{1}{1 + \frac{1}{Kn[P]} + \frac{[D]}{n[P]}} \tag{6.49}$$

It can be seen from this equation that the fraction of bound drug increases as K, *n* and P increase and decreases as D increases.

Albumin is the commonest protein involved in the binding of drugs in plasma, although the α-globulins sometimes show a specificity for particular drugs. However, the binding capacity of the latter is relatively low compared with that of albumin.

The binding of drugs to proteins gives rise to various effects, some examples of which are given below.

1. Protein bound drugs are unable to penetrate membranes and therefore cannot reach the sites where excretion or biotransformation to other compounds occurs. Since the binding to proteins is a reversible interaction, the bound drug acts as a reservoir which releases unbound drug to replace that lost from the system by excretion or biotransformation. This effect is important in maintaining the concentration of unbound drug at a given level and, therefore, in prolonging the duration of action of the drug. For example, the different degrees of protein binding that occur with sulphonamides enable these drugs to be divided into long and short acting types; e.g. strongly bound drugs, such as sulphamethoxydiazine, exhibit a prolonged activity compared with those, such as sulphathiazole, which are only weakly bound to proteins. This effect is also responsible for the more prolonged action of hydroxocobalamin compared with that of cyanocobalamin.

2. Protein binding may influence the solubility of drugs in the plasma; e.g. the solubility of dicoumarol in plasma is much greater than it is in physiological saline solutions (Gourley, 1967).

3. The response to a drug is caused by the build-up of a particular concentration of the drug at its site of action. Its total concentration in the plasma (i.e. bound + unbound drug) is, however, often used as an estimate of the amount of drug required to produce a particular response because it is much easier to determine this concentration. This procedure is often inaccurate since the different degrees of binding to proteins in the plasma and tissue proteins at the site of action may lead to an unequal distribution of the drug between the plasma and the particular tissue; for example, the strong binding of chlorpromazine to the lipoproteins in brain tissue produces a total concentration of this drug in the brain that is approximately fifty times greater than its total concentration in the plasma (Gourley, 1967).

4. Different substances may compete for the same binding sites on proteins. Such competition may lead to a variety of effects. For example, (*a*) the displacement of antibiotics from protein binding may be advantageous since a higher concentration of free drug is available to exert the required activity; (*b*) the displacement of protein bound anticoagulants may produce a high concentration of free drug and may therefore lead to excessive bleeding; (*c*) the acidic binding sites on albumin in premature babies are usually saturated with bilirubin, because the concentration of albumin is relatively low. The increase in free bilirubin that results when the bound form is displaced by acidic drugs such as salicylates and sulphonamides may cause brain damage (Gourley, 1967).

5. The response produced by a drug is caused by interaction between the drug and a receptor site. This interaction is similar to protein binding except that the latter type of binding involves relatively non-specific interactions while the drug-receptor interaction is very specific. However, substances with a similar structure to naturally occurring compounds may compete for a particular receptor site and so prevent a normal response; e.g. atropine inhibits the effect of acetylcholine by competing for the receptor sites on the cholinergic nerve endings.

BINDING TO OTHER TISSUE CONSTITUENTS

Interaction between the constituents of other tissues and drugs may lead to effects similar to those caused by binding to proteins. For example, very lipid soluble drugs (e.g. thiopentone) may accumulate in the body fat and tetracyclines may accumulate in hard tissues such as teeth, although the nature of the latter interaction is unknown.

Blood Concentrations

The concentration of a drug in the plasma is often related to the level of activity of the drug. It can be seen from Fig. 6.5 that the blood concentration will be affected mainly by the rate at which drug is absorbed into the plasma and by the rate at which the drug is eliminated by excretion or biotransformation. In addition, the distribution of drug between the plasma and other tissues will affect the blood concentration. Measurement of the concentration of a drug in the blood at various times after administration of a single dose produces results that can usually be represented by a curve similar to that shown in Fig. 6.8.

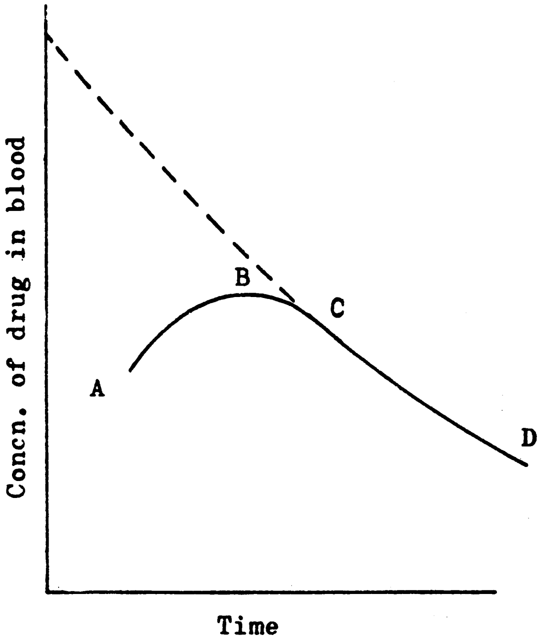


Fig. 6.8 The change in blood concentration of a systemically-active drug after administration of a single dose

The individual points on such a curve represent the difference between the amount of drug that has entered the plasma and the amount that has been eliminated in a given time. The shape of the curve indicates the rate of change in blood concentration (dc/dt), which can be expressed in terms of Eqn (6.50), where dA/dt and dE/dt are the rates of entry into and elimination from the plasma respectively.

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \frac{\mathrm{d}A}{\mathrm{d}t} - \frac{\mathrm{d}E}{\mathrm{d}t} \qquad (6.50)$$

The initial part, AB, of the curve in Fig. 6.8 indicates that the rate of absorption exceeds the rate of elimination, and this part of the curve is dependent on the nature of the dosage form and the route of administration. Obviously, in intravenous injections this part is absent and the maximum blood concentration is obtained immediately. The effects of other routes of administration and dosage forms that were included in Fig. 6.5 will be discussed later.

The latter part, BD, of the curve indicates that absorption is ceasing and the rate of elimination is now greater, and the smooth section of this part, i.e. CD, may be extrapolated backwards to indicate the apparent concentrations at shorter times, as shown by the dotted line. The elimination curve DC and its extrapolation can often be expressed by the first order kinetics, the linear form of which is given by Eqn (6.51)

$$\log C = \log C_0 - \frac{kt}{2{\cdot}303} \qquad (6.51)$$

where C is the concentration of drug in the blood, at time t, C_0 is the hypothetical concentration of drug when $t = 0$, and k is the rate constant for elimination of drug from the blood. Thus, a plot of log C versus t yields a straight line with a slope of $-k/2{\cdot}303$, and an intercept equal to $\log C_0$ if the disappearance of drug follows first order kinetics.

BIOLOGICAL HALF-LIFE

The rate of disappearance of a drug from body fluids such as blood is often expressed conveniently in terms of its biological half-life, $t_{\frac{1}{2}}$, which is the time taken for its concentration in the fluid to fall to half its initial value, assuming that the disappearance follows first order kinetics. This value is related to the rate constant k in Eqn (6.51), as shown by Eqn. (6.52)

$$t_{\frac{1}{2}} = \frac{0{\cdot}693}{k} \qquad (6.52)$$

The half-life can also be determined from inspection of graphs (*see* above).

EFFECT OF DOSAGE FORM AND ROUTE OF ADMINISTRATION ON THE BLOOD CONCENTRATIONS OF DRUGS

The rapidity with which maximum blood concentration of a drug is attained is affected markedly by the route of administration. In addition, the proportion of the actual dose administered that is absorbed into the plasma is also affected. In general, in drugs administered as solutions, these differences are dependent on the number and the nature of the membranes across which the drug must be transported before it reaches the plasma. Thus, a drug that is administered intravenously is not subject to any initial transport processes so that the maximum blood concentration is obtained immediately and the complete dose enters the plasma. This route of administration therefore offers the greatest degree of control of blood concentrations.

Absorption of drugs into the plasma after their administration as intramuscular or subcutaneous injections involves transport across membrane barriers. Such absorption is usually more rapid in intramuscular injections than with subcutaneous ones because the skeletal muscles have a better blood supply than subcutaneous tissue. In general both routes provide rapid absorption and involve little loss of drug, so that doses required to achieve a certain blood concentration are less than those in orally administered solutions. It should be borne in mind that the nature of specific drugs may alter these general observations. For example, the binding of phenylbutazone to tissue proteins after intramuscular injection reduces its absorption into the plasma to a considerably slower rate than that obtained after oral administration (Gourley, 1967).

The absorption of dissolved drugs from the stomach and intestine is dependent on the rates at which they are transported across the membrane lining the gastro-intestinal tract. The particular influence of pH on this transport process has already been discussed. The rate of absorption is also affected by the contents of the stomach and intestine and by the rate at which the stomach empties and fills. Rapid emptying will decrease absorption of a drug that is absorbed mainly from the stomach and enhance absorption of a drug that is absorbed readily from the intestine. Since the above effects vary tremendously between individuals, and even in the same individual under different conditions, the control of blood concentrations of orally administered drugs is difficult. Similar effects also apply to the absorption of drugs administered rectally as suppositories or retention enemas, poor control of blood concentrations again resulting.

The above discussion indicates that parenteral administration, particularly intravenous injection, provides the most rapid attainment of maximum blood concentration and allows the most satisfactory control over this concentration. However, the advantages and disadvantages of particular routes of administration must also be taken into account, as indicated by the examples in the following list.

(*a*) Injections are much less acceptable than oral preparations for self-administration by the patient because of: (i) the inconvenience in administration, (ii) the possibility of sensitisation, (iii) the possibility of discomfort, and (iv) the need for sterility of syringes etc.

(*b*) Inactivation of drugs by digestive enzymes or by the acid pH in the stomach may preclude oral administration. Parenteral or rectal administration may be used to overcome these effects.

(*c*) Difficulties in swallowing may prevent the use of oral administration.

When drugs are administered in dosage forms other than solutions the rate of release of drug from a dosage form as a solute in aqueous solution may be more important than its rate of transport across membranes in determining the rate of appearance in the plasma. In oral dosage forms in which the drug is present as a solid (e.g. powders, tablets, capsules, suspensions), the rate of dissolution in the gastro-intestinal fluids is often the rate determining process for absorption. Pharmaceutical manipulations associated with the control of absorption rates are therefore usually concerned with methods of increasing or decreasing the dissolution rates of drugs. Some examples of these methods that are relevant to the dosage forms given in Fig. 6.5 are indicated in the following list.

1. *Particle Size*

A decrease in particle size leads to an increase in the surface area of a powder and therefore to an increase in rate of dissolution. More rapid absorption from the stomach of a slowly soluble or relatively insoluble drug can therefore be achieved by the use of dosage forms that produce suspensions of smaller particles of the drug in the stomach. This effect is of particular importance in the formulation of tablets of slowly soluble materials since the surface area of a tablet is small. Such tablets therefore usually incorporate disintegrating agents that cause disruption of tablets into small particles when in contact with water. The drug is released by dissolution more rapidly from such particles than from the entire tablet. The rate of disintegration may also affect the rate of absorption of drugs administered in tablet form, and the BP includes a test that is designed to limit the time taken for disintegration of official tablets.

The absence of a disintegrating agent in a tablet that consists mainly of a slowly soluble substance produces the converse effect; i.e. a reduced rate of dissolution of the tablet. This effect is made use of in the formulation of tablets that provide a slow release of drug. Such slowly released doses are usually administered together with an initial dose of drug contained in part of the tablet that disintegrates and therefore is made available for rapid absorption. The slow release from the non-disintegrating part of the tablet maintains the concentration of drug in the plasma by allowing sufficient absorption to replace drug lost by excretion or biotransformation.

The increased absorption rate that results from the use of a smaller particle size may increase the efficiency of absorption; i.e. increase the proportion of drug that is absorbed from the administered dose. This effect is demonstrated by the observations of Atkinson *et al.* (1962) which showed that a reduction in the particle size of griseofulvin from 10 to 2·7 μm enabled the required effect of this antibiotic to be obtained with half the original dose.

The increased rate of solubility of chloramphenicol from solid solutions of this drug in urea is attributed to the fact that rapid dissolution of urea from these solid solutions leaves the chloramphenicol in a very finely dispersed form that is more rapidly soluble than normal particles of this drug (p. 30).

The more efficient absorption of salts of poorly soluble, weakly acidic drugs compared with the absorption of the corresponding free acids is attributed to the fact that although the free acids are liberated from their more soluble salts by the action of hydrochloric acid in the stomach, precipitation of the free acids, if it occurs, gives rise to very fine particles only. The latter redissolve rapidly when the concentration of the free acid in solution is decreased by absorption (Gorringe and Sproston, 1964).

2. *Changes in Chemical Structure*

Desirable changes in dissolution rates can often be achieved by the use of simple derivatives, different polymorphic forms or different hydrated forms of a drug. The effect of using the more soluble salts of sparingly soluble weak acids has already been discussed above. In addition, the use of chloramphenicol palmitate and other fatty acid esters of chloramphenicol instead of the free drug provides an example of the effect of change in solubility obtained from the use of a simple derivate of a drug. Chloramphenicol is extremely bitter and is therefore unsuitable for paediatric suspensions. The fatty acid esters of this drug are so insoluble that they do not stimulate the taste receptors on the tongue and are therefore satisfactory for paediatric use. The free drug is absorbed after liberation from the ester by hydrolysis in the stomach.

The effect of polymorphism is illustrated by the solubilities of the three polymorphic forms of riboflavine, which are 60 mg, 80 mg, and 1200 mg/litre, respectively. The effect of hydration of compounds (i.e. presence of water of crystallisation) usually causes a decrease in the rate of dissolution compared with that of the anhydrous forms; e.g. the anhydrous forms of caffeine and theophylline are more rapidly soluble than their hydrated forms.

3. *Properties of Dosage Form*

The importance of control of tablet disintegration rates in relation to methods of accelerating or retarding the rates of dissolution of ingredients has already been discussed. Retardation of dissolution rates is also achieved in certain dosage forms intended to provide a prolonged release of drug by the use of physical barriers that usually impede the diffusion of solvent into the dosage form and the outward diffusion of drug in solution. These barriers often consist of water insoluble fats, waxes or plastics and they may be used as a coating around the drugs or as matrices in which the drug is embedded. The rate of penetration of water and aqueous solutions through such barriers depends on the thickness of the diffusion pathway, the nature of the barrier materials (i.e. the more hydrophobic the material the slower the penetration of water), and on the porosity of the barrier. A similar effect is also achieved by the use of ion exchange resins from which bound drug is released when ion exchange occurs in the gastro-intestinal contents. The rate of exchange and the outward diffusion of drug is again dependent on the porosity of the ion exchange beads.

In certain cases the barrier may be insoluble in the gastric fluid but soluble in the intestinal juice. Such barriers, which may consist of shellac or cellulose acetate phthalate, are referred to as enteric coatings and may be used for orally administered drugs that are irritant to the gastric mucosa or that are destroyed by acid pH or gastric constituents.

Drugs administered in the form of suppositories are released into the rectum when the base melts at body temperature (e.g. theobroma oil, Witepsol bases) or when the base dissolves in the aqueous rectal liquid (e.g. polyethylene glycols). The small amount of water usually available in the rectum tends to retard the release of medicaments contained in the latter type of base. The release rates of the former type may be affected by the inclusion of surface active agents in the base. It is suggested that such agents alter the distribution of drugs between the molten fatty bases and aqueous body fluids and may cause emulsification at the interface between these two phases. The latter effect leads to an increase in the interfacial area across which partition can occur and, therefore, increases the rate of release of drug.

BIBLIOGRAPHY

BINNS, T. B. (Ed.) (1964) *Absorption and Distribution of Drugs.* Livingstone, London.

GARRETT, E. R. (1967) *Advances in Pharmaceutical Sciences. Vol. 2.* (H. S. Bean, A. H. Beckett, and J. E. Carless, Eds). Academic Press, London.

HOOVER, J. E. (Ed.) *Remington's Pharmaceutical Sciences* (1970) 14th ed. Mack Publishing Co., Pennsylvania.

KUNIN, R. (1958) *Ion Exchange Resins.* 2nd ed. Wiley and Sons, London.

LUNDBERG, W. O. (Ed.) (1961) *Autoxidation and Antioxidants. Vol. 1.* Interscience, New York.

MARTIN, A. N., SWARBRICK, J. and CAMMARATA, A. (1969) *Physical Pharmacy.* 2nd ed. Henry Kimpton, London.

SAUNDERS, L. (1966) *Principles of Physical Chemistry for Biology and Pharmacy.* University Press, Oxford.

SCOTT, G. (1965) *Atmospheric Oxidation and Antioxidants.* Elsevier, Amsterdam.

TROTMAN-DICKENSON, A. F. and PARFITT, G. D. (1966) *Chemical Kinetics and Surface and Colloid Chemistry.* Pergamon, London.

WILLIAMS, V. R. and WILLIAMS, H. B. (1967) *Basic Physical Chemistry for the Life Sciences.* Freeman and Co., London.

REFERENCES

ATKINSON, R. M., BEDFORD, C., CHILD, K. J., and TOMICH, E. G. (1962) *Antibiotics Chemother.*, **12**, 232–238.

BRODIE, B. B. (1964) *Absorption and Distribution of Drugs.* (T. B. Binns, Ed.). Livingstone, London, 16–48.

GARRETT, E. R. (1967) *Advances in Pharmaceutical Sciences. Vol. 2.* (H. S. Bean, A. H. Beckett, and J. E. Carless, Eds). Academic Press, London, 2–94.

GOLDSTEIN, A. (1949) The interactions of drugs and plasma proteins. *Pharmac. Rev.*, **1**, 102–165.

GORRINGE, J. A. L. and SPROSTON, E. M. (1964) *Absorption and Distribution of Drugs.* (T. B. Binns, Ed.). Livingstone, London. 128–139.

GOURLEY, D. R. H. (1967) *Modern Trends in Pharmacology and Therapeutics. Vol. 1.* (W. F. M. Fulton, Ed.). Butterworths, London, Chapter 1.

GUNN, C. and CARTER, S. J. (1965) *Dispensing for Pharmaceutical Students.* Pitman Medical, London.

HIGUCHI, T. and LACHMAN, L. (1955) Inhibition of hydrolysis of esters in solution by formation of complexes. I. Stabilization of benzocaine with caffeine. *J. Amer. pharm. Ass.*, (*Sci. Ed.*), **44**, 521–526.

JOHNSON, C. A. (1967) *Advances in Pharmaceutical Sciences. Vol. 2.* (H. S. Bean, A. H. Beckett, and J. E. Carless, Eds). Academic Press, London, 224–310.

JONES, W. and GRIMSHAW, J. J. (1963) Accelerated stability testing of pharmaceutical products. Papers 1 and 2. *Pharm. J.*, **191**, 459–465.

MACEK, T. J. (1970) *Remington's Pharmaceutical Sciences.* 14th ed., Mack Publishing Co., Pennsylvania. Chapter 79.

MARTIN, A. N., SWARBRICK, J. and CAMMARATA, A. (1969) *Physical Pharmacy.* 2nd ed. Henry Kimpton, London. Chapter 14.

NACHOD, F. C. and SCHUBERT, J. (1956) *Ion Exchange Technology.* Academic Press, New York.

SAUNDERS, L. (1966) *Principles of Physical Chemistry for Biology and Pharmacy.* University Press, Oxford, Chapter 8.

TOOTILL, J. P. R. (1961) A slope-ratio design for accelerated storage tests. *J. Pharm. Pharmac.*, **13**, 75T–86T.

7

Rheology

RHEOLOGY is the science concerned with the deformation of matter under the influence of stresses, which may be applied perpendicularly to the surface of a body (a tensile stress), tangentially to the surface (a shearing stress), or at any other angle to the surface. The deformations that result from the application of a stress may be divided into two types:

1. spontaneously reversible deformations or elastic deformations, and
2. permanent or irreversible deformations that are referred to as flow and are exhibited by viscous bodies.

The work used in producing an elastic deformation is recoverable when the body returns to its original shape after removal of the applied stress. However, in irreversible deformations the work used in maintaining deformation is dissipated as heat and is not recoverable mechanically when the stress is removed.

The majority of pharmaceutical systems where rheological properties are important range from simple liquids to semi-solids such as ointments, gels, creams, and pastes. Most attention has been paid to the flow properties of these systems, i.e. to their viscous behaviour. The importance of contributions from elastic effects to the rheological properties of semi-solid systems has gained more attention in recent years. However, this chapter is restricted to a consideration of the flow properties of pure liquids and disperse systems containing a liquid continuous phase since the complex nature of the rheological properties of semi-solids is outside the scope of the present discussion; information on these systems may be obtained from the booklist at the end of this chapter and from the review given by Barry and Warburton (1968).

Viscosity

Application of a shearing force to a fluid (i.e. a liquid or a gas) usually causes it to flow because fluid cannot support a strain for very long periods of time. When the force is removed, the fluid does not return to its original state; i.e. irreversible deformation has occurred. Such a deformation may be considered to involve a shearing action between infinitely thin layers (laminae) within the fluid. The deformation that results is expressed in terms of the rate of shear, which is the change in velocity of flow with a distance measured at right angles to the direction of flow. The shear stress that causes a particular rate of shear is obtained by dividing the shearing force by the area of the surface to which the shearing force is tangentially applied. The ratio of the applied shear stress to the rate of shear is known as the coefficient of viscosity. The effect of rate of shear on this ratio varies for different systems which has led to these systems being classified into the following types.

1. Newtonian Fluids

Flow of this type of fluid can be illustrated by Fig. 7.1, which represents two parallel planes each of area A separated by a fluid of depth x. The upper plane is caused to move with a velocity u relative to the lower plane by the application of a shearing force F. This movement causes the fluid in the laminae separating the two planes to be displaced relative to adjacent layers as shown in Fig. 7.1. The rate of movement of the laminae varies from a maximum value in the layer adjacent to the upper plane to a value that is close to zero in the layer adjacent to the lower plane. The mean rate of shear is given by the overall difference in fluid velocity, u, divided by the distance between the two planes, x; (i.e. the mean shear rate $= u/x$). In more general terms, for any point in a layer that is separated by a distance $\mathrm{d}x$ from the upper plane the rate of shear is given by $\mathrm{d}u/\mathrm{d}x$. The shear stress, S, is given by F/A (i.e. the applied force per unit area) and the flow of Newtonian fluids may be expressed quantitatively by the Newton equation, Eqn (7.1)

$$S = \eta \frac{\mathrm{d}u}{\mathrm{d}x} \tag{7.1}$$

where η is a constant that is known as the coefficient of viscosity or simply the viscosity of the fluid.

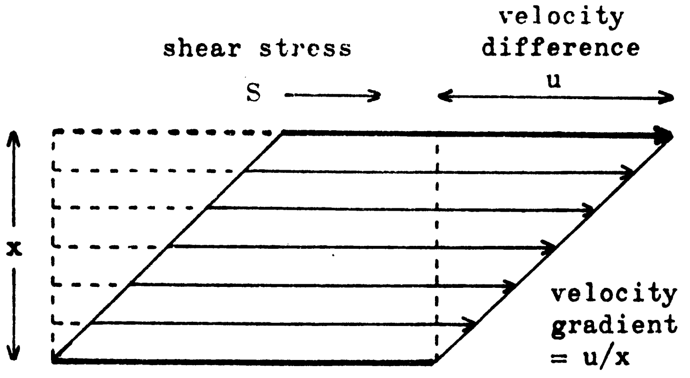


Fig. 7.1 Movement of planar laminae in simple shear

The dimensions of shear stress are force per unit area, and shear rate is measured in terms of reciprocal time; (shear rate = velocity/distance = distance × time^{-1}/distance = time^{-1}). Thus, the dimensions of η are given by Eqn (7.2)

$$\eta = \frac{S}{du/dx} = \frac{F/A}{du/dx} = \frac{\text{force}}{\text{area} \times \text{time}^{-1}} \qquad (7.2)$$

However, since

$$\text{force} = \frac{\text{mass} \times \text{length}}{\text{time}^2}$$

and

$$\text{area} = \text{length}^2$$

the dimensions of η that are indicated by Eqn (7.2) reduce to mass/length × time (i.e. $ML^{-1}T^{-1}$).

The basic SI unit is therefore the $\text{kg m}^{-1}\text{ s}^{-1}$. However, in the cgs system it was usual to express viscosities in units of dyne sec cm^{-2} in accordance with Eqn (7.2) and such units were termed poises in honour of Poiseuille, who was an early worker in this field. A similar derived SI unit is therefore used, the N s m^{-2}. It can be shown that 1 P (poise) = $0{\cdot}1\text{ N s m}^{-2}$, and 1 centipoise, cP, which is a convenient unit for use with fluids of low viscosity, is equal to $1 \times 10^{-3}\text{ N s m}^{-2}$.

The rheological properties of liquids are usually expressed in the form of flow diagrams, which consist of graphs showing the variation of shear rate with shear stress. Equation (7.1) indicates that for a Newtonian fluid such a graph should be linear, as shown by line (*a*) in Fig. 7.2, the slope of which is equal to the reciprocal of viscosity, η, of the fluid, a value referred to as the fluidity, ϕ, i.e.

$$\phi = \frac{1}{\eta} \qquad (7.3)$$

The linear nature of the flow diagram for a Newtonian fluid shows that η is, in fact, a true constant unaffected by the value of the rate of shear. Thus, a single determination of η from the shear stress at any given shear rate is sufficient to characterise the flow properties of a Newtonian fluid. Liquids that show this type of behaviour include water, simple organic liquids, true solutions, and dilute suspensions and emulsions.

2. Non-Newtonian Fluids

Many liquids encountered in pharmaceutical practice (e.g. concentrated suspensions and emulsions) do not follow Eqn (7.1) because the value of η varies with the rate of shear. It is therefore usual to consider the apparent viscosities of these systems at particular rates of shear, where the apparent viscosity, η_{app}, is the ratio of shear stress to shear rate

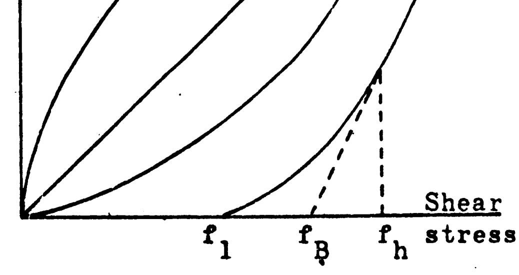


Fig. 7.2 Flow curves for Newtonian and Non-Newtonian systems

(*a*) Newtonian (*b*) Plastic
(*c*) Pseudoplastic (*d*) Dilatant

at a given point on the flow diagram. In addition, the influence of shear on the apparent viscosity may be time-dependent and this effect provides a further variation in the flow properties of some non-Newtonian fluids. The results of such time-dependent effects will be discussed later.

The lines (*b*), (*c*), and (*d*) in Fig. 7.2 represent the flow behaviours of the three common types of non-Newtonian fluid, in which time-dependent effects are absent. These types are referred to as plastic, pseudoplastic, and dilatant flow, respectively. The changing slopes of these three lines indicate that the apparent viscosity varies with rate of shear. A single determination of η_{app} from the shear stress at any given shear rate would therefore be of little use, and the whole flow diagram must be determined in order to characterise these non-Newtonian systems.

(a) PLASTIC FLOW

This is represented by curve (*b*) in Fig. 7.2 in which it can be seen that the line does not pass through the origin of the graph but arises at some point on the shear stress axis. This indicates that a certain shearing stress must be exerted before flow begins. This stress is termed the yield value, the system behaving like a solid when small stresses that are lower than this value are applied to it; i.e. the system exhibits elastic deformations that are reversible when these small stresses are removed.

Materials that show plastic behaviour are often termed Bingham bodies in honour of Bingham, who carried out many of the early studies on these materials. The quantitative behaviour of these systems is usually expressed in terms of the Bingham equation, Eqn (7.4), where f_B is the Bingham yield value, η_{pl} or U is the plastic viscosity, the other terms being defined as before.

$$U \quad \text{or} \quad \eta_{\text{pl}} = \frac{S - f_B}{\mathrm{d}u/\mathrm{d}x} \qquad (7.4)$$

This equation implies that the flow diagram is a straight line that arises on the shear stress axis at the yield value. In practice, however, flow usually occurs at a lower shear stress, the flow curve gradually approaching the theoretical line shown in Fig. 7.2. It is often more satisfactory to use the lower yield value, f_1, in practical applications of plastic materials since this indicates when actual flow begins. In addition, a higher yield value, f_h, is sometimes used. This corresponds to the shear stress beyond which the flow curve becomes linear. The three yield values are indicated in Fig. 7.2.

(b) PSEUDOPLASTIC FLOW

This type of behaviour is represented by line (*c*) in Fig. 7.2. It can be seen that the curve arises at the origin of the graph, i.e. no yield value exists, and flow begins immediately on application of a shearing stress. The slope of the curve gradually increases until it reaches a maximum value. Since the apparent viscosity at any shear rate is given by the reciprocal of the slope it can be seen that the apparent viscosity decreases as the shear rate increases, until a constant value is reached.

Pseudoplastic flow cannot as yet be satisfactorily expressed quantitatively by equations based on fundamental principles. An empirical equation, Eqn (7.5), is often used and this shows reasonable agreement with experimental results except for those obtained over extended ranges of shearing stresses.

$$S^n = k\frac{\mathrm{d}u}{\mathrm{d}x} \qquad (7.5)$$

In this equation k and n are constants for a particular system. If $n = 1$ the equation is similar to the Newtonian equation, Eqn (7.1), but the greater the value of n the more pseudoplastic the behaviour. The values of these constants can be obtained by plotting log $\mathrm{d}u/\mathrm{d}x$ versus log S. If Eqn (7.5) is applicable, then a straight line is obtained with a slope of n and an intercept on the log $\mathrm{d}u/\mathrm{d}x$ axis that is equal to log k.

The decrease in apparent viscosity with increasing rates of shear in plastic and pseudoplastic systems results from the breakdown, under the influence of shear, of structures (e.g. aggregates of dispersed particles) in the system. Greater breakdown occurs at higher shear rates, although beyond certain high shear rates no further structural breakdown can occur and the apparent viscosity then becomes constant. When the shear stress is reduced or removed, reformation of the structures in these systems occurs immediately and the flow curve obtained from measurements at decreasing shear rates is superimposable on that obtained from measurements taken at increasing shear rates. The occurrence of a yield value in plastic systems indicates that stronger forces than those in pseudoplastic systems must first be overcome before flow can occur.

To distinguish between plastic and pseudoplastic behaviour it is necessary to obtain measurements at low shear rates. Extrapolation of linear portions of flow curves obtained at high rates of shear should not be carried out to provide possible yield values since the system may, in fact, be a pseudoplastic one. Thus, differentiation between these two types of behaviour often depends upon the sensitivity of the method of measurement at low shear rates.

(c) DILATANT FLOW

This type of behaviour is represented by curve (*d*) in Fig. 7.2. It can be seen that the slope of this curve gradually decreases to a constant value, which indicates that the apparent viscosity must increase with increase in shear rate up to a maximum value.

A power equation similar to that used for pseudoplastic flow (Eqn 7.5) may be used to describe the behaviour of dilatant materials but in this case the constant n is less than one.

Dilatancy is usually exhibited by concentrated dispersions of deflocculated particles (*see* Chapter 5). It is suggested that in these systems the particles are arranged in a state of close packing and the small amount of liquid present is sufficient to fill the narrow spaces between adjacent particles. These thin liquid films allow the system to flow like a liquid when the rate of shear is low. However, at higher shear rates the particles will become displaced from their close-packed arrangement, which results in the formation of larger void spaces in the system. The liquid continuous medium is now insufficient to fill all the spaces between particles, hence the movement of the latter relative to each other involves a greater amount of friction and the apparent viscosity therefore increases. This effect may be troublesome in high-speed milling processes since the viscosity of dilatant suspensions may increase so much at the high rates of shear involved in the operation of these mills that overloading of the motors may occur.

Time-dependent Effects

In certain cases the breakdown by shearing forces of structures in disperse systems is markedly time-dependent; i.e. the amount of breakdown increases with time so that the apparent viscosity gradually decreases even when the shear rate is maintained constant. The effect may be reversible or irreversible after removal of the shearing stress. If it is reversible, i.e. if the structure reforms on removal of the stress, the reformation process is usually time-dependent also, so that a time lag occurs before the original viscosity is achieved.

The redetermination of a flow curve after an initial determination may be used to indicate the difference in flow properties of a system before and after structural breakdown by shearing forces has occurred; i.e. the two curves are not superimposable and a hysteresis loop is obtained in the flow diagram (*see* Fig. 7.3). The area of such a loop may be used as an indication of the relative amount of structural breakdown that has occurred in the system.

The previous paragraphs have considered only time-dependent effects that result in a decrease in apparent viscosity with increase in shear rate. The reverse process may also occur. The various types of time-dependent effect are as follows.

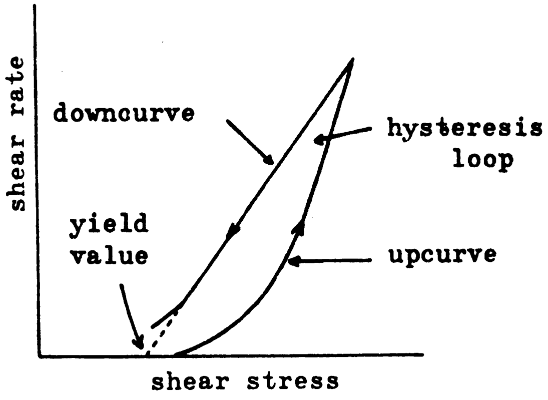


Fig. 7.3 Flow curve for a system that exhibits thixotropy superimposed upon plastic behaviour

(a) THIXOTROPY

This term means 'to change by touch'. It is usually defined as a reversible isothermal transition from a gel to a sol (i.e. a colloidal dispersion). However, it is applied to any reversible time-dependent decrease in apparent viscosity that results from the application of shearing forces. The decrease in viscosity arises from a breakdown of structures (e.g. a gel network) within a system when it is sheared. After the shearing forces are removed a time lag occurs before structural re-formation is complete. Bentonite gel is a good example of a system that exhibits thixotropy, and Fig. 7.3 illustrates the flow diagram of the system. It is believed that the breakdown of the random network (*see* Chapter 5) formed by the hydrated bentonite particles on application of a shearing force is caused by the elongated particles aligning with their long axes parallel to the direction of flow of the continuous aqueous phase. This orderly arrangement causes the interparticle links to be broken; the network therefore disintegrates and the apparent viscosity decreases. On removal of the shearing forces the arrangement of dispersed particles gradually becomes less orderly under the influence of Brownian motion, and the gel network re-forms after a time lag.

A similar effect, referred to as irreversible thixotropy, is sometimes encountered. Application of shearing forces causes breakdown of structures within the system but these structures do not re-form on removal of the shear stress, or the time lag is so long that from a practical point of view the effect is irreversible.

(b) RHEOPEXY

The time lag that occurs before a reversible thixotropic system returns to its original state after removal of the shearing stress may often be reduced by applying a gentle rolling or rocking motion. It is suggested this motion provides a mild turbulence that aids the return of dispersed particles to a random orientation when network re-formation can occur.

(c) NEGATIVE THIXOTROPY

The distinction between this effect and rheopexy is confused (Bauer and Collins, 1967). Both effects lead to a time-dependent increase in apparent viscosity on application of a shearing stress. Samyn and Wan (1967) suggest that the negative thixotropy they observed in their studies on clay suspensions was caused by breakdown of relatively large compact floccules, the breakdown leading to an increase in the interparticle contact in the system, which therefore exhibited an increase in apparent viscosity. These workers also observed an effect, analogous to the phenomenon of rheopexy in reversible thixotropic systems, which caused the clay suspensions to become more mobile under the influence of a mild turbulence; they suggested that it be called negative rheopexy.

The frequent occurrence of time-dependent effects in the rheological behaviour of disperse systems, and the long lag times that are often required before an effect is reversed after removal of shearing forces indicate that a knowledge of the histories of these systems is important in any rheological study. Standardised pre-treatments and handling techniques should therefore be employed whenever possible in such studies.

THE DETERMINATION OF FLOW PROPERTIES

Many types of viscometer have been devised for measuring flow properties. Some of these are capable of providing data for calculating viscosities in terms of fundamental units. However, the design of many instruments prevents their use in this manner since they are capable of providing data only in terms of empirical units that cannot be transformed into fundamental ones. These empirical instruments therefore provide relative measurements only; they are often used in quality control processes.

A review of the many types of viscometer that are available commercially is beyond the scope of this chapter, which is therefore limited to a consideration of the general principles of the more common types of instrument and to the methods mentioned in the BP (1968). The most frequently used viscometers involve the measurement of: (*a*) the rate of flow of liquids through capillary tubes, or (*b*) the rate of shear or shearing stresses in liquids moving between concentric cylinders or between cones and plates in rotational viscometers. These methods are therefore considered initially in the following sections.

1. Capillary Viscometers

These are the most commonly used instruments for Newtonian liquids. They measure the rate of flow of liquids through capillary tubes under the influence of gravity or an externally applied pressure. Applied pressure increases the range of usefulness of these instruments which can be used for investigating non-Newtonian behaviour.

(a) THE OSTWALD U-TUBE VISCOMETER (Fig. 7.4)

This is the most common type of instrument, in which flow through a capillary occurs under the influence of gravity. It is specified by the British Standards Institution and is described in the BP (1968). Various sizes are designated in the BP for use in determining the viscosity of Liquid Paraffin, Dextran Injection, and Iron Sorbitol Injection.

Liquid is introduced into the viscometer through the arm *V* until the level reaches the mark *G*. A fine pipette is used in this operation to avoid wetting the sides of the tube above *G*. The viscometer is fixed vertically in a thermostatted bath and allowed to attain the required temperature. The sample volume is adjusted and the liquid is sucked or blown into arm *W* until the meniscus is just above mark *E*. The suction or pressure is released and the time taken for the bottom of the meniscus to fall from *E* to *F* is noted.

(b) THE UBBELOHDE SUSPENDED LEVEL VISCOMETER (Fig. 7.5)

This is one of the many modifications of the U-tube viscometer. It is described in the BP (1968) where it is specified for use in the determination of the viscosity of a standard aqueous solution of Methylcellulose 450. A volume of liquid sufficient to fill bulb *C* during use is introduced through tube *V*. This volume must not be too large, otherwise the lower end of the ventilating tube *Z* will be blocked. The viscometer is fixed vertically in a thermostatted bath and allowed to reach temperature equilibrium. The ventilating tube *Z* is then closed and liquid is drawn into *C* by applying suction at *W* until the meniscus is just above the mark *E*. The liquid is held at this level by closing *W*, and *Z* is opened so that liquid drains away from below the capillary. *W* is finally opened, and the time taken for the meniscus to fall from *E* to *F* is observed. If the end

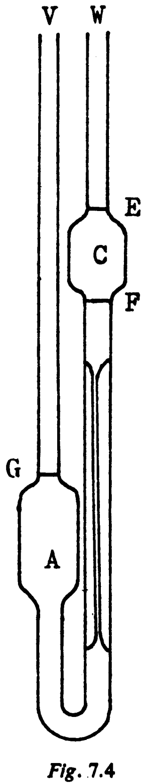


Fig. 7.4
Ostwald U-tube viscometer

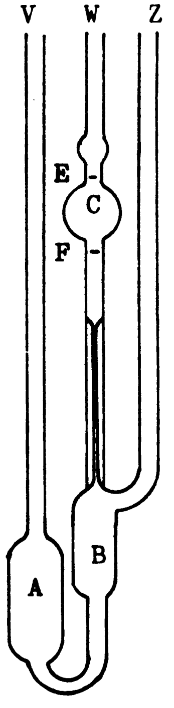


Fig. 7.5
Suspended level viscometer

of tube *Z* becomes blocked by the liquid at any time while the liquid is flowing through the capillary the determination must be repeated.

The suspended level viscometer avoids the difficulties associated with readjustment of the volume of liquid in the viscometer when measurements are taken at a series of temperatures, since the lower liquid level is fixed automatically at the bottom of the capillary.

CALCULATION OF RESULTS

The rate of flow of liquids through tubes is given by Poiseuille's equation, Eqn (7.6), where V is the volume of liquid flowing in time t, p is the pressure difference across the ends of the tube, r and l are the radius and length of the tube, respectively, and η is the coefficient of viscosity of the liquid.

$$\frac{V}{t} = \frac{\pi p r^4}{8\eta l} \qquad (7.6)$$

This equation applies to conditions of streamline flow only; i.e. when Reynolds' number is less than 1000 (*see* p. 146), and the dimensions of a chosen viscometer should therefore ensure that this type of flow does occur in the liquid under test. In a particular viscometer the values of r and l are constant, and Eqn (7.6) may therefore be written as—

$$\frac{V}{t} \propto \frac{p}{\eta} \qquad (7.7)$$

The pressure p will depend on the hydrostatic head of liquid, which depends, in turn, on the difference in heights of the liquid in the two arms of the U-tube viscometer. However, this difference will decrease as the liquid flows from the reservoir through the capillary. If h is the average difference in heights of a liquid with a density ρ, then the average hydrostatic pressure p is given by Eqn (7.8), where g is the gravitational constant.

$$p = h\rho g \qquad (7.8)$$

Substitution for p in Eqn (7.7) yields—

$$\frac{V}{t} \propto \frac{h\rho g}{\eta} \qquad (7.9)$$

However, V and h will be constant for a given viscometer,

$$\therefore \frac{1}{t} \propto \frac{\rho}{\eta} \qquad (7.10)$$

The ratio of viscosity, η, to density, ρ, of a liquid is termed the kinematic viscosity (ν); i.e.

$$\nu = \frac{\eta}{\rho} \qquad (7.11)$$

and it is usually expressed in terms of units known as stokes (St). The SI units are, in fact, $m^2 s^{-1}$ and $1 \text{ St} = 1 \times 10^{-4} m^2 s^{-1}$. Substitution of kinematic viscosity into Eqn (7.10) shows that

$$\frac{1}{t} \propto \frac{1}{\nu} \qquad (7.12)$$

or

$$\nu = ct \qquad (7.13)$$

where c is a constant that depends on the dimensions of the viscometer. The value of c can be determined from observations on the rate of flow of a liquid with a known kinematic viscosity. Oils with standard viscosities are available commercially for direct calibration of viscometers. Alternatively, a step-up procedure may be employed in calibrating a series of viscometers of increasing size. This involves using water as a primary standard for calibrating a viscometer with the slowest efflux time. The kinematic viscosity of a more viscous liquid is determined using the same viscometer, the liquid then being used to calibrate the next largest viscometer. The procedure is repeated until all the viscometers in the series are calibrated. However, the method is satisfactory only if the correction terms (see later) are negligible.

EXPERIMENTAL PRECAUTIONS AND CORRECTIONS

A suitable viscometer must be chosen to ensure that: (*a*) streamline flow occurs in the capillary, and (*b*) unnecessarily long flow times are avoided. The viscometer must be well-cleaned before use, and air bubbles must be excluded from the liquid inside it. Adequate temperature control should be provided and care should be taken to ensure that the viscometer is fixed vertically.

For accurate determinations corrections should be made to allow for effects that arise from: (i) the changes in velocity that occur when liquid enters a capillary from a reservoir and flows from the capillary into a lower reservoir; (ii) the use of part of the applied pressure to impart kinetic energy to the liquid; and (iii) surface tension, which tends to raise or lower the meniscus of a liquid in a tube and therefore affects the hydrostatic pressure. Details of methods of allowing for these corrections are given in the references cited in the Bibliography at the end of this chapter.

2. Rotational Viscometers

The basic design of this type of viscometer involves the relative rotation of two components, separated by the material under test, about the same axis of symmetry. Many designs are available but the most common geometries are those that involve concentric cylinders or a cone and plate.

(a) CONCENTRIC CYLINDER VISCOMETERS

A simple example of this type of viscometer is shown in Fig. 7.6. The outer cylinder *A*, which acts as a container for the liquid, can be rotated at different speeds, and the inner cylinder *B* is suspended freely by a torsion wire. Rotation of *A* produces movement of the liquid in the gap separating the cylinders, and a torque is transmitted to the inner cylinder. The resultant stress exerted on this latter cylinder is

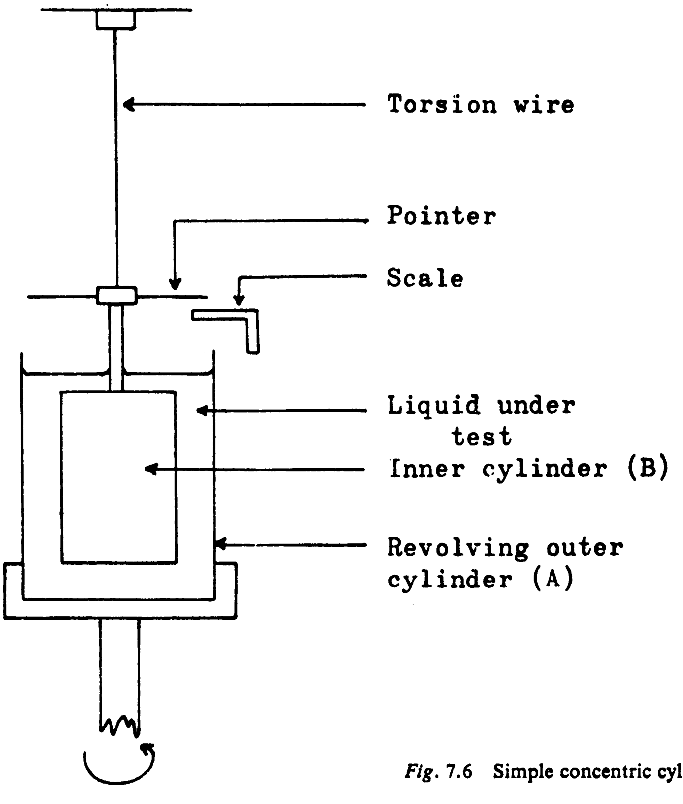


Fig. 7.6 Simple concentric cylinder viscometer

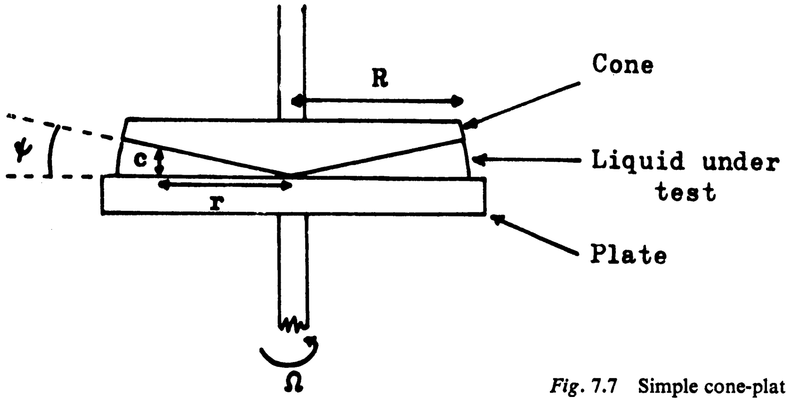


Fig. 7.7 Simple cone-plate viscometer

indicated by the angular deflection, θ, of a pointer which moves past a scale. Provided a steady laminar flow occurs in the liquid then the viscosity may be calculated from Eqn (7.14),

$$c\theta = \frac{4\pi h\omega\eta}{\dfrac{1}{r_1^2}-\dfrac{1}{r_2^2}} \qquad (7.14)$$

where c is the torsional constant of the wire, h is the height of inner cylinder covered by the liquid, ω is the angular velocity of the outer cylinder, and r_1 and r_2 are the radii of the inner and outer cylinders, respectively.

Other types of concentric cylinder viscometer are available in which: (i) the inner cylinder rotates under constant stress and the ensuing rate of shear is measured; (ii) the inner cylinder rotates at a constant rate and the shear stress developed at the inner cylinder wall is measured; and (iii) the inner cylinder rotates at a constant rate and the stress exerted on the outer cylinder is measured.

Concentric cylinder viscometers are particularly useful in studies on the flow properties of non-Newtonian systems. They allow measurements to be made over a wide range of shear rates so that flow diagrams can be constructed and used to characterise the behaviour of a fluid. However, these viscometers suffer from certain disadvantages; for example, a uniform rate of shear is not obtained throughout the sample, particularly if the gap between the cylinders is large. End effects are also important because Eqn (7.14) applies only to the transmission of torque between the walls of the outer and inner cylinders. The transmission of effects between the ends of the cylinders must therefore be taken into account by correction factors or eliminated by satisfactory viscometer design. Frictional heating may be troublesome at high rates of shear and constant temperature jackets are therefore necessary. In addition, filling and cleaning are often difficult if the gap width is narrow, and large quantities of liquid may be needed if the gap is wide or if the cylinders are of large dimensions.

(b) CONE-PLATE VISCOMETERS

One type of cone-plate viscometer, illustrated in Fig. 7.7, consists of a slightly conical disc, the apex of which just touches a flat plate. In this type the plate is rotated at a fixed speed and the torque transmitted through the sample to the cone is measured. The viscosity (η) may be calculated from Eqn (7.15)

$$\eta = \frac{3G/2\pi R^3}{\Omega/\psi} \qquad (7.15)$$

where, G is the torque on the cone, R is the radius of the cone, Ω is the radial velocity of the plate, and ψ is the angle between the cone and plate. This angle is usually less than $\pi/360$ rad. Greater angles have been used but edge effects then become important.

Similar variations in design to those of concentric cylinder viscometers are available. Thus, the cone or plate may be driven at a fixed speed and the torsional couple on the other component may be measured as in Fig. 7.7, or the torque that develops on the cone when it is driven at constant speed may be determined. The advantage of this latter design is that it allows easier temperature control by circulation of water from a thermostatted bath through the rigid plate. In addition, the mechanical construction of the plate adjustment mechanism is facilitated and greater precision of gap width control is obtained.

In any type of cone-plate viscometer the rate of shear at a given radius r is equal to the ratio of the linear velocity, Ωr, to the gap width, c, at that radius; i.e.

$$\text{shear rate} = \Omega r/c \qquad (7.16)$$

Since both these quantities are proportional to the radius r it follows that the shear rate is constant throughout the entire sample under test. This is an advantage compared with concentric cylinder viscometers. In addition, only a small amount of liquid is required in cone-plate viscometers, and cleaning is easy.

3. Other Viscometers Specified in the BP

(a) FALLING SPHERE VISCOMETER

This viscometer is based on Stokes' law, which states that a body falling through a viscous medium experiences a resistance or viscous drag that tends to oppose the motion of the body. Thus, when a body falls through a liquid under the influence of gravity, an initial period, during which acceleration of the falling motion occurs, is followed by motion at a uniform terminal velocity when the gravitational force is balanced by the viscous drag. Equation (7.17), in which g represents the acceleration due to gravity, may be applied when this terminal velocity, u, is attained by a sphere of diameter d and density ρ_s falling through a liquid of viscosity η and density ρ_l.

$$3\pi\eta du = \frac{\pi}{6} d^3 g(\rho_s - \rho_l) \qquad (7.17)$$

The left-hand side of this equation represents the viscous drag experienced by the sphere, and the right-hand side represents the force responsible for the downward motion of the sphere under the influence of gravity. Rearrangement of Eqn (7.17) shows that the viscosity may be calculated if the velocity u is measured and r, ρ_s, and ρ_l are known; i.e.

$$\eta = \frac{d^2 g(\rho_s - \rho_l)}{18u} \qquad (7.18)$$

In addition, since $\nu = \eta/\rho_l$ then Eqn (7.19) allows the kinematic viscosity (ν) of the liquid to be calculated.

$$\nu = \frac{d^2 g(\rho_s - \rho_l)}{18u\rho_l} \qquad (7.19)$$

In the derivation of these equations it is assumed that the sphere is falling through a medium of infinite dimensions. However, in practice the liquid is contained in a cylinder or fall tube of a finite size. A correction factor, known as the Faxen term, F, must therefore be introduced into these equations to allow for the effect of the walls of the fall tube on the motion of the sphere; e.g.

$$\nu = \frac{d^2 g(\rho_s - \rho_l)F}{18u\rho_l} \qquad (7.20)$$

The correction factor F may be calculated from Eqn (7.21), where D is the diameter of the fall tube.

$$F = 1 - 2{\cdot}104d/D + 2{\cdot}09d^3/D^3 \qquad (7.21)$$

The measurement of u may be carried out in the apparatus illustrated in Fig. 7.8. The liquid is placed in the fall tube, which is clamped vertically

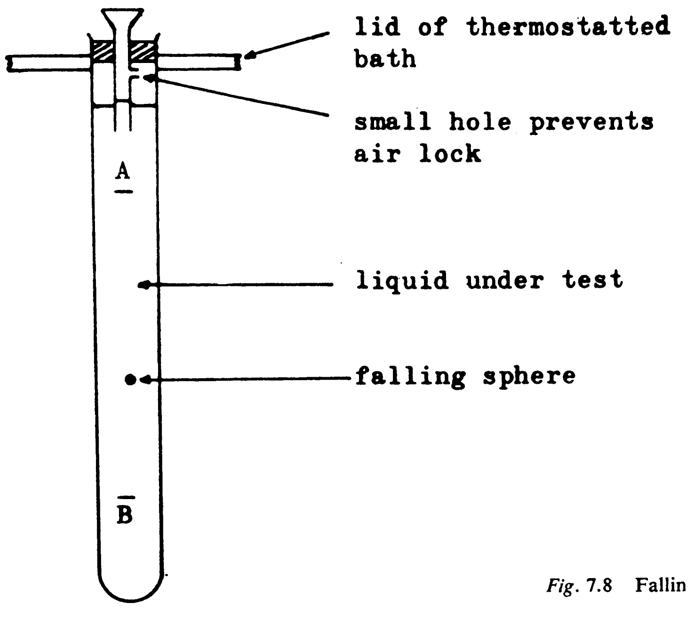


Fig. 7.8 Falling sphere viscometer

inside a constant temperature bath. Sufficient time is allowed for temperature equilibrium to be attained and for air bubbles to be removed from the liquid. A steel sphere is cleaned and allowed to reach the temperature of the test before being introduced into the fall tube through a narrow guide tube. The motion of the sphere is followed by a suitable telescope and the time taken for it to pass between the marks *A* and *B* is noted.

If the same sphere and fall tube are used, Eqn (7.20) reduces to

$$\nu = Kt(\rho_s/\rho_l - 1) \qquad (7.22)$$

where *K* is a constant that may be determined from observations using a liquid of known kinematic viscosity.

The falling sphere viscometer is described in the BP (1968), which specifies its use in the determination of the viscosity of a solution of Pyroxylin in acetone.

This type of viscometer is usually employed for Newtonian fluids, and a similar viscometer, known as the rolling sphere viscometer, involves the determination of the velocity of a rolling motion of a sphere down an inclined tube filled with liquid.

(b) REDWOOD VISCOMETER

This involves determining the time taken for a given volume of liquid to flow through a narrow orifice. The geometry of this orifice precludes the application of Poiseuille's equation, Eqn (7.6), to this viscometer and the efflux times are not related in a simple manner to the viscosities of the liquids as they are in capillary viscometers. The Redwood viscometer is an empirical instrument and the efflux times are therefore arbitrary measurements of viscosities, usually expressed as Redwood seconds.

Two sizes of viscometer are available (Redwood No. 1 and No. 2), and the BP specifies the No. 1 instrument for use in determining the viscosity of a standard solution of tragacanth relative to that of water.

Many other fundamental and empirical viscometers are available commercially. Reviews of these instruments are given by Van Wazer *et al.* (1963) and by Martin *et al.* (1964).

THE EFFECT OF TEMPERATURE

The viscosity of a liquid usually decreases with rise in temperature. The amount of such a decrease is often of the order of 1 to 10 per cent per °C.

The opposite effect may occur in certain cases; e.g. aqueous solutions of synthetic polymers, such as methylcellulose, which exhibit gel formation when the temperature is increased.

The commonly observed decrease in viscosity with rise in temperature may often be expressed by Eqn (7.23), where *A* and *B* are constants for a given liquid, *R* is the gas constant and *T* is the thermodynamic temperature.

$$\eta = Ae^{B/RT} \qquad (7.23)$$

THE FLOW PROPERTIES OF DISPERSE SYSTEMS

The presence of dispersed particles in a liquid continuous medium causes the streamlines of liquid around the particles to be distorted which, in turn, causes a further dissipation of energy. The viscosity of the system is therefore increased. The Einstein equation (7.24) provides a quantitative expression of this effect.

$$\frac{\eta}{\eta_0} = (1 + 2{\cdot}5\phi) \qquad (7.24)$$

In this equation, η and η_0 are the viscosities of the disperse system and the pure continuous medium, respectively, and ϕ is the volume fraction of the disperse phase; i.e. the volume of the disperse phase divided by the total volume of the system.

The assumptions made in the derivation of Eqn (7.24) restrict its application to dilute disperse systems. In fact, its use is theoretically limited to systems containing disperse particles which consist of rigid spheres. The equation has been verified experimentally using suspensions of glass beads and spherical fungal spores.

Several modifications of the Einstein equation have been suggested for use with systems containing higher concentrations of dispersed particles. These modifications usually take the form shown in Eqn (7.25)

$$\frac{\eta}{\eta_0} = (1 + 2{\cdot}5\phi + x\phi^2) \qquad (7.25)$$

Many values of the multiplying factor, *x*, have been reported but the effects of particle size distribution, collisions between globules, and degree of flocculation were not realised for a long time. Modified forms of the Einstein equation have been reviewed by Rutgers (1962) and by Sherman (1964, 1968).

Further effects arise from the solvation of the dispersed particles, since this increases their effective size and therefore affects the value of ϕ, from the shape and rigidity of the particles, and from electrostatic interactions between the double layers around particles and between charges that may be present on different parts of a flexible macromolecule.

1. Colloidal Dispersions

The following viscosity coefficients, often referred to in the literature, may be defined with respect to the Einstein equation.

(*a*) *Relative Viscosity* (or viscosity ratio)

This is given the symbol η_{rel} and is defined as the ratio of the viscosity, η, of a dispersion to that of its liquid continuous medium, η_0; i.e.

$$\eta_{\text{rel}} = \frac{\eta}{\eta_0} = 1 + 2{\cdot}5\phi \qquad (7.26)$$

(*b*) *Specific Viscosity* (or viscosity ratio increment)

This is given the symbol η_{sp} and is defined as the relative increase in viscosity caused by the presence of the dispersed phase; i.e.

$$\eta_{\text{sp}} = \frac{\eta - \eta_0}{\eta_0} = \frac{\eta}{\eta_0} - 1 \qquad (7.27)$$

$\therefore$ from Eqn (7.26)

$$\eta_{\text{sp}} = 2{\cdot}5\phi \qquad (7.28)$$

In addition, since the volume fraction is directly related to the concentration of the disperse phase, then Eqn (7.28) may be written in the form

$$\eta_{\text{sp}} = kc \qquad (7.29)$$

where k is a constant and c is the concentration, usually expressed in terms of grammes of disperse phase in 100 cm^3 of total dispersion.

It should be remembered that the above equations apply to dilute systems only. For more concentrated dispersions the last equation is usually written as a power series, Eqn (7.30), where α and β are constants that must be calculated from experimental observations.

$$\eta_{\text{sp}} = \alpha c + \beta c^2 + \cdots \qquad (7.30)$$

(*c*) *Reduced Viscosity* (or Staudinger's viscosity number)

This is the ratio of specific viscosity to concentration obtained by dividing the previous equation through by c; i.e.

$$\frac{\eta_{\text{sp}}}{c} = \alpha + \beta c + \cdots \qquad (7.31)$$

(*d*) *Intrinsic Viscosity* (or the Kraemer viscosity number)

This is given the symbol $[\eta]$. It is the intercept obtained by extrapolation of a graph of reduced viscosity versus concentration, as shown in Fig. 7.9.

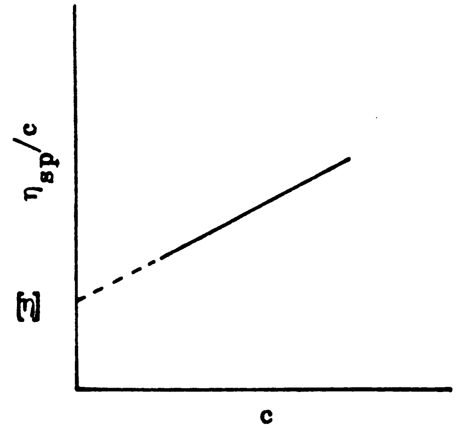


Fig. 7.9 Plot of reduced viscosity versus concentration showing determination of intrinsic viscosity

This coefficient is useful in calculating the average molecular weight, $\bar{M}$, of polymers, to which it is related by Eqn (7.32), where K and a are constants that must be obtained from separate experiments at fixed temperature with a given polymer-solvent system.

$$[\eta] = K\bar{M}^a \qquad (7.32)$$

The determination of intrinsic viscosity is used as a test in the BP for the standardisation of Dextran Injection, which is a plasma substitute, since the average molecular weight of this material will influence its retention in the capillaries.

Suspensions

The rheological properties of suspensions are markedly affected by the degree of flocculation that occurs in a given system. The presence of floccules tends to reduce the amount of free continuous medium, since much of it is entrapped within the floccules. The apparent viscosity of a flocculated suspension is therefore usually higher than that of a similar suspension in which the particles are deflocculated. In addition, marked flocculation produces a considerable amount of structure within a system. If the forces responsible for the formation of such structures are able to withstand weak stresses then the system will possess a yield value, below which it will exhibit the characteristics of solids. When the applied stress is sufficient to overcome the forces responsible for floccule formation a progressively greater amount of structural breakdown will occur as the shear stress is increased. Thus, flocculated suspensions will tend to exhibit plastic or pseudoplastic behaviour. If the breakdown and reformation of floccules is time-dependent, the phenomenon of thixotropy will be superimposed on these types of behaviour.

Such effects are not observed in deflocculated systems because the formation of structures is precluded in them. In dilute deflocculated suspensions the effect of the disperse phase on the viscosity is mainly concerned with the distortion of the streamlines around the particles, when the Einstein equation may be applicable. More concentrated deflocculated systems may exhibit dilatant behaviour. The explanation of this effect has been given previously.

The rheological properties of suspensions are of particular importance in formulating pharmaceutical products of this type. These properties should be such that:

(i) the product is easily administered (e.g. easily poured from a bottle or injected through a syringe needle),
(ii) sedimentation is either prevented, or if it does occur, redispersion is easy, and
(iii) the product has an elegant appearance.

Several methods may be employed to obtain these desired properties.

(*a*) *Deflocculated Particles in Newtonian Vehicles*

Sedimentation in these systems will produce a compact sediment or cake that is difficult to redisperse. The rate of sedimentation can be decreased by using a liquid continuous medium that possesses a high Newtonian viscosity. Such a system may introduce problems associated with pourability, and if any sedimentation does occur, redispersion is likely to be even more difficult.

(*b*) *Deflocculated Particles in Non-Newtonian Vehicles*

A non-Newtonian dispersion medium that exhibits plastic or pseudoplastic behaviour will retard the sedimentation of small particles since its apparent viscosity will be high at the small stresses associated with sedimentation. In addition, pourability and redispersion are facilitated by the decrease in apparent viscosity that occurs at the higher shear rates involved in these processes.

The commonly used suspending agents such as acacia, tragacanth, methylcellulose, and sodium carboxymethylcellulose impart such properties to aqueous solutions. Suspending agents such as bentonite and similar clays also impart thixotropic properties to aqueous dispersions. The deflocculated particles of a medicament become trapped in the gel network that forms in such dispersions during resting periods, and their sedimentation is retarded. Breakdown of the gel network by shaking reduces the apparent viscosity for a time sufficient to allow easy administration. Subsequent sedimentation is retarded by reformation of the gel network after a time lag.

(*c*) *Flocculated Particles in Newtonian Vehicles*

Flocculation of suspended particles will produce large volume sediments with a porous nature that are easy to redisperse. Separation of the product into two layers, i.e. sediment and supernatant liquid, is likely to occur, however, and produces an inelegant appearance. This can sometimes be prevented if flocculation is sufficient to produce a sedimentation volume that occupies the whole system. Careful control of flocculation is necessary to ensure that the product is easy to administer. Such control is usually achieved by using satisfactory concentrations of electrolytes, surface active agents or polymers. Variations in these concentrations will lead to effects that range from complete flocculation to deflocculation. In addition, the observed effect of a given concentration of one of these additives on a particular dispersion will depend on the characteristics of the disperse particles (e.g. specific surface area, presence of trace amounts of impurities). Since these factors are likely to vary between rather wide limits for commercially available materials, the use of controlled flocculation as the sole means of obtaining the desired properties in a pharmaceutical suspension is less satisfactory than using a structured vehicle, for large-scale production.

(*d*) *Flocculated Particles in Non-Newtonian Vehicles*

This type of system allows the advantages of the previous two methods to be utilised. In addition, variations in the properties of the material to be suspended are less likely to influence the performance of a product made on a large scale, and less difference will be observed between batches of product made from the same flow sheet.

Further details of these methods are given by Martin and Swarbrick (1966), and the theory of coarse particle suspension formulation has been reviewed by Hiestand (1964).

Emulsions

The flow properties of emulsions have a marked influence on their usefulness as dosage forms for medicinal compounds. Except for very dilute systems, emulsions are non-Newtonian, and in pharmaceutical emulsions the most commonly observed behaviour involves a decrease in apparent viscosity with increase in rate of shear. This may range from pseudoplastic behaviour in emulsions with a moderate concentration of disperse phase to plastic behaviour with considerable yield values in concentrated emulsions and creams. In addition, pharmaceutical emulsions often exhibit some degree of thixotropy. These properties are particularly satisfactory from a user's point of view since they allow easy application of a topical preparation or easy administration of a product intended for oral use. Furthermore, the physical stability of emulsions against the effects of creaming is aided by the high apparent viscosities of these systems at low rates of shear.

Many factors contribute to the rheological properties of emulsions. These have been listed and discussed by Sherman (1964, 1968) and include the effect of the viscosities and chemical natures of the continuous and disperse phases, the volume concentration of the disperse phase and its globule size distribution, and the chemical structures and concentrations of the emulsifying agent and additional stabilising agents. Talman, Davies, and Rowan (1967, 1968) have suggested that the observed rheological properties of emulsions containing a long-chain alcohol as one of the components of a complex interfacial film could be accounted for by the migration of this oil-soluble component into the aqueous phase, where it produces a viscous gel.

BIBLIOGRAPHY

DINSDALE, A. and MOORE, F. (1962) *Viscosity and its Measurement*. The Institute of Physics and the Physical Society, London.

MARTIN, A. N., BANKER, G. S., and CHUN, A. H. C. (1964) *Advances in Pharmaceutical Sciences. Vol. 1.* (H. S. Bean, A. H. Beckett, and J. E. Carless, Eds). Academic Press, London. Chapter 1.

VAN WAZER, J. R., LYONS, J. W., KIM, K. Y., and COLWELL, R. E. (1963) *Viscosity and Flow Measurement*. Interscience, New York.

REFERENCES

BARRY, B. W. and WARBURTON, B. (1968) Some rheological aspects of cosmetics. *J. Soc. cosmet. Chem.*, **19**, 725–744.

BAUER, W. H. and COLLINS, E. A. (1967) *Rheology; Theory and Applications. Vol. IV.* (F. R. Eirich, Ed.). Academic Press, London. Chapter 8.

HIESTAND, E. N. (1964) Theory of coarse suspension formulation. *J. pharm. Sci.*, **53**, 1–18.

MARTIN, A. N., BANKER, G. S., and CHUN, A. H. C. (1964) *Advances in Pharmaceutical Sciences. Vol. 1.* (H. S. Bean, A. H. Beckett, and J. E. Carless, Eds). Academic Press, London. Chapter 1.

MARTIN, A. N. and SWARBRICK, J. (1966) *American Pharmacy*. 6th ed. (J. B. Sprowls and H. M. Beal, Eds). Lippincott, Philadelphia. Chapter 8.

RUTGERS, I. R. (1962) Relative viscosity and concentration. *Rheol. Acta*, **2**, 305–348.

SAMYN, J. C. and WAN, Y. J. (1967) Negative thixotropy in flocculated clay suspensions. *J. pharm. Sci.*, **56**, 188–191.

SHERMAN, P. (1964) The flow properties of emulsions. *J. Pharm. Pharmac.*, **16**, 1–25.

SHERMAN, P. (1968) *Emulsion Science*. (P. Sherman, Ed.). Academic Press, London. Chapter 4.

TALMAN, F. A. J., DAVIES, P. J., and ROWAN, E. M. (1967) The rheology of some oil-in-water emulsions stabilized by complex condensed films. *J. Pharm. Pharmac.*, **19**, 417–425.

TALMAN, F. A. J., DAVIES, P. J., and ROWAN, E. M. (1968) The effect of the concentration of the water-soluble component on the rheology of some emulsions containing long-chain alcohols. *J. Pharm. Pharmac.*, **20**, 513–520.

VAN WAZER, J. R., LYONS, J. W., KIM, K. Y., and COLWELL, R. E. (1963) *Viscosity and Flow Measurement*. Interscience, New York.