

2 – APPLICATIONS OF ANALYTICAL CHEMISTRY

2.1 – Introduction

What do the following branches of science have in common?

- archaeology
- environmental science
- food science
- forensic science
- medical research
- perfume science
- astrobiology

The answer is that they all involve the identification of chemical compounds that occur in very small quantities in mixtures containing possibly several hundred different substances. This chapter looks at how we can separate these highly complex mixtures into their individual components, and how we can then find out the chemical structures of these compounds.

The techniques of separation, purification and identification described in this chapter have been the subject of continuous development over the last 80 years. During that time they have become increasingly more powerful, accurate, sensitive, and miniaturised.

Astrobiology

The extreme portability of some of these techniques was demonstrated by the fact that both the Viking Mars landers, and the abortive Beagle 2 Mars lander, included on-board gas chromatographs and mass spectrometers that weighed just a few kg, but their sensitivity was such that they could detect nanogram (1×10^{-9} g) quantities of substances.

Why did these craft contain these instruments?

Life on Mars?

The purpose of the Mars probes was to search for signs of life (past or present) on the Red Planet. Once the presence of water (albeit frozen) had been confirmed in the surface rocks of Mars, scientists have been conjecturing that there might be some form of life there. But how could they be sure? The easiest way was to try to discover whether the small molecules which are the hallmarks of life could be found in the Mars environment.

The 1976 Viking lander was able to shovel up soil samples from the Martian surface and use an automated gas chromatograph-mass spectrometer (GCMS) to investigate them. The GCMS baked a small soil sample in an oven to drive off any volatile gases present, separated these volatiles using a gas chromatograph, and analysed their composition with a mass spectrometer. The experimental set-up worked perfectly, but disappointingly found none of the small amino and carboxylic acids that might have been expected if living organisms had been present in the soil.

The Beagle 2 Mars lander unfortunately never sent back any signals after it left the orbiting craft in December 2004. It is assumed that it crashed into the planet's surface and was damaged or destroyed. It had on board a sophisticated mass spectrometer that could analyse the $^{12}\text{C}:^{13}\text{C}$ ratio in samples of CO_2 . By heating up soil samples with pure oxygen, any "organic" compounds derived from life forms would be oxidised to CO_2 . By comparing the $^{12}\text{C}:^{13}\text{C}$ ratio in this sample of CO_2 to that of a CO_2 sample from the Martian atmosphere (which contains 95% CO_2), it would have been possible to determine whether the carbon in the soil sample was of biological origin. This is because it has been found that in every biosystem on Earth, organisms concentrate ^{12}C at the expense of ^{13}C , and it was assumed Martian organisms would do likewise.

The jury is still out on the question of life on Mars. Some results give a strong indication that there could be simple forms of life there, whereas other experiments have given negative results. One thing is certain, however: the use of chemical instrumentation and analysis, like the methods described in this chapter, will play a crucial part in providing a definite answer to this question.

2.2 – Determining structures

By the end of this section, you should be able to:

- outline the principles of the mass spectrometer
- explain the meaning of the terms base peak, molecular ion and fragmentation pattern as applied to mass spectra
- suggest the identity of a molecule based on its fragmentation products
- deduce the number of carbon atoms in a compound using the $M+1$ peak
- deduce the presence of chlorine or bromine in a molecule by use of the $M+2$ peak
- outline the principles of ^1H nuclear magnetic resonance
- explain how the chemical environment of a proton affects the magnetic field it experiences, and hence the frequency of radiation absorbed at resonance
- explain the use of the δ scale, and the use of TMS as an internal standard
- describe the effects of adjacent protons on the magnetic field experienced by a given proton
- predict, from the integration of the peak areas in an NMR spectrum, the number of protons in each group present in a given molecule
- predict, from the splitting pattern of a particular peak in an NMR spectrum, the number of protons adjacent to a given proton
- suggest, from an NMR spectrum, possible structures for a given molecule
- describe how the addition of D_2O can be used to identify protons on oxygen or nitrogen atoms
- outline the principles of X-ray crystallography
- appreciate the application of X-ray crystallography in the elucidation of the structures of large molecules of biochemical importance

Mass Spectrometry

The principle of mass spectrometry is very simple, although modern mass spectrometers are very sophisticated, precision-made instruments, capable of determining molecular masses to an accuracy of 1 part in 100 000.

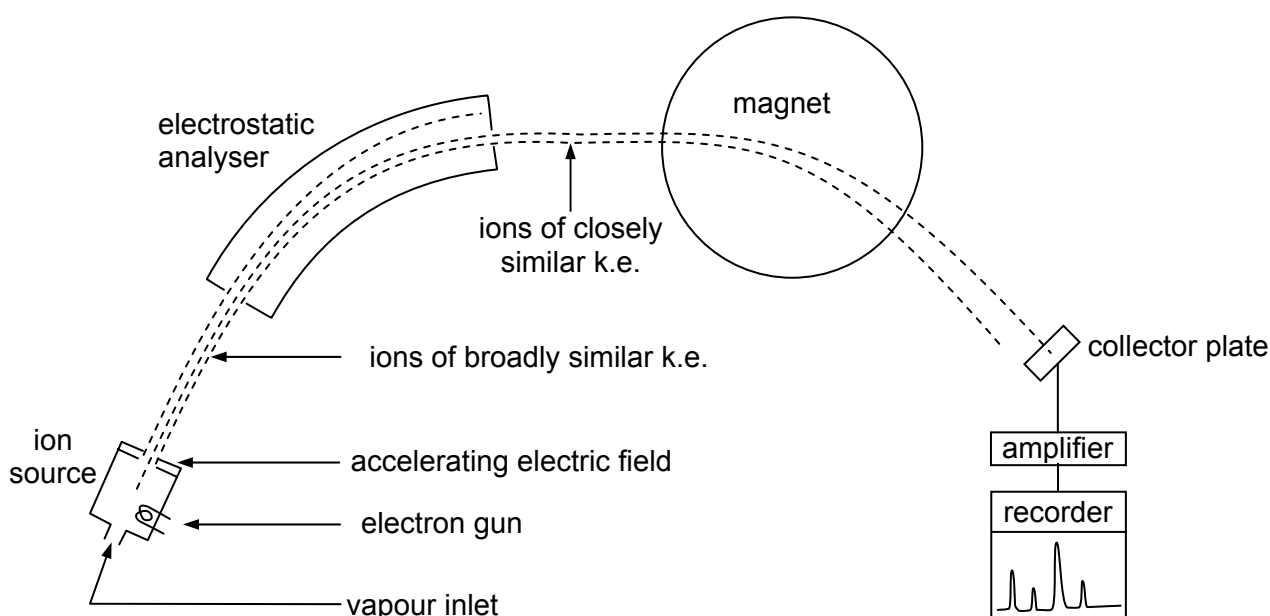
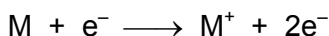


Figure 2.1 – diagram of a mass spectrometer

Six processes occur in a mass spectrometer (see Figure 2.1).

1. If not already a gas, the compound is vaporised in an oven. Only a small vapour pressure is required, since the interior of the instrument is kept under a high vacuum.
2. Electrons are fired at the gaseous molecules. These knock off other electrons from some of the molecules:



3. The gaseous ions are accelerated by passing through an electric field (at a voltage of 5-10 kV). At this stage they can be travelling at up to $2 \times 10^5 \text{ m sec}^{-1}$ (about 1/1000 the speed of light)
4. They then pass through an electrostatic analyser, which selects ions of kinetic energy within a narrow range by using an electric field.
5. The fast-moving ions now pass through the poles of an electromagnet, where they are deflected (see below).
6. The deflected ions pass through a narrow slit and are collected on a metallic plate connected to an amplifier. For a given strength of magnetic field, only ions of a certain mass pass through the slit and hit the collector plate. As the (positive) ions hit the plate, they cause a current to flow through the amplifier. The more ions there are, the larger the current.

The ions may travel a metre or so through the spectrometer. In order for them to do this without hitting into too many air molecules (which would deflect them from their course), the inside of the spectrometer is evacuated to a very low pressure.

The equation governing the deflection of ions in the magnetic field is as follows:

$$r = \sqrt{\frac{2mV}{eB^2}} \quad \text{where } r = \text{radius of circular path in the magnetic field}$$

m = mass of ion

V = accelerating voltage

e = electrical charge on the ion

B = strength of magnetic field

The accelerating voltage V is usually kept constant. We thus see that the radius of curvature is proportional to $\sqrt{m/e}$, but inversely proportional to B . To obtain a mass spectrum, therefore, the current through the electromagnet is changed at a steady rate. This causes the magnetic field, B , to change its strength, and hence allows ions of different mass/charge values to pass successively through the slit. A mass spectrum is produced, which plots (ion current) against (electromagnetic current), which is equivalent to (relative abundance) against (mass/charge (m/e) ratio). In practice, most ions that are formed in a mass spectrometer have a charge of +1, and so the x-axis is a measure of the masses of the ions. The y-axis normally shows the abundances of the peaks as a percentage of that of the most abundant peak (known as the base peak). The base peak usually corresponds to a particularly stable fragment of the molecules under investigation.

There are three main ways in which mass spectrometry is applied to the determination of the structures of organic compounds.

1. By measuring the relative heights of the molecular ion (M) peak and the ($M+1$) peak we can determine the number of carbon atoms in a molecule, and by using the ($M+2$) and ($M+4$) peaks (if any) we can identify halogen-containing compounds.
2. By measuring the accurate mass of a molecular ion we can determine its molecular formula.
3. By identifying the fragments produced when an ion breaks up inside a mass spectrometer we can often piece together the structure of the parent molecule.

We shall look at each of these techniques in turn.

$^{12}\text{C} : ^{13}\text{C}$ ratio

Naturally-occurring carbon is composed of 98.9% ^{12}C and 1.1% ^{13}C (along with extremely small, and variable, amount of ^{14}C). Although the $^{12}\text{C} : ^{13}\text{C}$ ratio is very small for compounds like methane which contain just one carbon atom, the ratio increases in proportion to the number of carbon atoms, as the

chances of a molecule containing at least one ^{13}C atom increase. The formula relating the (M+1)/(M) ratio the number of carbon atoms is:

$$n = \frac{100}{1.1} \left(\frac{A_{M+1}}{A_M} \right) \quad \text{where } n = \text{number of carbon atoms}$$

A_{M+1} = the abundance of the M+1 peak

and A_M = the abundance of the molecular ion, M, peak.

Example: Compound A has a molecular ion at an m/e value of 120, and relative abundance 23%, and a peak at m/e 121 with a relative abundance of 2%. How many carbon atoms are in a molecule of A?

Answer: $n = (100/1.1) \times (2/23) = 7.91$

The nearest whole number is 8. Therefore compound A has 8 carbon atoms per molecule.

SAQ 1. Compound B contains carbon, hydrogen and oxygen only. Its mass spectrum contains a molecular ion peak at m/e = 102 (relative abundance 35%) and an M+1 peak at m/e 103 (1.5%).

Calculate the number of carbon atoms in the molecule, and hence deduce the number of oxygen atoms it contains, and its molecular formula.

(M), (M+2) and (M+4) peaks

Both chlorine and bromine naturally occur as mixtures of two isotopes, with the relative abundances shown in the table.

element	isotope	relative abundance	approximate ratio
chlorine	^{35}Cl	75.8%	3:1
	^{37}Cl	22.4%	
bromine	^{79}Br	50.5%	1:1
	^{81}Br	49.5%	

The mass spectrum of a compound containing one of these elements should therefore show two molecular ions, one with an m/e value two mass units higher than the other. If the molecule contains **two** chlorine atoms, (or two bromine atoms, or one of each) we should expect to see **three** molecular ions, at m/e values of M, M+2 and M+4. The ratio of the M/(M+2) peak should reflect the natural abundances given in the table (i.e. 3:1 for chlorine; 1:1 for bromine). The situation is more complicated for molecules containing two halogen atoms. For example dibromomethane, CH_2Br_2 , shows three molecular ion peaks at m/e 172, 174 and 176 in the ratio 1:2:1, corresponding to $\text{CH}_2^{79}\text{Br}_2$, $\{\text{CH}_2^{79}\text{Br}^{81}\text{Br} + \text{CH}_2^{81}\text{Br}^{79}\text{Br}\}$ and $\text{CH}_2^{81}\text{Br}_2$. The three molecular ion peaks in dichloromethane, CH_2Cl_2 , however, appear in the ratio 9:6:1, due to the greater natural abundance of the ^{35}Cl isotope.

SAQ 2. Calculate the M : M+2 : M+4 ratio for CH_2BrCl

Molecular formulae from accurate masses

Using very high-resolution mass spectrometry, we can measure m/e ratios to an accuracy of 5 significant figures (1 part in 100 000). By this means, it is not only possible to measure the M_r value of a compound, but also to determine its molecular formula. We can do this because the accurate relative atomic masses of individual atoms are not exact whole numbers.

For example:

The following three compounds all have an approximate M_r of 70:

name	structure	molecular formula
pentene	$\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}=\text{CH}_2$	C_5H_{10}
aminopropanonitrile	$\text{CH}_3\text{CH}(\text{NH}_2)\text{CN}$	$\text{C}_3\text{H}_6\text{N}_2$
but-1-ene-3-one	$\text{CH}_2=\text{CHCOCH}_3$	$\text{C}_4\text{H}_6\text{O}$

By using the following accurate atomic masses we can calculate their accurate M_r values.

element	accurate relative atomic mass
H	1.0078
C	12.000
N	14.003
O	15.995

Answer: The accurate masses are as follows:

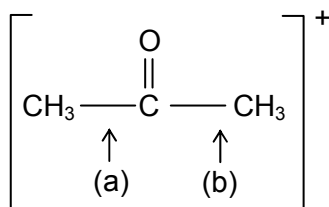
$$\begin{aligned}
 \text{C}_5\text{H}_{10} &= 5 \times 12.000 + 10 \times 1.0078 &= 70.078 \\
 \text{C}_3\text{H}_6\text{N}_2 &= 3 \times 12.000 + 6 \times 1.0078 + 2 \times 14.003 &= 70.054 \\
 \text{C}_4\text{H}_6\text{O} &= 4 \times 12.000 + 6 \times 1.0078 + 15.995 &= 70.045
 \end{aligned}$$

The last two are quite close together. They differ by 9 parts in 70 000, or about 0.13%. This is well within the capabilities of a high-resolution mass spectrometer.

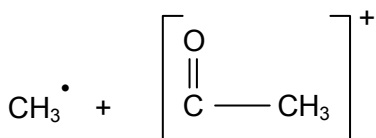
SAQ 3. Explain whether a molecule having an accurate mass of 60.0574 is 1,2-diaminoethane, $\text{C}_2\text{H}_8\text{N}_2$, or propan-1-ol, $\text{C}_3\text{H}_8\text{O}$.

The use of fragmentation patterns

If the ionising electron beam in a mass spectrometer has enough energy (anything from 25 – 70 electron-volts), the molecular ions formed by the loss of an electron can undergo bond fission, and molecular fragments are formed. Some of these will carry the positive charge, and therefore appear as further peaks in the mass spectrometer. Take propanone as an example:

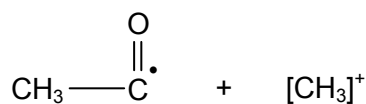


cleavage at (a)
gives:



$m/e = 43$

cleavage at (b)
gives:



$m/e = 15$

We therefore expect the mass spectrum of propanone to contain peaks at m/e 15, and 43, as well as the molecular ion peak at 58 (see Figure 2.2).

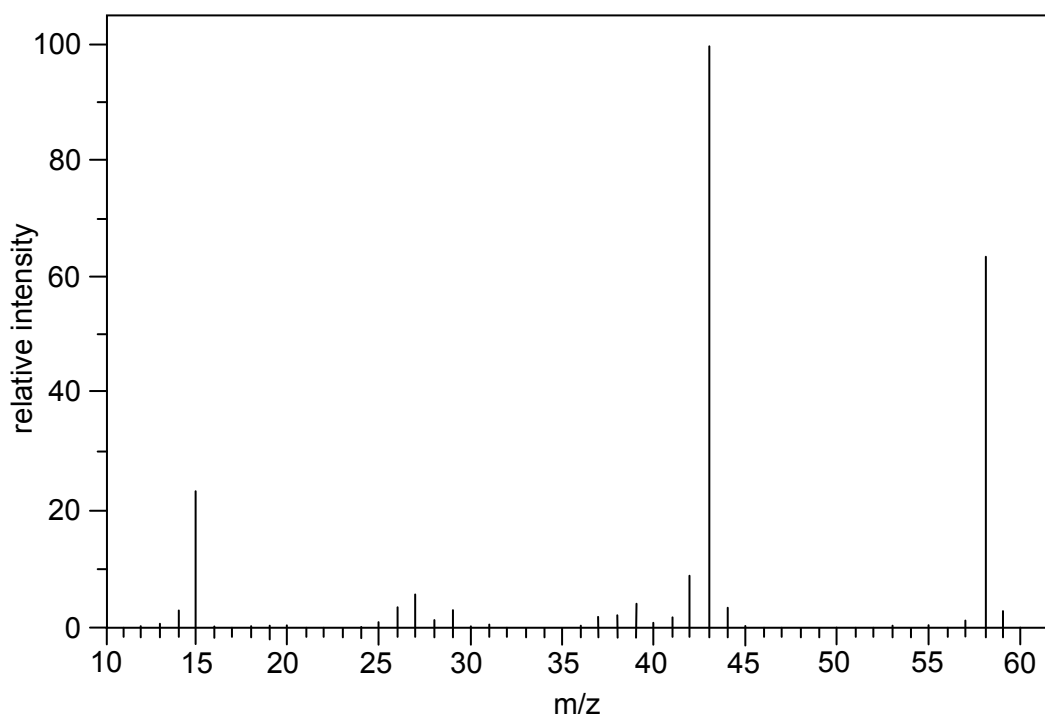


Figure 2.2 – mass spectrum of propanone

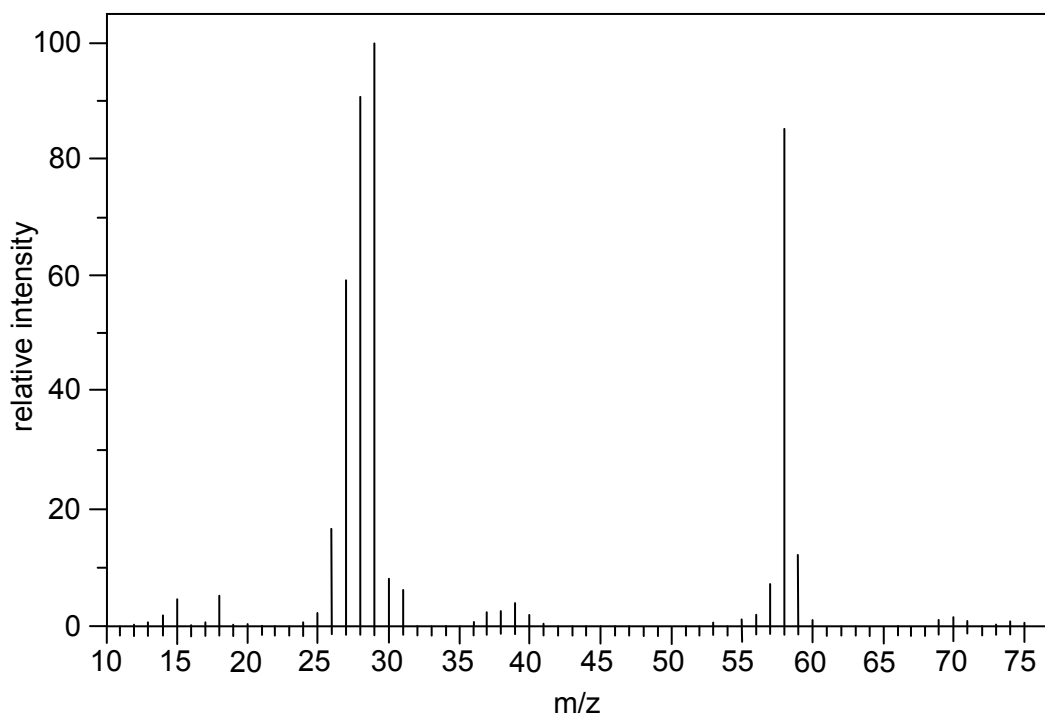
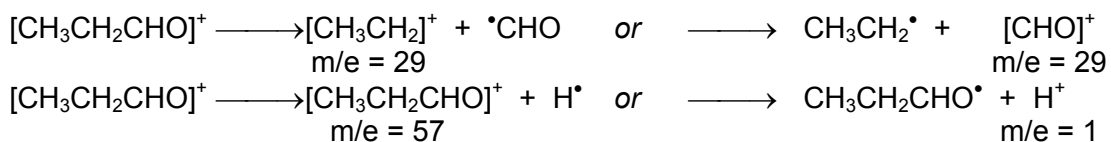


Figure 2.3 – mass spectrum of propanal

The fragmentation pattern can readily distinguish between isomers. Compare Figure 2.2 with Figure 2.3 which shows the mass spectrum of propanal. Here there is no peak at $m/e = 15$, nor one at $m/e = 43$. Instead, there are peaks at $m/e = 57$ and several from $m/e = 26$ to $m/e = 29$. This is readily explained by the following fragmentations.



SAQ 4. Use the values of accurate relative atomic masses in the table above to see whether it would be possible to decide whether the peak at $m/e = 29$ is due to CH_3CH_2^+ or CHO^+ .

An example of how these three techniques can help us to determine the structure of an unknown compound is given by the following spectrum of compound B.

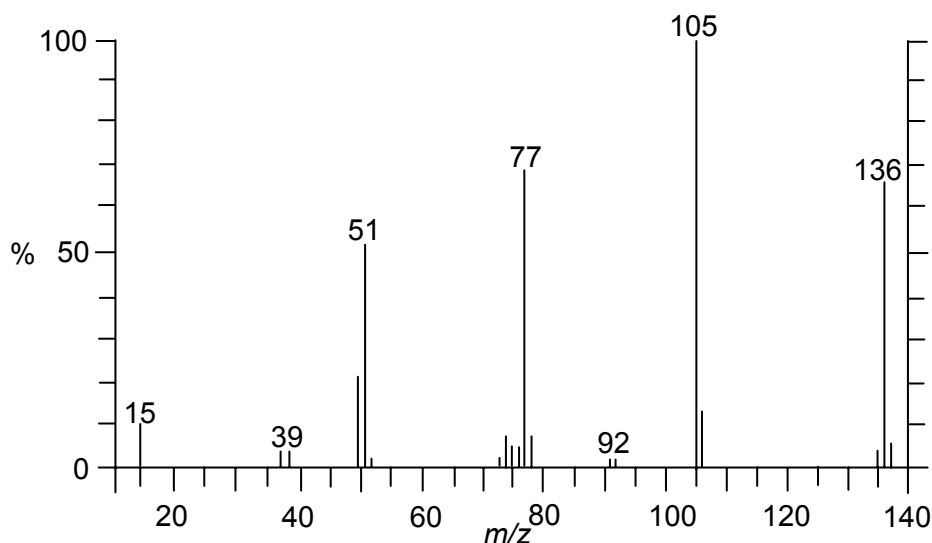
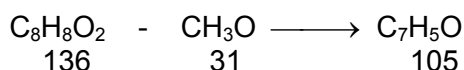


Figure 2.4 – mass spectrum of compound B

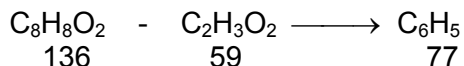
The M:M+1 ratio of 11.5:1 suggests that a molecule of compound B contains 8 carbon atoms, and an accurate determination of its relative atomic mass suggests its molecular formula is $\text{C}_8\text{H}_8\text{O}_2$. The high carbon : hydrogen ratio suggests a compound containing an aryl ring. There are, however, many isomers with this formula: phenyl ethanoate, methyl benzoate and methyl benzoic acid are just three examples.

The base peak (= most stable fragment) in the spectrum is at $m/e = 105$. The other significant peaks are at m/e 15, 51 and 77. We can analyse these peaks by either taking their M_r values and guessing at their molecular formulae, or by calculating the molecular formula of the fragment(s) that have been lost from the molecular ion at $m/e = 136$

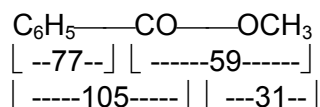
Thus: the peak at $m/e = 105$ represents the loss of $136 - 105 = 31$



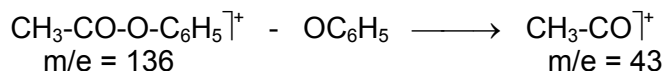
and the peak at $m/e = 77$ represents the loss of $136 - 77 = 59$



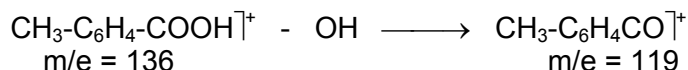
The peak at m/e 105 (28 units higher than C_6H_5 at m/e 77) can be identified with $\text{C}_6\text{H}_5\text{CO}^+$ (acyl cations are particularly stable, so are often the base peaks in mass spectra). This suggests that compound B is methyl benzoate:



Just as important as the assigning of structures to the fragments of the molecule is the awareness of the *absence* of fragments that might have been expected in the mass spectrum of alternative structures. Thus, bearing in mind the stability of acyl cations mentioned above, we might have expected phenyl ethanoate to have produced a fragment at m/e 43,



and methylbenzoic acid to have produced a fragment at m/e 119.



The absence of both these peaks confirms that B cannot be either of these compounds.

Nuclear Magnetic Resonance (NMR) spectroscopy

Magnetic properties of the nucleus

The nuclei of hydrogen atoms spin about an axis. Because the nuclei are positively charged, this spin is associated with a circulation of electric charge. Circulating charges give rise to magnetic fields, so the spinning ^1H nucleus has a magnetic moment, rather like the magnet of a compass needle. When put in an *external* magnetic field the nuclei tend to turn to a preferred orientation in which the nuclear magnet is aligned **with** the external field (like a compass needle does in the Earth's magnetic field). Another, less favoured, orientation is when the nuclear magnet is aligned **against** (i.e. opposed to) the external field. Nuclei obey the laws of quantum mechanics, and according to quantum laws these are the only two orientations allowed for nuclei such as ^1H , which have a nuclear spin of $\frac{1}{2}$ a unit. This situation is illustrated in Figure 2.5.

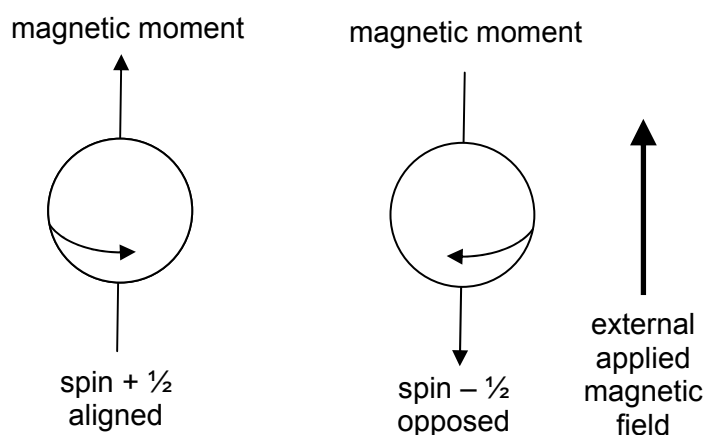


Figure 2.5 - nuclear spin

In the absence of an applied field, the spin states of a given nucleus are of equal energy. As can be seen in Figure 2.6, however, the spin states in an external magnetic field are no longer of equivalent energy and the two spin states occupy two different energy levels.

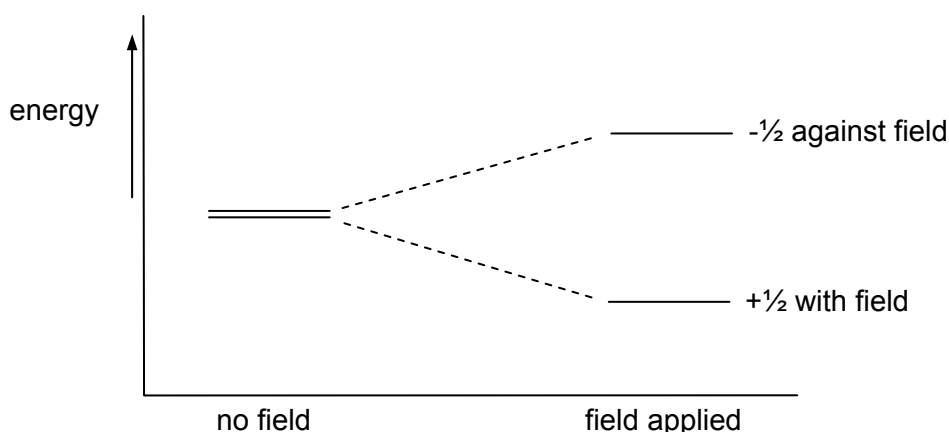


Figure 2.6 – spin states in an external magnetic field

The phenomenon of nuclear magnetic resonance occurs when nuclei aligned with the applied field absorb energy (ΔE in Figure 2.6) and change their spin orientation with respect to the field. The

energy difference between the two spin states is related to the frequency of electromagnetic radiation by the Planck equation:

$$\text{energy absorbed} = \Delta E = hf$$

where f = frequency of radiation, and h = Planck's constant.

The energy gap separating the two levels increases as the strength of the external field increases. The strength of the magnetic field is measured in teslas, T (one tesla is about 10 000 times as strong as the Earth's magnetic field). For a typical field of 9.4 Tesla, T, $\Delta E = 3 \times 10^{-25}$ J per nucleus, and the frequency of radiation absorbed is about 400 MHz, which is in the UHF radio-frequency region of the spectrum. Because the energy gap, ΔE , is so small (only 0.2 J mol^{-1} , compared to the average thermal energy of 1300 J mol^{-1} at room temperature) the two spin states are approximately equally populated, the difference in population being only 1 part in 30 000. That is, for every 30 000 nuclei in the lower energy spin state, there are 29 999 nuclei in the upper spin state. The consequence of this is that the net absorption of radiation is very small, and so a very sensitive detection system is required. In addition, as we shall see later, the frequencies at which different protons within molecules absorb energy differ only very slightly from each other, so the detection system has to be able to measure frequencies to an accuracy of 1 part in 100 million, and the electromagnet has to provide a magnetic field of comparable stability.

The NMR spectrometer

The basic elements of a typical NMR spectrometer are shown in Figure 2.7. About 20mg of the sample is dissolved in about 0.5 cm^3 of a solvent that contains no protons, e.g. tetrachloromethane, CCl_4 , or deuteriochloroform, CDCl_3 , or heavy water, D_2O , and a drop of tetramethylsilane (TMS), $(\text{CH}_3)_4\text{Si}$, is added to act as an internal reference. [TMS is used as a reference because it is volatile, inert and produces a strong singlet peak at a higher field than most ^1H absorptions in organic molecules, so its peak does not interfere with the other peaks.] The solution is placed in a precision glass tube of 0.5 cm diameter, to a depth of 2-3 cm. The sample tube is then suspended between the poles of the permanent magnet. The effect of slight variations in the magnetic field is minimised by spinning the sample at 20-30 revolutions per minute. Also in the gap between the pole pieces of the magnet is a coil attached to a radio-frequency (UHF) oscillator which supplies the energy for the nuclear spins to "flip". At right angles to this energising or oscillator coil, there is a second coil coupled to an amplifier and recorder. It is this second coil which acts as the detector. When no energy is being absorbed by the sample, the detector coil picks up none of the energy from the oscillator coil. When the sample absorbs energy, however, the re-orientation of nuclear spins induces a signal in the detector coil; this signal is recorded as a resonance signal. Most modern NMR spectrometers are Fourier Transform machines, which are designed to operate with a fixed external field and a range of frequencies.

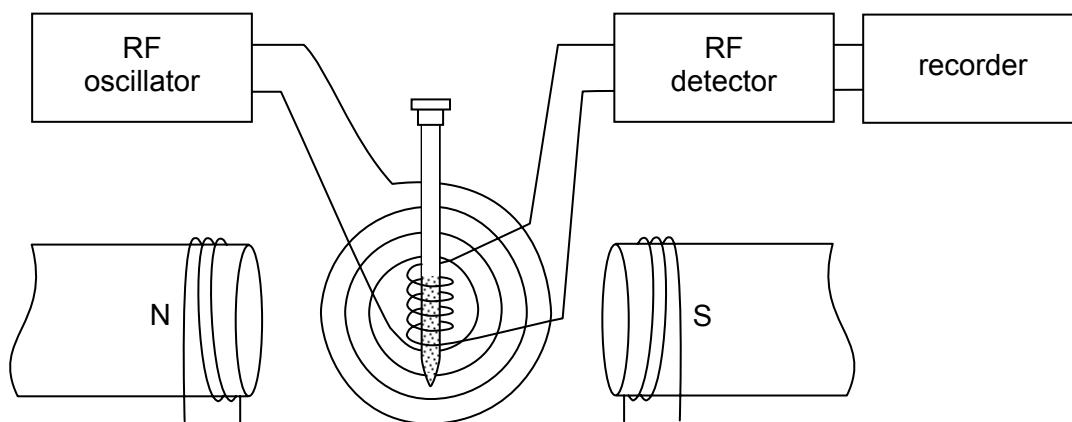


Figure 2.7 – the NMR spectrometer

The ^1H NMR spectra of organic compounds

When the frequency of radiation supplied corresponds to the energy difference ΔE , the population of the higher energy state increases as radiation is absorbed. The equilibrium population ratio is re-established by redistributing the energy amongst other energy states in the molecule (e.g. the translational movement of molecules). This takes a fraction of a second.

Not all the ^1H atoms within a molecule absorb energy at the same frequency. This is because the frequency at which a proton absorbs radiation depends on the strength of the *local* magnetic field around it. Even in a *constant external* field, protons in different chemical environments within a molecule absorb at different frequencies, because the local magnetic field they experience depends on the electrical and magnetic environment around them.

The electrons within molecules are usually 'paired' (i.e. they occur as pairs of electrons spinning in opposite directions). When a molecule is placed in an external field, the electron pairs rotate in their orbits in such a way so that they produce a magnetic field which opposes the external field. This phenomenon is called *diamagnetism*. The effect is to 'shield' nearby protons from the external field. This in turn reduces the frequency at which they absorb energy when they flip back from their lower to their higher energy state (see Figure 2.8).

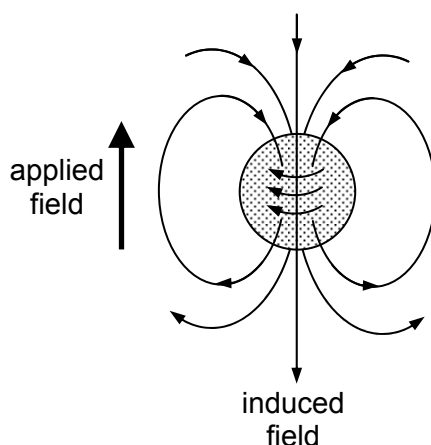


Figure 2.8 – shielding

When, however, a proton is near an electronegative atom within a molecule, the bonding electrons are drawn away from the proton to the electronegative atom. The proton is less shielded from the external magnetic field, and hence it absorbs radiation at a higher frequency. The effect is very pronounced if the proton is attached to a benzene ring. In this situation the mobile delocalised π electrons in the ring can create a strong diamagnetic effect, opposing the external field. This has the effect of *strengthening* the magnetic field within the vicinity of the protons (see Figure 2.9).

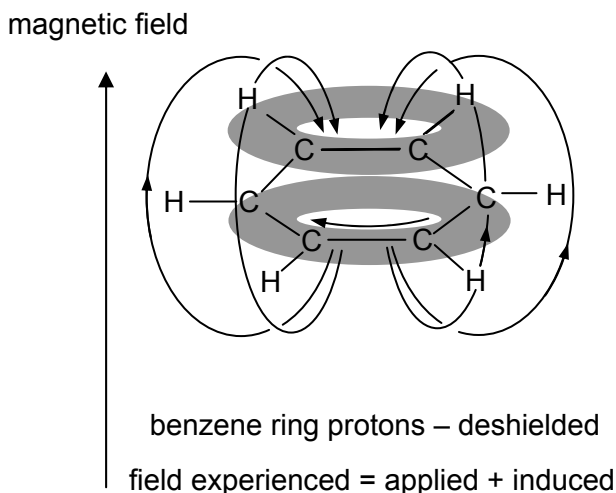


Figure 2.9 – deshielding

The ^1H NMR spectrum of ethanol

An NMR spectrum consists of a graph of absorbance against frequency. The frequency scale is measured in units of **chemical shift** (symbol δ). The chemical shift of a proton is the difference between its absorption frequency and that of TMS, measured in part per million (ppm). Most protons in organic molecules resonate within 10 ppm of TMS, and by convention, the zero point of the scale (TMS) is on the right hand side (Figure 2.10).

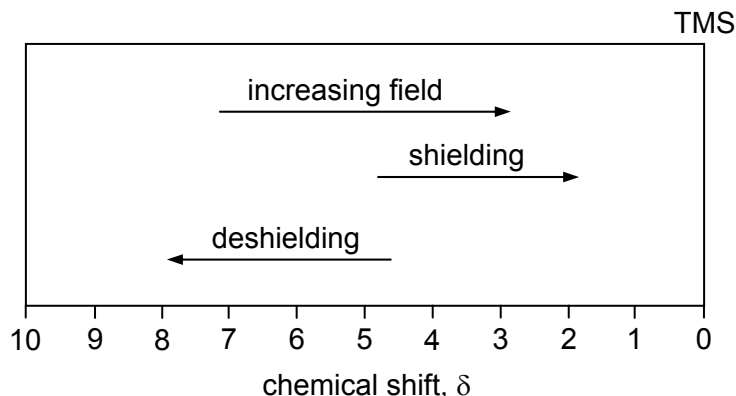


Figure 2.10 – chemical shift

The low resolution NMR spectrum of ethanol contains three different absorption peaks occurring at slightly different values of the applied field (Figure 2.11). The areas under these peaks are found to be in the ratio 1:2:3. These different peaks are explained by assuming that the single -OH proton, the two -CH₂- protons and the three -CH₃ protons all resonate at slightly different values of applied field: the nearer a proton is to the electronegative oxygen atom, the less shielded it is from the external magnetic field. This is because the oxygen atom attracts the bonding electrons, and their diamagnetic effect on the resonating proton is reduced.

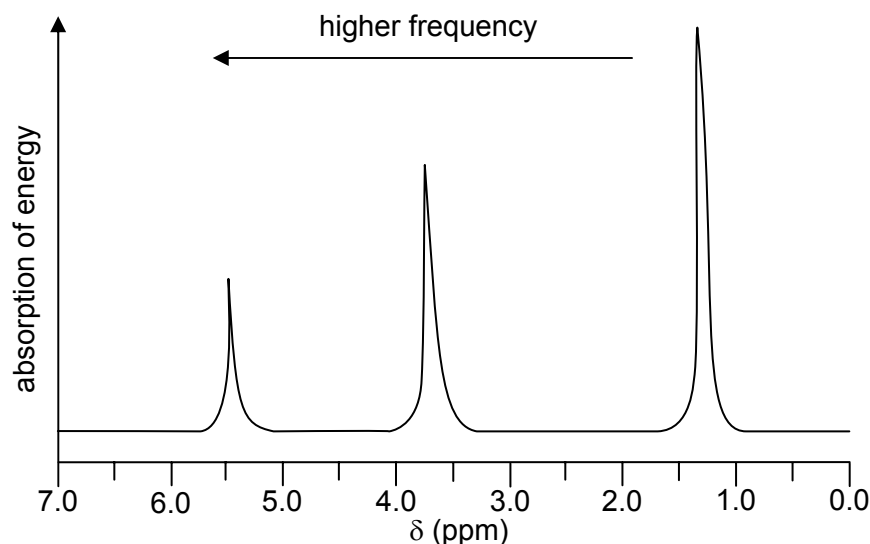


Figure 2.11 – low resolution NMR spectrum of ethanol

A high-resolution spectrum of ethanol (see Figure 2.12) shows the same three peaks, but this time the peak at δ 1.2 is split into three peaks, and that at δ 3.7 is split into four. This is due to a phenomenon known as **spin-spin coupling**.

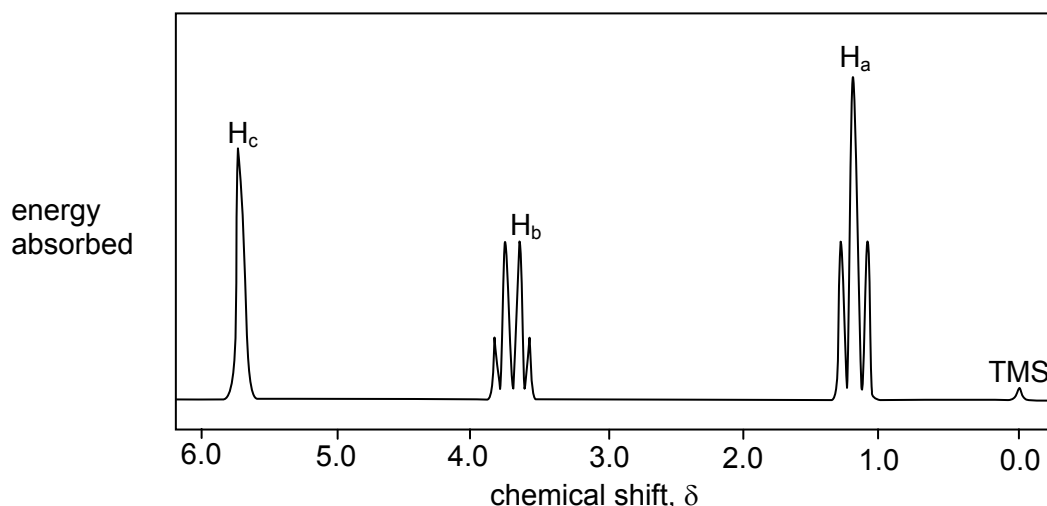


Figure 2.12 – high resolution NMR spectrum of ethanol

The local magnetic field experienced by a resonating proton is not only influenced by the presence of the electrons in the bonds around it, but is also influenced by the orientation (with respect to the external field) of the magnetic moments of nearby protons. If the magnetic moment of a nearby proton is aligned **with** the external field, the local magnetic field experienced by the resonating proton will be increased. The opposite will occur if the magnetic moment of the nearby proton is aligned **against** the external field. The resonating proton will therefore absorb energy at two different frequencies. Because the populations of the two energy levels ($+\frac{1}{2}$ and $-\frac{1}{2}$) are almost equal, the absorbances at the two frequencies of the resonating proton will be equal, and a doublet of lines will be seen. The situation is a little more complicated if there are **two** nearby protons. The magnetic moment of each proton can be either with or against the field, and the possible combinations are shown in Figure 2.13.

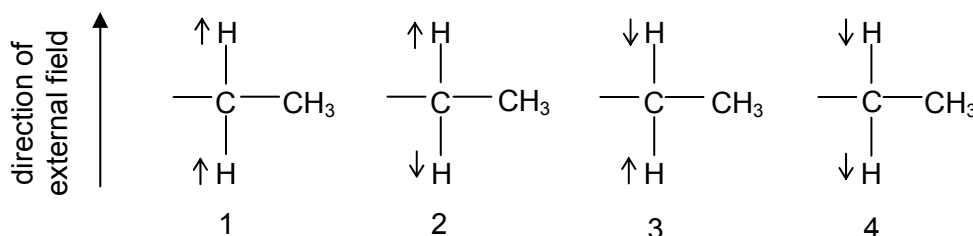


Figure 2.13 – combinations of magnetic moment of two nearby protons

In situations 2 and 3 the magnetic moments of the two $-\text{CH}_2-$ proton cancel each other out, so the field experienced by the $-\text{CH}_3$ protons will not change. In situations 1 and 4, however, the magnetic moments of the $-\text{CH}_2-$ protons reinforce each other, and so the field experienced by the $-\text{CH}_3$ protons will be higher or lower. We would therefore expect the $-\text{CH}_3$ protons to absorb at three different frequencies. Because the probabilities of the four states 1-4 are all equal, there is twice the chance of the $-\text{CH}_3$ protons experiencing **no** change in field (states 2 and 3) as there is for the protons to experience either an enhanced or reduced field (states 1 or 4). We therefore expect the absorbance at the central (unchanged field) frequency to be twice as great as the other two, forming a 1:2:1 triplet of lines. This is exactly what is observed for the absorption centred on δ 1.2.

We can work out the splitting pattern of the peak at δ 3.7 in a similar way: the magnetic moments of the three protons in the $-\text{CH}_3$ group can be oriented in the following combinations:

	$\uparrow\uparrow\uparrow$	$\uparrow\uparrow\downarrow$ $\uparrow\downarrow\uparrow$ $\downarrow\uparrow\uparrow$	$\uparrow\downarrow\downarrow$ $\downarrow\uparrow\downarrow$ $\downarrow\downarrow\uparrow$	$\downarrow\downarrow\downarrow$
net moment:	+3	+1	-1	-3
number of combinations:	1	3	3	1

We can now see why the peak at δ 3.7 is split into a 1:3:3:1 quartet.

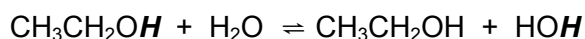
The general rules governing splitting patterns are as follows:

1. Protons in identical chemical environments (e.g. the three H atoms in the $-\text{CH}_3$ group) do not split their own absorption peak.
2. The absorption peak of resonating protons is only split by nearby protons on the **adjacent** carbon atom – protons further away (usually) have little effect
3. The absorption of protons adjacent to n protons is split into $(n + 1)$ peaks.
4. The intensities of the peaks in a multiplet are as in the following table.

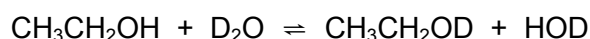
number of protons adjacent to the resonating proton	number of lines in multiplet	relative intensities of lines
1	2	1:1
2	3	1:2:1
3	4	1:3:3:1
4	5	1:4:6:4:1

- SAQ 5.** (a) In the molecule $\text{CH}^{\text{A}}_3\text{-CH}^{\text{B}}_2\text{-CH}^{\text{C}}\text{Cl-CH}^{\text{D}}\text{Cl}_2$, which protons would be split by the H^{B} protons?
- (b) Which would be split by the H^{C} proton?
- (c) Predict the splitting pattern you would observe for (i) the H^{C} proton, and (ii) the H^{B} protons.

You may have noticed that the peak in Figure 2.12 due to the $-\text{OH}$ proton is a singlet, whereas you might have expected it to be split by the adjacent $-\text{CH}_2-$ protons into a triplet. Likewise, you might have expected the $-\text{CH}_2-$ protons to appear as a quintet – split by the three $-\text{CH}_3$ protons and also by the $-\text{OH}$ proton. This splitting does not occur because protons on $-\text{OH}$ groups undergo rapid exchange with each other, and with protons on other $-\text{OH}$ groups such as those in water. This means that adjacent protons only experience an averaged-out field of all the exchanging protons. (The $-\text{OH}$ absorption in **ultra-dry** ethanol does in fact appear as a triplet).



A useful application of this ready exchange of $-\text{OH}$ protons is the disappearance of their absorption peak when an NMR sample is shaken with D_2O (D is deuterium, ^2H). The deuterium nucleus has a nuclear spin of ± 1 rather than $\pm \frac{1}{2}$, so does not absorb in the same frequency range as ^1H .



One further point can be seen from Figure 2.11. The area under each peak is proportional to the number of protons responsible for that absorption. Thus from the left hand side, the areas of the peaks are in the ratio 1:2:3. This is also true for the high-resolution spectrum in Figure 2.12, but in this case the sum of the areas underneath *all* the split peaks in a group is proportional to the number of protons. This information is usually presented on the spectrum either as an *integration trace* (see the dotted line on Figure 2.12) or as the indication $n\text{H}$ printed near the peak, where n is the number of protons (see Figure 2.12).

Some examples of ^1H NMR spectra

To illustrate the application of these principles let us look at the NMR spectrum of two unknown compounds X and Y, and see if we can use the spectra to identify X and Y. We shall refer to the *Data Booklet* for δ values. Try to identify them by yourself, and then look at the analysis of each spectrum given below.

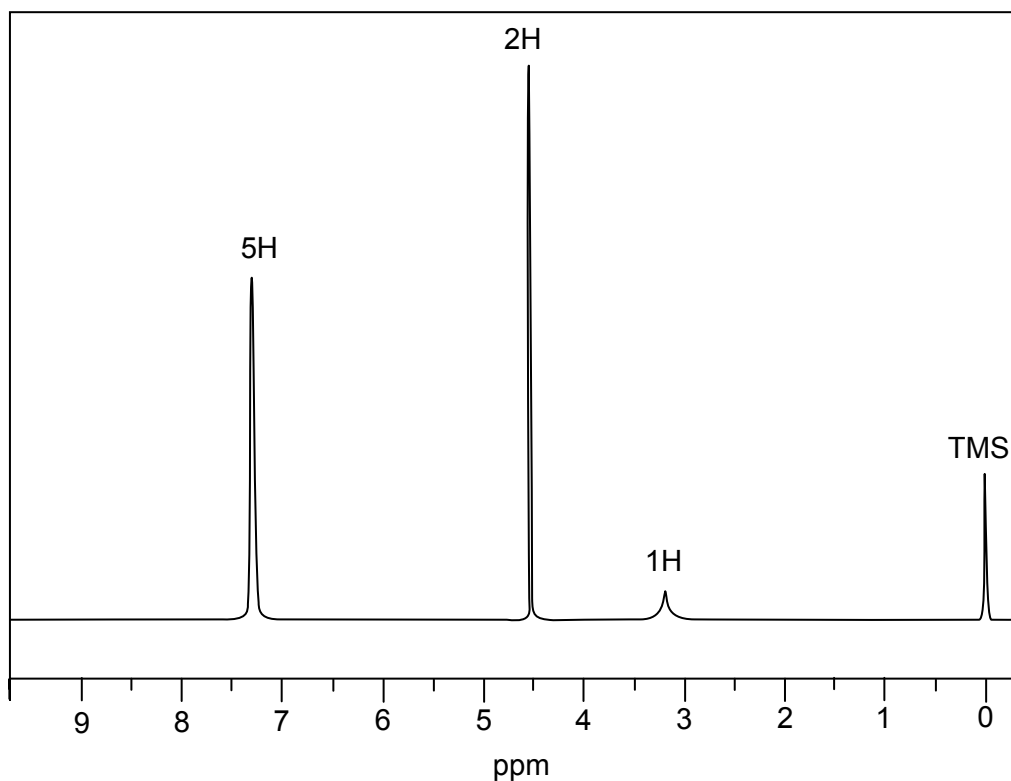
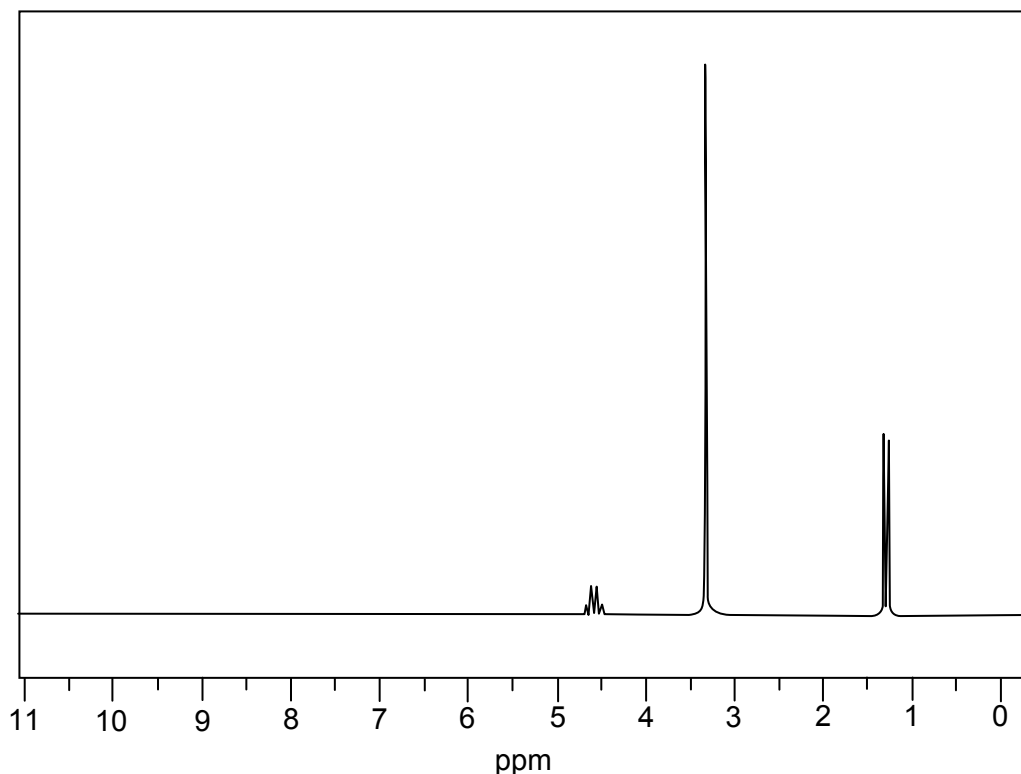


Figure 2.14 – ^1H NMR spectrum of X

Further information on X:

$M_r = 108$

the ^1H peak disappears on shaking with D_2O

Figure 2.15 – ^1H NMR spectrum of Y

Further information on Y: $M_r = 90$
no change in the spectrum on shaking with D_2O

Analysis of the data

Compound X

- None of the three peaks in the spectrum (don't forget to ignore the TMS peak, no matter how high it is!) is split into a multiplet, so compound X does not contain any non-equivalent protons on adjacent carbon atoms.
- The 1H peak at δ 4.5 (identical to *Data Booklet* value), which disappears in shaking with D_2O , is likely to be due to an $-\text{OH}$ (alcohol) group.
- The 5H peak at δ 7.3 is in the aromatic C-H region of the spectrum (identical to *Data Booklet*). Its integration of 5 protons suggests an aryl ring with just one group attached to it.
- The 2H peak at δ 4.3 is consistent with a CH_2 group joined to an oxygen atom (*Data Booklet* for $\text{O}-\text{CH}_3$ is 3.8) but brought downfield by its proximity to an aromatic ring.

Compound Y

- The 3H doublet at δ 1.2 is in the region of the spectrum we expect to see $\text{C}-\text{CH}_3$ protons. (The slightly higher δ value compared to the *Data Booklet* value of 0.9 could be due to the proximity of an aromatic ring or an electronegative atom such as oxygen). The fact that this peak is split into a doublet suggests it is adjacent to a carbon atom with one hydrogen on it.
- The fact that there are six protons responsible for the singlet at δ 3.3 suggests two equivalent CH_3 groups. The δ value is near that of an $-\text{O}-\text{CH}_3$ group (*Data Booklet* 3.8).
- The 1H peak at δ 4.5 is split into a quartet due to its proximity to the $\text{C}-\text{CH}_3$ protons mentioned above (if one peak in a spectrum is split into a multiplet, another split peak is bound to be seen also). Although its δ value is the same as that expected for an $-\text{OH}$ proton, this cannot be the case because D_2O has no effect on the spectrum. It is, however, consistent with a C-H proton adjacent to two electronegative oxygen atoms (*Data Booklet*: the change from $\text{R}-\text{CH}_3$ (δ 0.9) to $\text{R}_3\text{C}-\text{H}$ (δ 2.0) increases the δ value; $-\text{O}-\text{CH}_3$ is at δ 3.8).

Answers: Compound X is phenylmethanol, $\text{C}_6\text{H}_5\text{-CH}_2\text{-OH}$
 Compound Y is 1,1-dimethoxyethane, $(\text{CH}_3\text{O})_2\text{CH-CH}_3$

Magnetic Resonance Imaging (MRI)

The application of NMR in medicine is becoming increasingly commonplace. The technique most commonly used is magnetic resonance imaging. A patient is placed on a steady table and the part of the body to be scanned (often the head) is immobilised. (Occasionally the patient is anaesthetised.) The table is slowly passed through the magnetic field, where a fine beam of UHF radiation is passed through and analysed. This shows the ^1H absorption pattern of a thin section (about 1 cm thick) of the body. The table continues to move gently into the magnetic field, and after each 1 cm another absorption pattern is recorded. A typical brain scan containing 20-30 slices takes only 10 minutes to complete.

Most of the hydrogen atoms within cells are in water molecules or lipid molecules, and it is the environment of the water molecules that gives an indication of the medical state of the cell. MRI is useful in pinpointing brain tumours and sites of injury, and diagnosing hydrocephalus, multiple sclerosis, Alzheimer's and other brain diseases.

X-ray crystallography

X-rays are very short wavelength electromagnetic rays. Their wavelength of about 0.1 nm is comparable to the interatomic distances in solids. If a beam of monochromatic X-rays (i.e. rays of a single wavelength) is passed through a crystal, some X-rays will be reflected from the layers of atoms in the crystal lattice. (In fact, the X-rays are *diffracted* by the planes of atoms within a crystal, but geometrically this is identical to the reflection described here, and the concept of reflection is easier to understand.)

If the angle at which the incoming beam hits the layers is just right, the peaks in the waves of the reflected rays from the layers below will coincide with the peaks in the waves of the ray reflected from the first layer of atoms, and a **reinforcement** of intensity will result. For this to occur the difference in the path length of the two rays (see Figure 2.16) must be a whole number of wavelengths. The path length difference depends on the distance between the layers of atoms, d , the wavelength of the X-rays, λ , the angle of incidence of the rays to the atomic layer, θ , according to the equation

$$n\lambda = 2d\sin\theta \quad (\text{where } n = \text{a whole number, } 1, 2, 3 \text{ etc})$$

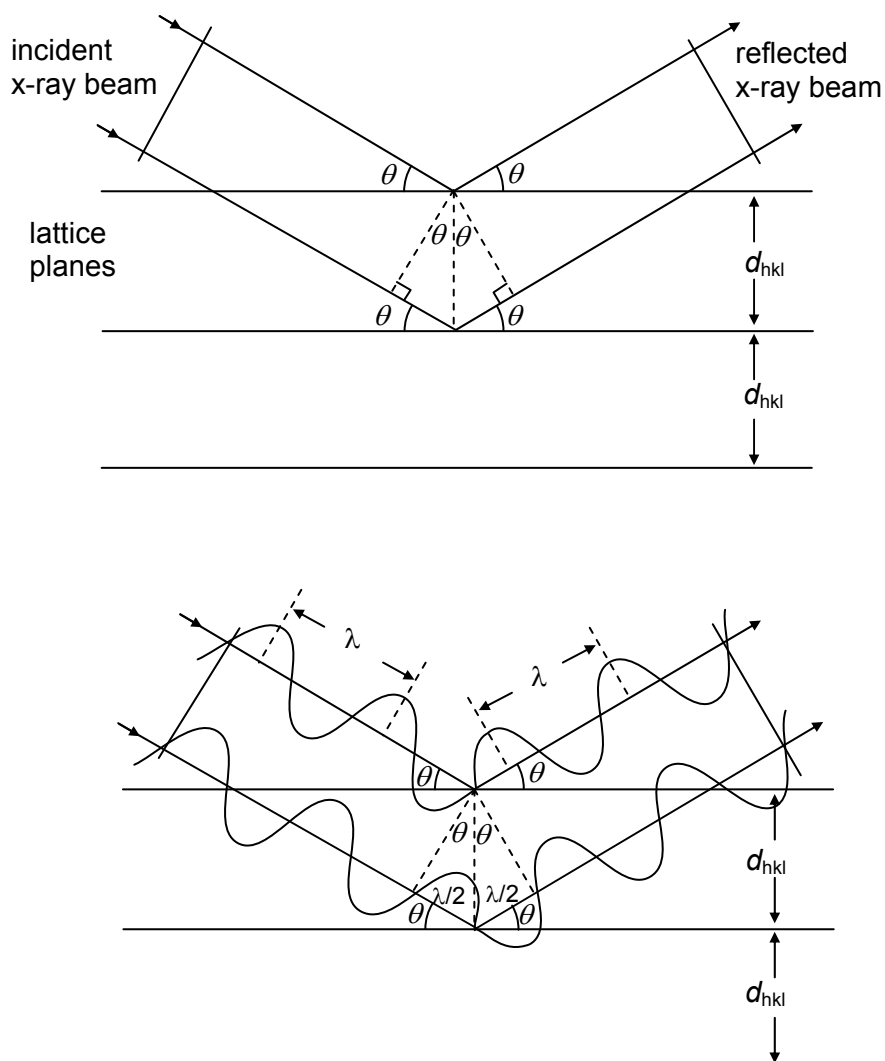


Figure 2.16 – X-rays reflecting from layers of atoms

To understand how X-ray crystallography can help us understand the chemical structures of compounds we first need to look at how crystals are built up. Crystals are characterised by containing a **lattice** of ions or molecules. This lattice is a regular 3-dimensional repeating pattern of particles. The smallest 3-dimensional repeating unit within a crystal is called the **unit cell**. The orientation of atoms within one unit cell is identical to that within all the other unit cells in the crystal. If the planes of atoms responsible for the reinforcement of intensity of X-rays indicated in Figure 2.16 are parallel to one face of the unit cell, the distance, d , is equal to the length of the side that is perpendicular to that face (for simplicity, we shall assume the unit cell has rectangular faces). This is shown in Figure 2.17.

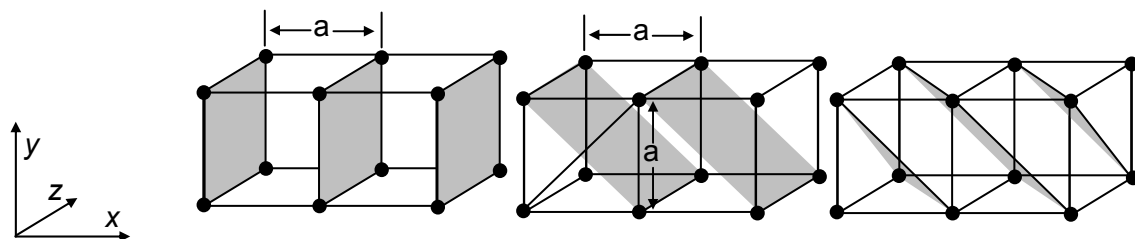


Figure 2.17 – diagram of unit cell within an array of atoms, showing d

An X-ray spectrometer consists of three main parts:

- an X-ray source
- a stable support on which the crystal is mounted, and which can be rotated by a measured angle around all three axes
- a system for detecting the reflected X-rays. This can be either a photographic film or a scintillation counter.

The X-rays are produced when a high-energy beam of electrons (accelerated by a potential of 60 000 V) strike a metal target. This produces X-rays of many different wavelengths, but on passing through filters made of thin metal foils, a beam of single-wavelength (monochromatic) X-rays is produced. The diffractometer itself, consisting of the rotatable support and the counter/detector, is shown diagrammatically in Figure 2.18.

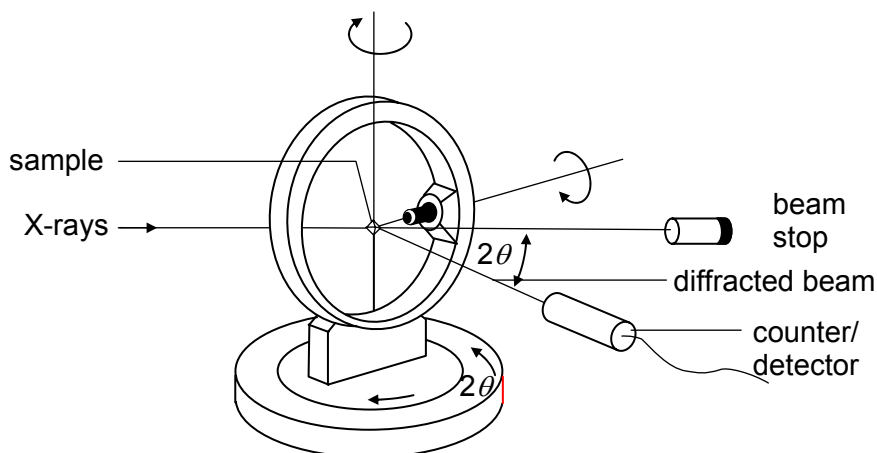


Figure 2.18 – X-ray diffractometer

The X-rays are diffracted by the electron clouds in the atoms of the crystal. The electron clouds interact with the electric field of the X-rays. All atoms except hydrogen contain enough electrons to diffract the X-rays, and the heavier the atom (i.e. the greater the number of electrons around it), the more intense is the reinforcement of the beam.

Thus, by measuring the angles between the incident beam and the diffracted beams, and the relative intensities of those beams, we can piece together a picture of the electron density at all points in the unit cell. This process requires an enormous number of calculations to be done, requiring many hours of high speed computer time. It also relies on some good luck and informed judgement of the part of the X-ray crystallographer. The technique is often to guess at a structure, and get the computer to predict the diffraction pattern that would be formed by that structure. This prediction is then compared to the observed pattern. Repeated modifications of the suggested structure eventually produce a diffraction pattern that is identical to the one observed. With molecules containing many hundreds of atoms (e.g. a protein), the process can be quite time consuming. Often, though, the limiting factor is the production of a sufficiently pure crystal in the first place.

The structures of simpler molecules such as ethanoylaminoethanoic acid (see Figure 2.19) can be worked out much more quickly. The stages involved were as follows:

1. The product was synthesised and purified by recrystallisation from water.
2. Batches of crystals were grown from aqueous solution.
3. An X-ray diffraction pattern was collected. The unit cell size was estimated. The spots were indexed and the lengths of the three sides of the unit cell were measured.
4. The density of the substance was measured which enabled the number of molecules per unit cell to be calculated.
5. The reflections from one set of planes were found to be exceptionally intense, which suggested that all the atoms in the molecule were concentrated in those planes.
6. The perpendicular distance between the planes was calculated to be $3.24 \times 10^{-10}\text{m}$. This small distance suggested that the molecules within the layers had a planar arrangement.

7. Trial structures were suggested based on the known chemistry of ethanoylaminoethanoic acid.
8. The electron density map was interpreted and related to the arrangement of atoms within the molecules.

The result is shown in Figure 2.19

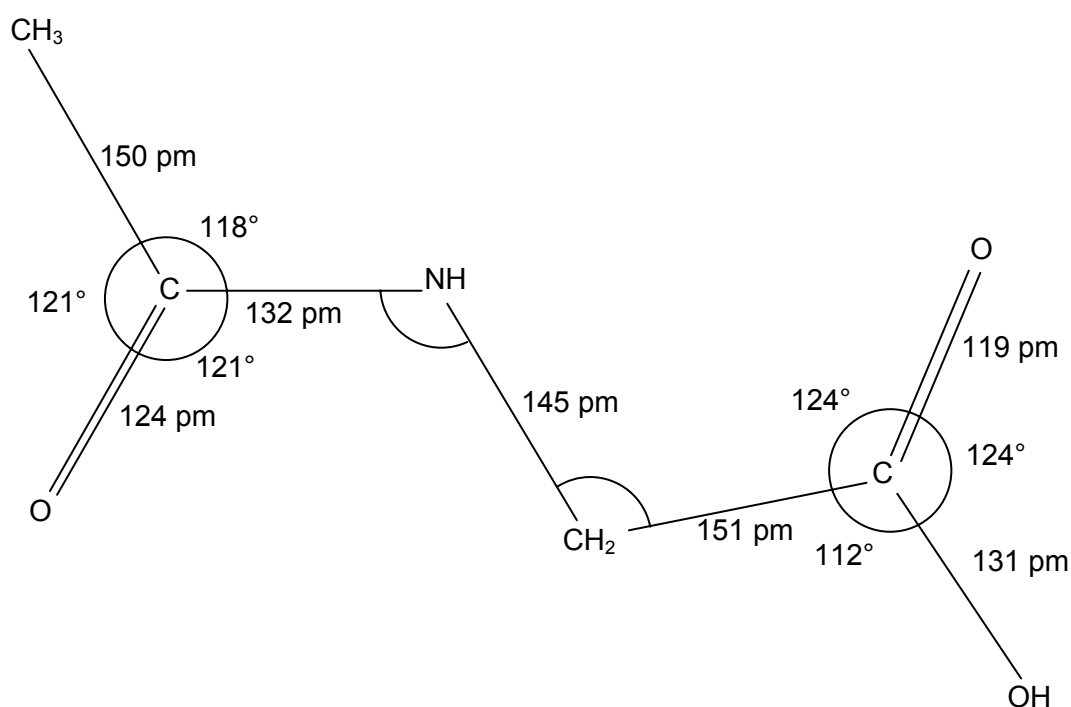
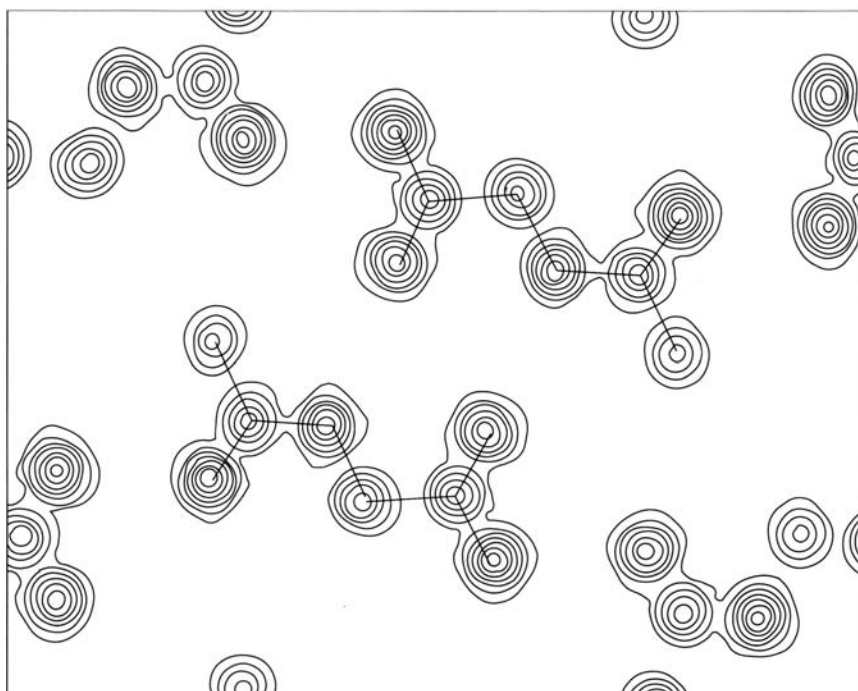


Figure 2.19 – electron density map of N-ethanoylaminoethanoic acid

SAQ 6. Work out which atoms are at the centres of the electron density contours in Figure 2.19. Using a protractor, estimate the C-N-C and the N-C-O bond angles.

The role of X-ray crystallography in the discovery of the structure of DNA by Watson and Crick in 1953 is well known. The DNA molecule has a good deal of symmetry, with the planar base-pairs taking up positions parallel to each other. The relatively simple X-ray diffraction photograph of a hydrated DNA fibre (Figure 2.20) shows the central spots arranged in a cross pattern – typical of a helical structure. Key dimensions obtained by X-ray crystallography are shown in Figure 2.21.

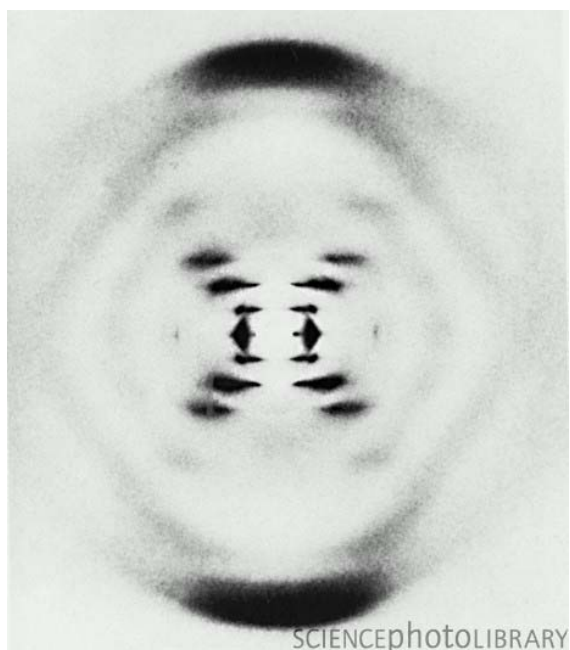


Figure 2.20 – X-ray diffraction photograph of a hydrated DNA fibre

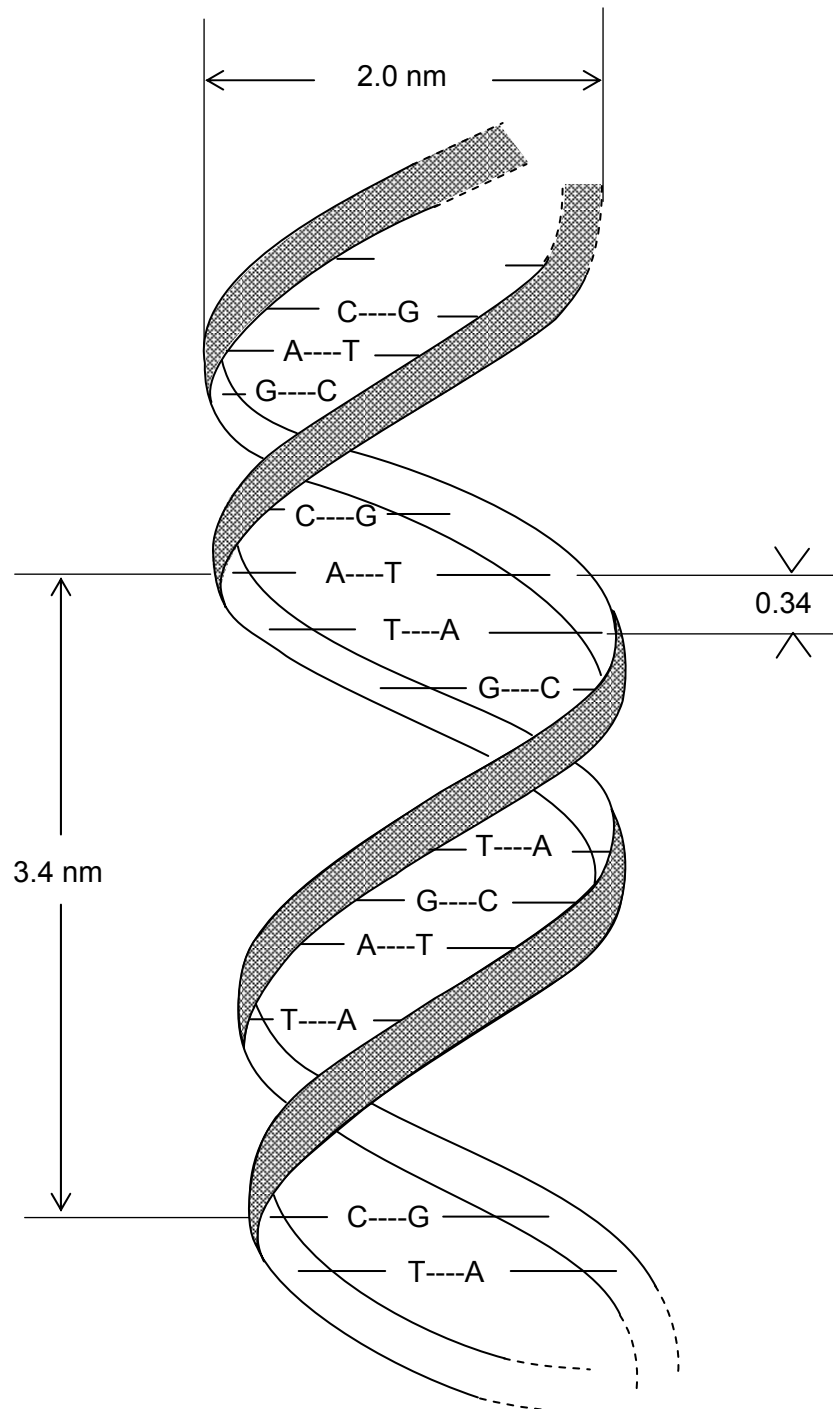


Figure 2.21 – key dimensions obtained for DNA by X-ray crystallography

Case study: The structure and function of myoglobin

Myoglobin is an oxygen transporting agent and oxygen reservoir found in the muscular tissues of vertebrates. It achieves its function by the use of a haem group. This consists of a Fe^{2+} ion surrounded by a protoporphyrin ring. The iron is bound to the polypeptide chain by an Fe-N covalent bond to a histidine residue. The oxygen molecule is transported by forming a dative bond to the Fe^{2+} ion (Figure 2.22).

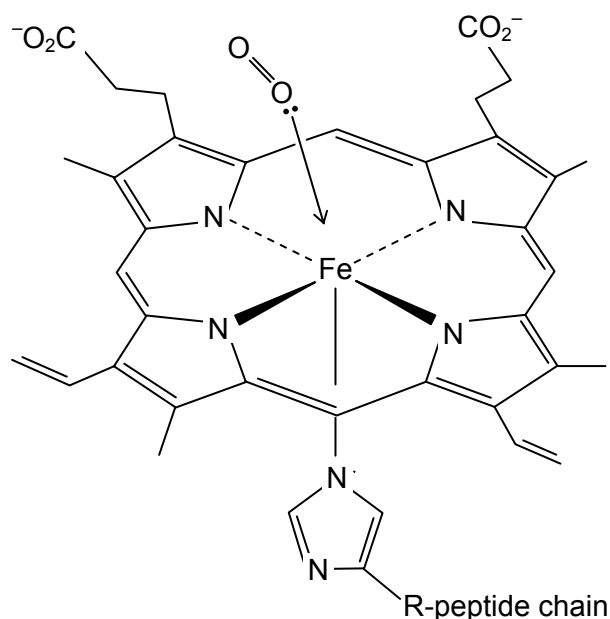


Figure 2.22 – iron protoporphyrin with O_2 and histidine

One problem that has fascinated chemists for some time is why, in the presence of oxygen, the Fe^{2+} ion does **not** become oxidised to Fe^{3+} . If iron(II) protoporphyrin is dissolved in water, and oxygen is bubbled through the solution, an almost immediate oxidation to Fe^{3+} occurs. A key intermediate in this oxidation is a complex of an oxygen molecule sandwiched between two haem groups (Figure 2.23).

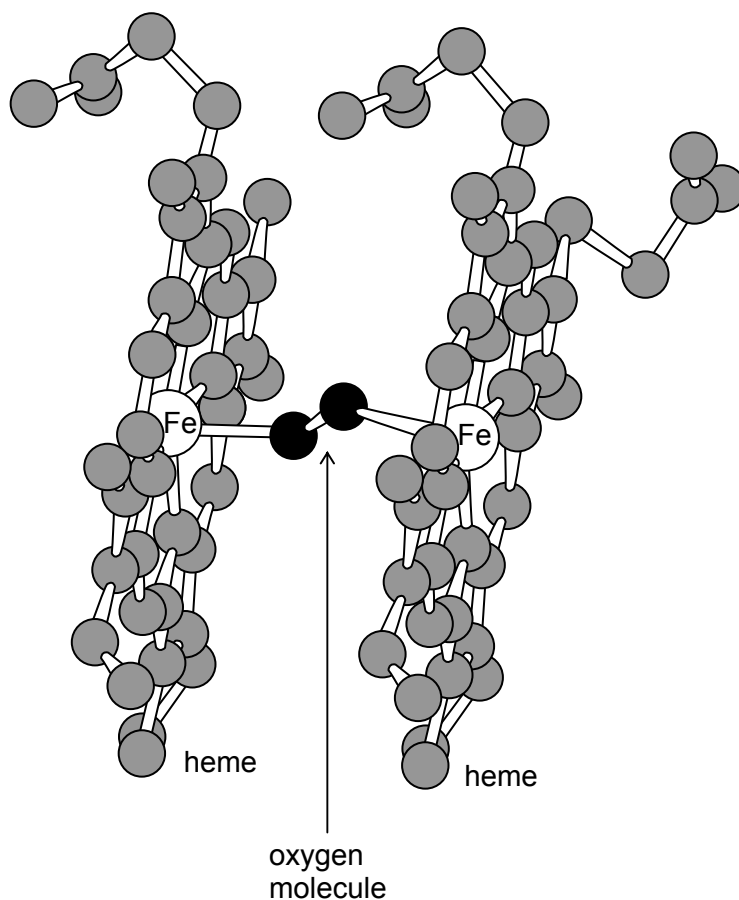


Figure 2.23 – two iron protoporphyrins with O_2 sandwiched in between

Like many iron complexes, haem has a large affinity for carbon monoxide. This displacement of O_2 by CO is why carbon monoxide is so poisonous. But another curious property of the haem groups in both myoglobin and its blood counterpart, haemoglobin, is that their affinities for carbon monoxide, although large, are very much smaller than that of a isolated haem group in solution.

Both these observations were readily explainable once the 3-dimensional structure of myoglobin had been determined.

This was achieved by John Kendrew and his colleagues between 1957 and 1962. It was the first protein to have its structure determined by X-ray crystallography. Kendrew chose myoglobin for two important reasons:

- it is a comparatively small protein ($M_r \approx 18,000$) which has the advantage of containing a heavier (iron) atom. Since this is surrounded by a large number of electrons (26) it should diffract X-rays well.
- it readily forms good quality crystals, and is available in quantity.

After several years of intense work and many hundreds of hours of calculation, involving the measurement and analysis of the intensities of 25 000 spots in the X-ray photograph (see Figure 2.24), Kendrew succeeded in working out its structure (Figure 2.25). As can be seen, the myoglobin molecule is extremely compact, with a large percentage of its amino acids joined in the α -helical configuration. This produced several stable, stiff “rods”, which are joined by small lengths of more flexible parts of the amino acid chain.

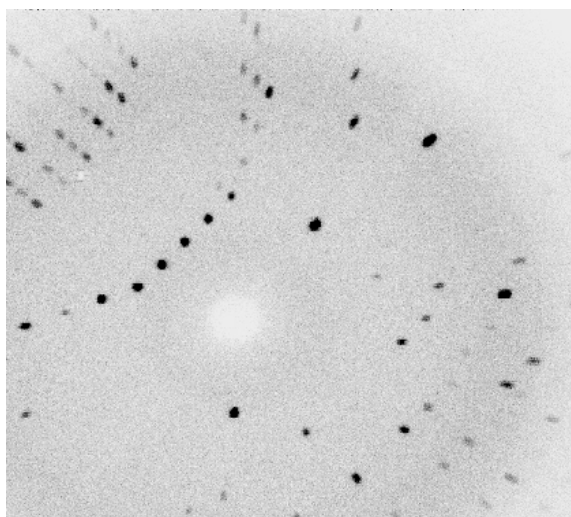


Figure 2.24 – X-ray diffraction photograph of myoglobin

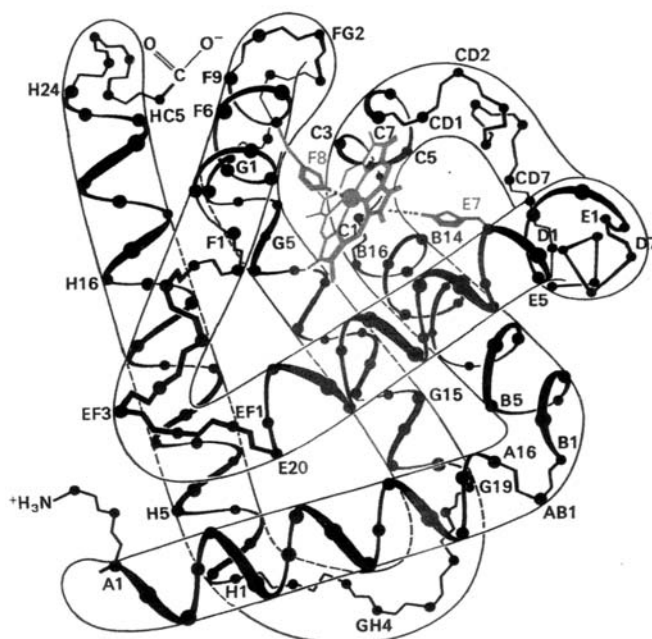


Figure 2.25 – structure of myoglobin molecule

In particular, the oxygen-binding site is very hindered by the amino acids surrounding it, and the haem group is on the inside of the molecule. Furthermore, there is a second histidine group positions just over (but slight to one side of) the 6th coordination position of the Fe^{2+} iron, where the O_2 molecule sits.

This structure now explains the two observations described above. The fact that the haem group is on the inside of the molecule makes it impossible for two such groups to come together with an oxygen molecule bridge between them, as in Figure 2.23. In addition, the second histidine group, above the Fe^{2+} ion, forces the complexing molecule to bond with the Fe^{2+} ion at an angle. This favours O_2 , but hinders the bonding of CO (see Figure 2.26).

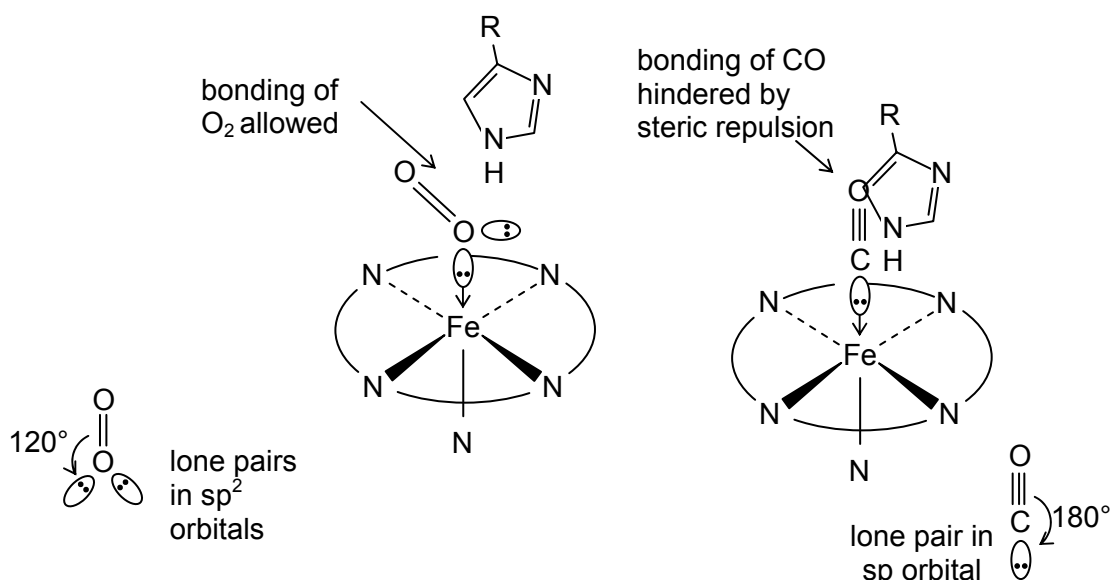


Figure 2.26 – arrangement of molecules within the myoglobin complex

Thus the specific nature of the function of myoglobin can be readily understood once its structure is known.

Summary

- Mass spectrometry, nuclear magnetic spectroscopy and X-ray crystallography are three methods of determining the structures of molecules.
- Mass spectrometry can tell us the number of carbon atoms in a molecule (from the $M : M+1$ ratio); the molecular formula (from an accurate determination of the M_r of the molecular ion) and something about its structure (from the fragmentation pattern).
- Nuclear magnetic spectroscopy tells us the number of hydrogen atoms in each chemical environment in a molecule (from the integration trace and the chemical shift (δ) values), and the number of their nearest neighbours (from the splitting patterns). If a particular type of proton has n nearest neighbours, its peak is split into $(n+1)$ lines.
- X-ray crystallography tells us how the atoms are arranged in a crystal of the compound. It can determine the positions of all atoms except hydrogen.

2.3 – Separating and identifying substances

By the end of this section, you should be able to:

- understand what is meant by the term partition coefficient
- calculate a value for the partition coefficient, and use it in calculations of concentrations of solute in various solvents
- appreciate the use of solvent extraction in the extraction of plant products and the analysis of drug and pesticide residues
- outline the principles of paper chromatography (PC), thin layer chromatography (TLC), gas-liquid chromatography (GLC) and high performance liquid chromatography (HPLC)
- understand the meaning of the terms retention ratio (R_f) and retention time.
- outline the principles of electrophoresis
- appreciate the uses of the various forms of chromatography and of electrophoresis in the separation and analysis of proteins, drugs, pesticides, explosives, stimulants and other food additives.

The partition of a solute between two immiscible solvents

We are all familiar with the most common liquid on Earth, water. Water is an excellent solvent for many substances. Some, like salt, sugar, and ammonia, dissolve in water well; whereas others, like oxygen or iodine, dissolve only partially; still others, like silica, benzene or helium, are virtually insoluble in water.

If we take another solvent, such as hexane, we find a different pattern of solubilities. Silica and helium are still insoluble, but so also are salt and sugar. Iodine and benzene dissolve well, but ammonia and oxygen are only slightly soluble.

It is worth digressing for a moment to look at the reasons why some substances dissolve and others do not. In general we can say that a substance dissolves because the total energy of the solute+solvent system is less when the solute is in solution than when it is not. This tends to be the case if the interactions between the solute molecules and those of the solvent are **similar** to those between the solvent molecules themselves.

In general, there are four different ways that molecules attract each other. These are:

- ion-ion attractions
- ion-dipole attractions
- van der Waals' (induced dipole) attractions
- hydrogen bonding.

Although there are van der Waals' attractions between water molecules, these are small because of the small number of electrons in the H_2O molecule. The main force of attraction is due to hydrogen bonding. This explains why sugar ($\text{C}_{12}\text{H}_{22}\text{O}_{11}$) dissolves well – the 11 oxygen atoms in the sucrose molecule are all hydrogen bonded to water molecules, as are the 8 hydrogen atoms on the OH groups. The H_2O molecule also possesses a dipole, which is attracted to cations and anions by ion-dipole interactions. This explains why salt (NaCl) dissolves in water. Ammonia dissolves well due to extensive hydrogen bonding between NH_3 and H_2O molecules.

The main interactions between adjacent hexane molecules are van der Waals' (induced dipole) forces. Molecules such as benzene and iodine, which also have induced dipole forces between them, dissolve well in hexane. But benzene and iodine are insoluble in water because the energy gain by forming van der Waals' attractions between their molecules and H_2O molecules does not exceed the energy penalty of breaking the hydrogen bonds between two adjacent H_2O molecules in the solvent. Likewise, sugar and salt are insoluble in hexane because there is no possibility of hexane either hydrogen bonding with sugar molecules, or undergoing ion-dipole interactions with the Na^+ and Cl^- ions in salt.

Solvents can, in general, be listed according to their **polarity**. Polar solvents are more likely to be able to dissolve ions, hydrogen bonded molecules and molecules with dipoles; non-polar solvent are

more likely to dissolve those solutes whose molecules are only attracted to each other by van der Waals' (induced dipole) attractions.

SAQ 7. (a) Methanol, CH_3OH , is very soluble in water, whereas heptanol, $\text{C}_7\text{H}_{15}\text{OH}$, is almost insoluble in water. Why is this?

(b) Predict, with a reason, which alcohol would be more soluble in hexane.

If a solute such as iodine is soluble in two immiscible solvents (e.g. water and hexane), its solubility (usually measured in g cm^{-3} or mol dm^{-3}) is very unlikely to be the same in both solvents. One solvent will be better at dissolving it than the other.

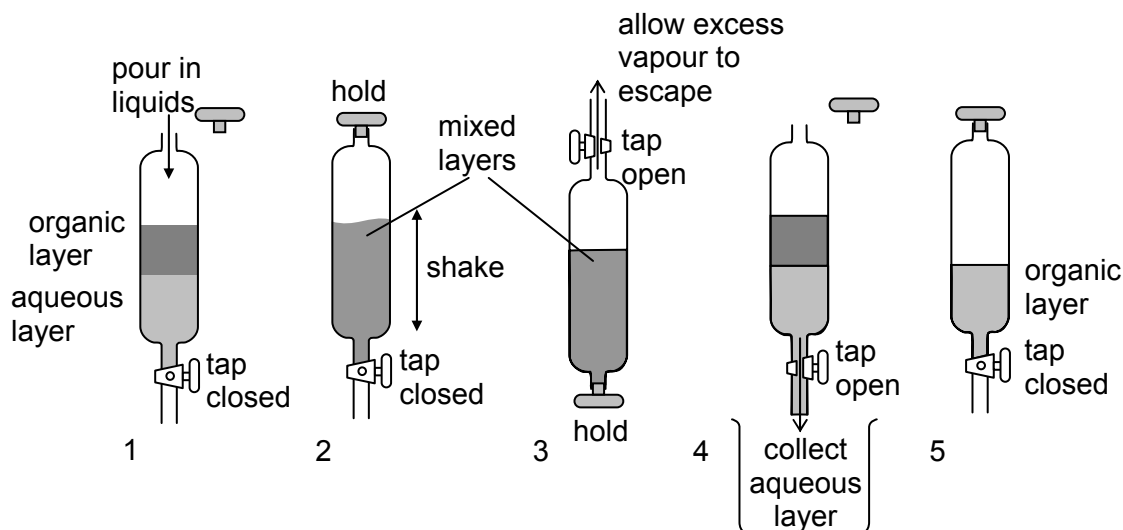
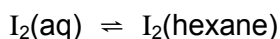


Figure 2.27 – solvent extraction

When some iodine crystals are shaken with a mixture of hexane and water until no further change takes place, and the two layers allowed to separate (see Figure 2.27), we find that the ratio of the concentrations of iodine in each layer is a constant, no matter how much iodine we start with. This constant is the equilibrium constant for the change:



$$K_c = [\text{I}_2(\text{hexane})]/[\text{I}_2(\text{aq})]$$

K_c is called the **partition coefficient** of iodine between hexane and water. The usual symbol for a partition coefficient is K_{pc} . Like all equilibrium constants, the value of a partition coefficient changes with temperature. Notice that, in general partition coefficients have no units.

Partition coefficients are useful in working out how much solvent we need in order to extract a minimum amount of solute from one solvent into another. An example will make this clear.

Example 1:

Question:

When 100 cm^3 of an aqueous solution containing 2.0 g of the organic dye X was shaken with 20 cm^3 of hexane, it was found that 1.6 g of the dye had been extracted into the hexane.

- Calculate the partition coefficient of X between hexane and water.
- Calculate the minimum volume of hexane needed to reduce the amount of X in the aqueous layer to less than 0.1 g .

Answer:

(i) If 1.6 g of X had been extracted by the hexane, 0.4 g must have been left in the aqueous layer. The two concentrations are therefore:

in the hexane layer: $1.6/20 = 0.08 \text{ g cm}^{-3}$ in the water layer: $0.4/100 = 0.004 \text{ g cm}^{-3}$

$$K_{pc} = [\text{X}(\text{hexane})]/[\text{X}(\text{aq})] = 0.08/0.004 = 20$$

(ii) If the mass left in the aqueous layer is to be less than 0.1 g, then the mass extracted must be at least 1.9 g. The corresponding concentrations are as follows:

in the hexane layer: $1.9/v \text{ g cm}^{-3}$ in the water layer: $0.1/100 = 0.001 \text{ g cm}^{-3}$

(where v = volume of hexane needed)

$$K_{pc} = [X(\text{hexane})]/[X(\text{aq})]$$

$$\therefore 20 = (1.9/v)/0.001$$

$$v = 1.9/(20 \times 0.001) = 95 \text{ cm}^3$$

Successive extractions

Solvents are often expensive and flammable, and can also be polluting to the environment. It is sensible, therefore, to use the minimum amount of solvent that is needed to achieve the intended goal. Thus in the example above, the use of 20 cm^3 of hexane allowed 1.6g of the dye X to be extracted from its aqueous solution, i.e. $(1.6/2.0) \times 100 = 80\%$.

We can extract more than this if we use two separate 10 cm^3 portions of hexane, rather than one 20 cm^3 portion. The following calculation will make this clear.

Example 2:

1st extraction: Let us assume that x grams of X have been extracted by the first 10 cm^3 . The equilibrium concentrations will therefore be:

in the hexane layer: $x/10 \text{ g cm}^{-3}$ in the water layer: $(2.0 - x)/100 \text{ g cm}^{-3}$

$$K_{sp} = \frac{x/10}{(2.0 - x)/100} = 20$$

$$\therefore x/10 = (40 - 20x)/100$$

$$\therefore 100x = 10(40 - 20x)$$

$$\therefore 100x = 400 - 200x$$

$$\therefore x = 1.33 \text{ g}$$

Thus the amount extracted = 1.33g, so the amount remaining in the aqueous layer = 0.67g

2nd extraction: We now separate the 10 cm^3 of hexane solution of X from the aqueous layer, and add another 10 cm^3 of hexane and shake again. Let us assume that the second 10 cm^3 will extract y grams of X, the equilibrium concentrations will be:

in the hexane layer: $y/10 \text{ g cm}^{-3}$ in the water layer: $(0.67 - y)/100 \text{ g cm}^{-3}$

$$K_{sp} = \frac{y/10}{(0.67 - y)/100} = 20$$

By a calculation similar to the one shown above, we find that

$$y = 0.44 \text{ g}$$

The total amount of X extracted by two successive 10 cm^3 portions of hexane is therefore $1.33 + 0.44 = 1.77\text{g}$. This represents **89%** of the original 2.0g amount of X in the 100 cm^3 of water.

Similar calculations show that if we had split the 20 cm^3 of hexane into four 5 cm^3 portions, the amounts of X extracted at each stage would have been as follows:

1st extraction by 5 cm^3 : 1.00g

2nd extraction by 5 cm^3 : 0.50g

3rd extraction by 5 cm^3 : 0.25g

4th extraction by 5 cm^3 : 0.125g

total extracted = 1.875 g (94%)

All these results are collected together in the following table:

extraction method	percentage extracted
1 x 20 cm ³ of hexane	80%
2 x 10 cm ³ of hexane	89%
4 x 5 cm ³ of hexane	94%

As can be seen, the amount extracted steadily increases as we split the amount of extracting solvent into smaller and smaller portions. However, it is impossible to extract **all** of a solute, no matter how many portions of solvent we use, since it is never possible to move any equilibrium **completely** to one side or the other. But, if the solvent is a volatile one, and if the solute is involatile and stable to heat, it is possible to 'automate' the process by using a continuous extraction apparatus as shown in Figure 2.28.

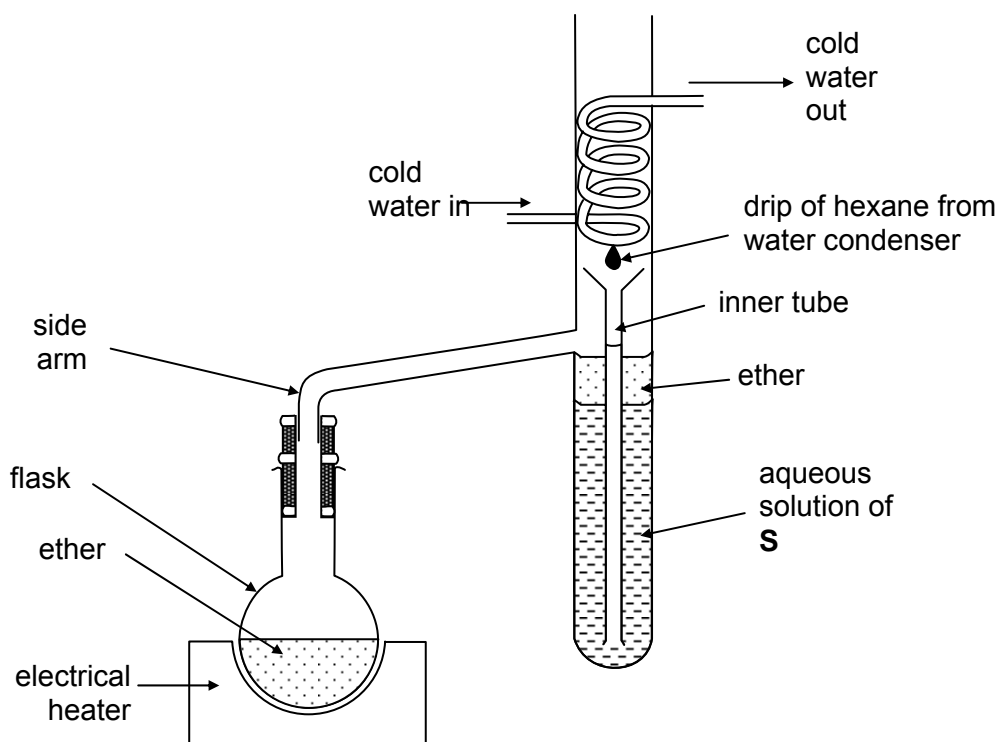


Figure 2.28 – continuous hexane extraction

SAQ 8. Explain in a few sentences how the continuous extraction apparatus shown in Figure 2.28 works.

Solvent extraction is used to extract perfumes and pharmaceutical precursors from plants, and in the analysis of insecticide residues and other pollutants in drinking water supplies, blood and milk. Various applications of solvent extraction are given throughout this chapter.

SAQ 9. A solution of iodine in trichloroethane (TCE) was shaken with water. The iodine content of the two layers was determined by titration with aqueous sodium thiosulphate. A 25 cm³ portion of the aqueous layer required 9.5 cm³ of 0.02 mol dm⁻³ thiosulphate to reach the end point, whereas 5 cm³ portion of the TCE layer required 17.5 cm³ of 0.2 mol dm⁻³ thiosulphate.
Calculate the K_{pc} for iodine between TCE and water.

Chromatography

The name "chromatography" comes from two Greek words meaning "colour picture", and shows the origin of the technique in the separation and analysis of coloured dyes and plant pigments. Nowadays, however, we do not have to rely on colour to detect the compounds we separate using the

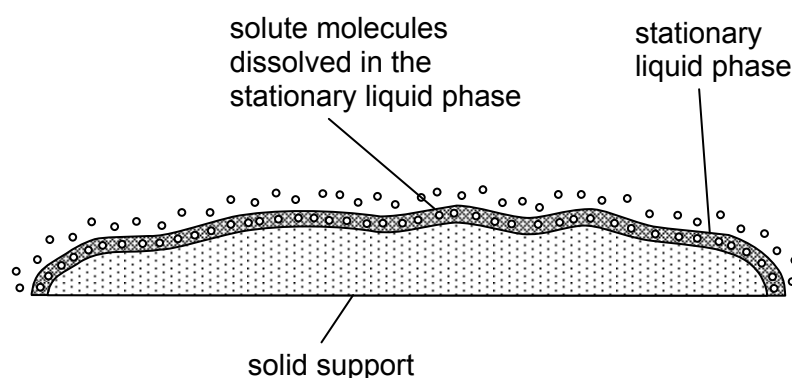
various techniques of chromatography, since there are many other detection methods available. Four techniques will be described here:

- paper chromatography (PC),
- thin layer chromatography (TLC),
- gas/liquid chromatography (GLC) and
- high performance liquid chromatography (HPLC).

The basic principles and techniques

All chromatographic methods use the same principle of a mobile phase (a liquid or a gas) moving past a stationary phase. The stationary phase may be a solid onto which the solutes are adsorbed (TLC, and some GLC and HPLC) or a liquid which is held in a thin film of the surface of an inert solid (PC, GLC and HPLC). The solute molecules partition themselves between the stationary phase and the moving phase. Unlike solvent extraction, however, the partition is not a true equilibrium, since the mobile phase is constantly moving past the area on the stationary phase where a particular solute is adsorbed/dissolved. Thus a fresh partition with pure mobile phase, is continually being set up. The mobile phase should never become saturated with the solute.

Separation by partition



Separation by adsorption

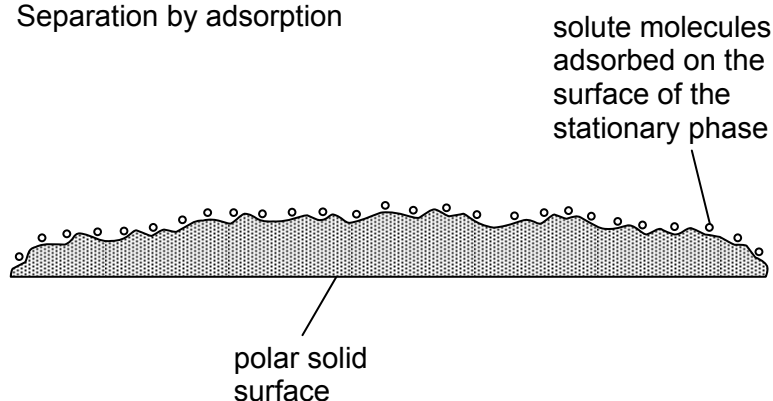


Figure 2.29 – the two types of chromatographic separation – partition and adsorption

Paper chromatography

The mixture to be separated and analysed is dissolved in a solvent, such as water or ethanol, and a small spot of the solution is placed about 1 cm from the edge of a rectangular sheet of chromatography paper (which is like filter paper, but is made to a more exact specification). Other spots of “reference” compounds may also be applied at the same distance from the edge. The solvent is allowed to evaporate. This allows the solutes to become adsorbed into the fibres of the paper (often by means of hydrogen bonding to the OH groups of the cellulose fibres, or more likely to the H₂O molecules that are still associated with cellulose, even when the paper feels dry).

The edge of the sheet is then immersed in the chromatography solvent (the moving phase), being careful at all times to keep the spots above the surface of the solvent. Capillary action draws the liquid up the sheet, and as it passes the point where the spot has been adsorbed, the compounds in the mixture will partition themselves between the cellulose surface and the moving solvent. The more strongly a compound is adsorbed, the less likely it is to be drawn off the cellulose surface. On the other hand, those solutes that are weakly adsorbed, or which have a high solubility in the moving phase, will be drawn up the paper as the solvent front ascends. The solvent is usually less polar than the cellulose surface and its associated water layer, so polar, strongly hydrogen bonded compounds only travel to a small degree, whereas the less polar and less hydrogen bonded compounds travel a larger distance.

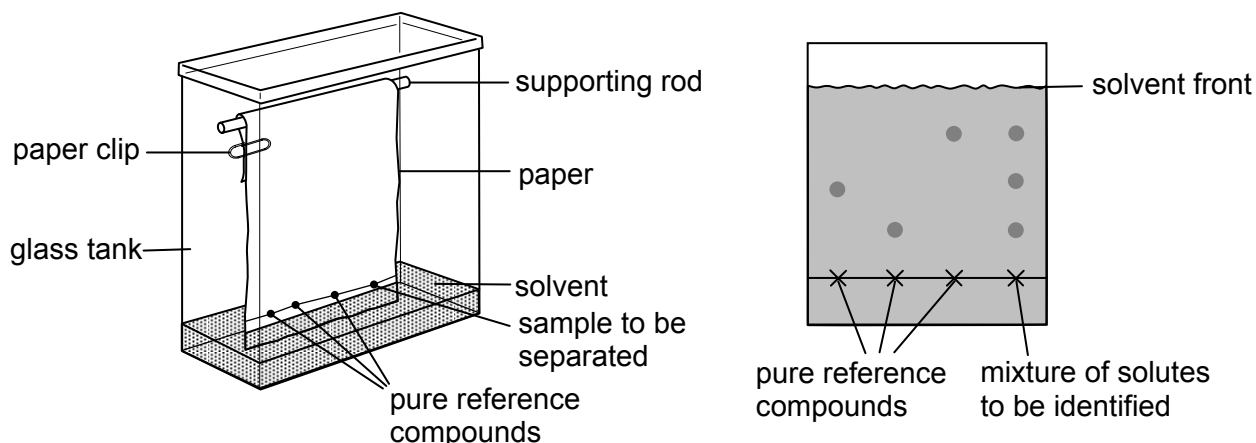


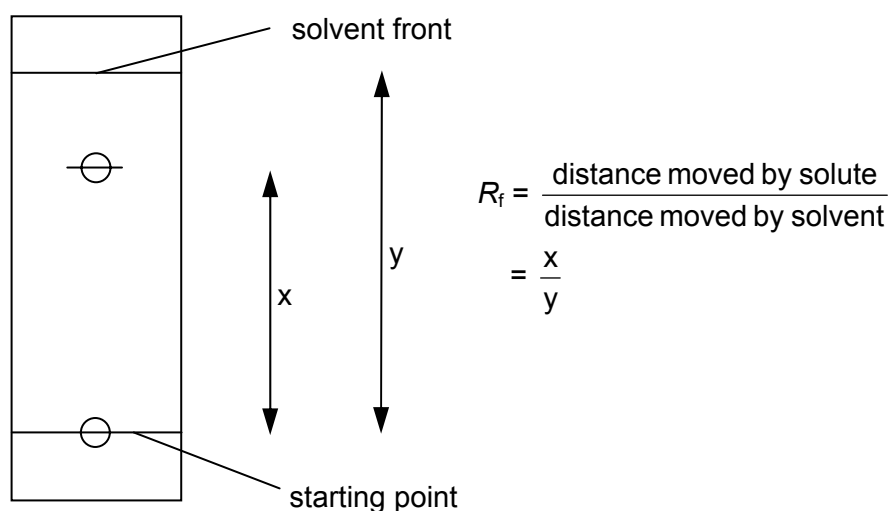
Figure 2.30 – paper chromatography and the resulting chromatogram

Once the solvent has reached the top end of the sheet of paper, the paper is removed and the solvent is allowed to evaporate. All solutes will now be re-adsorbed onto the area of the paper where they had reached, and their presence can be detected by their colour, or by their UV absorbance, if appropriate. For colourless compounds, an easy way of visualising the spots is to place the paper in a beaker with a few iodide crystals. The iodine evaporates slightly, and the vapour is preferentially absorbed by the less polar solute spots, making them appear brown, against a white paper background. Alternatively, the paper can be sprayed with a dilute solution of a reagent that forms colours with the compounds contained in the spots. Ninhydrin is used to visualise amino acids and small peptides; Molisch's reagent (a mixture of 1-naphthol and hydrochloric acid) can be used for sugars in general; Tollens' reagent can be used for reducing sugars such as glucose and maltose.

In Figure 2.31 the distance x that is travelled by a component in the solute mixture and distance y is that travelled by the solvent front. The ratio R_f , called the retention ratio (or retardation factor), is defined by

$$R_f = \frac{x}{y} = \frac{\text{distance moved by solute}}{\text{distance moved by solvent}}$$

Each solute or component has a characteristic R_f value for a given solvent. This means that the value of R_f can be used to identify a possible component. However, for this to be so, the chromatogram needs to be carried out carefully, under controlled conditions. For example, the solvent may need to be buffered, and draughts must be excluded by placing the chromatographic paper inside a glass tank closed with a lid.

Figure 2.31 – calculating the retention ratio, R_f

Sometimes, different compounds may have very similar R_f values in a particular solvent, and so no separation can be achieved. If a different solvent is tried, the partition coefficients of the two compounds between the new solvent and the stationary phase may be sufficiently different to allow a separation.

A complex mixture can be more thoroughly analysed by two-way paper chromatography. A spot of the mixture is placed towards one corner of a square sheet of chromatography paper (only one spot can be applied – no references can be used). One solvent is used first, and the solvent front allowed to reach the far edge of the paper, after which the paper is placed with the left side at the bottom for a second development using a different solvent. The spots are then more widely dispersed and can be recognised from their positions on the final chromatogram. This two-way technique is commonly used to identify the amino acids obtained from the hydrolysis of a protein (see Figure 2.32).

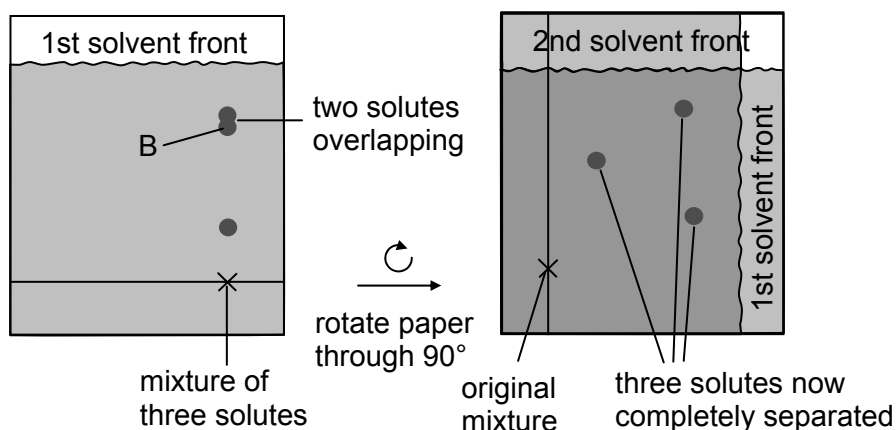


Figure 2.32 – two-way paper chromatography

SAQ 10. Calculate the R_f value in the **second** solvent of the spot labelled B.

Thin layer chromatography

Thin layer chromatography (TLC) is similar to paper chromatography, but the stationary phase is a thin layer of a solid such as alumina or silica supported on an inert base such as glass, aluminium foil or insoluble plastic. The plate can be prepared in the laboratory by using a slurry of the powder and then carefully oven-dried; practice is necessary to make standard and uniform coverings of the plates. They can also be obtained from laboratory suppliers, which is the normal practice. The plate is relatively small, e.g. about 15 cm x 5 cm.

The mixture to be analysed is 'spotted' at the bottom of the TLC plate and allowed to dry. The plate is placed in a closed vessel containing solvent (the mobile phase) so that the liquid level is below the spot.

TLC has advantages over paper chromatography in that its results are more reproducible, and that separations are very efficient because of the much smaller particle size of the stationary phase.

The solvent ascends the plate by capillary action, the liquid filling the spaces between the solid particles. This technique is usually done in a closed vessel to ensure that the atmosphere is saturated with solvent vapour and that evaporation from the plate is minimised before the run is complete. The partition here is between the solute adsorbed onto the SiO_2 or Al_2O_3 particles (via polar/hydrogen bonded interactions, but also some acid/base attractions, since SiO_2 is slightly acidic, and Al_2O_3 slightly basic) and the solute dissolved in the moving phase. The plate is removed when the solvent front approaches the top of the plate and the position of the solvent front is recorded before it is dried (this allows the R_f value to be calculated).

TLC has applications in industry in determining the progress of a reaction by studying the components present; and in separating reaction intermediates. In the latter case a line of the reaction mixture is 'painted' across the TLC plate instead of a single spot, and the line of product after separation is cut out of the plate and dissolved in an appropriate solvent.

The ways in which spots of colourless compounds on a TLC plate can be visualised are similar to those used for PC, but an additional technique is often used for compounds containing aromatic rings, or other systems that absorb UV radiation at 254 nm: the silica or alumina is impregnated with a fluorescent insoluble compound that absorbs UV light and emits it as visible light. When placed under a UV lamp, the plate emits a bright white light except where a UV absorbing compound is situated. Here, a dark spot is observed.

Gas/liquid chromatography (GLC)

This technique uses a gas as the mobile phase, and a non-volatile liquid coated onto small inert particles as the stationary phase. The particles are packed into a narrow glass column a few mm in diameter and between 1 m and 3 m long. The column is mounted inside an oven whose temperature can be controlled, or even programmed to increase as the separation is underway.

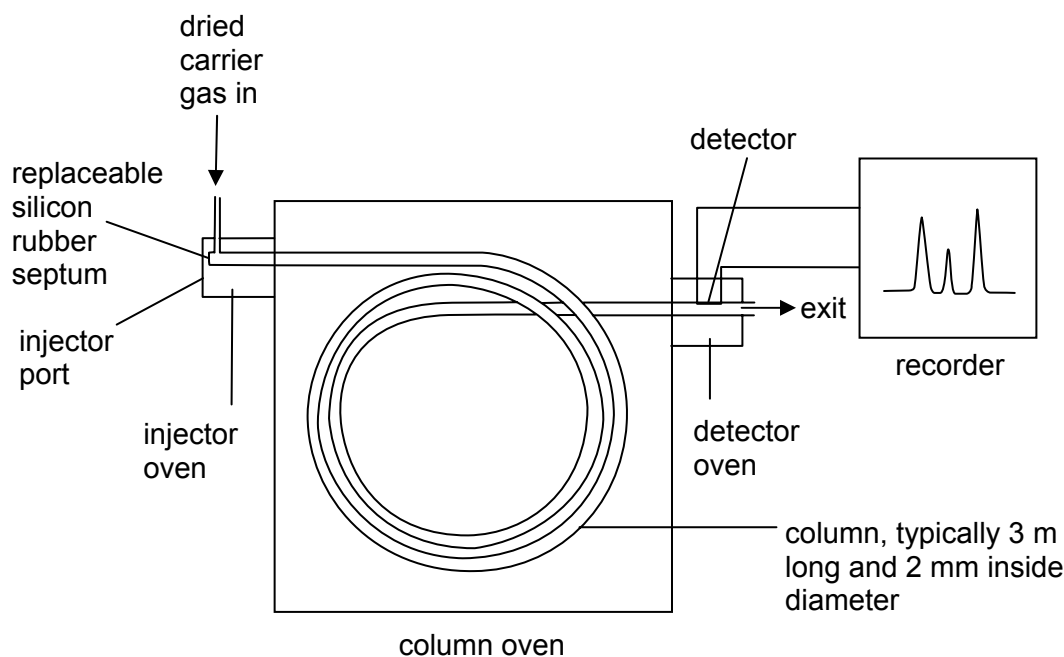


Figure 2.33 – GLC apparatus

For separation or identification the sample must be either a gas or have an appreciable vapour pressure at the temperature of the column - it does not have to be room temperature. The sample is

injected through a self-sealing disc (a rubber septum) into a small heated chamber where it is vaporised if necessary. Although the sample must all go into the column as a gas, once it is there the temperature can be below the boiling point of the components as long as they have appreciable vapour pressures inside the column. This ensures that all the fractions pass through the column over a reasonable time span. The injector oven is usually 50-100 °C hotter than the start of the column.

The sample is then taken through the column by an inert gas (known as the carrier gas) such as helium or nitrogen, which must be dry to avoid interference from water molecules. It can be dried by passing it through anhydrous copper(II) sulphate or self-indicating silica (silica impregnated with cobalt(II) chloride). Unwanted organic solvent vapours can be removed by passing the gas through activated charcoal. The column is coiled so that it will fit into the thermostatically controlled oven.

The temperature of the oven is kept constant for a straightforward separation, but if there is a large number of components, or if they have similar affinities for the stationary phase relative to the mobile phase, then it is common for the temperature of the column to be increased gradually over a required range. This is done by using computer control, and gives a better separation if the boiling points of the components are close. It also gives a faster separation if some components are relatively involatile.

The fractions progress to the end of the column, and then to a detector. Two types of detector are commonly used: thermal conductivity detectors and flame ionisation detectors. Only the flame ionisation detector (FID) will be described here (see Figure 2.34). It is the one that is most commonly used, and is particularly useful for detecting organic compounds. The gas from the column is mixed with hydrogen and air, and is then burned. Some CH^* radicals, which are formed on combustion, are then oxidised to CHO^+ ions. These ions produce a current of about 1 μA through a cathode to a sensitive amplifier and then to a computer or chart recorder.

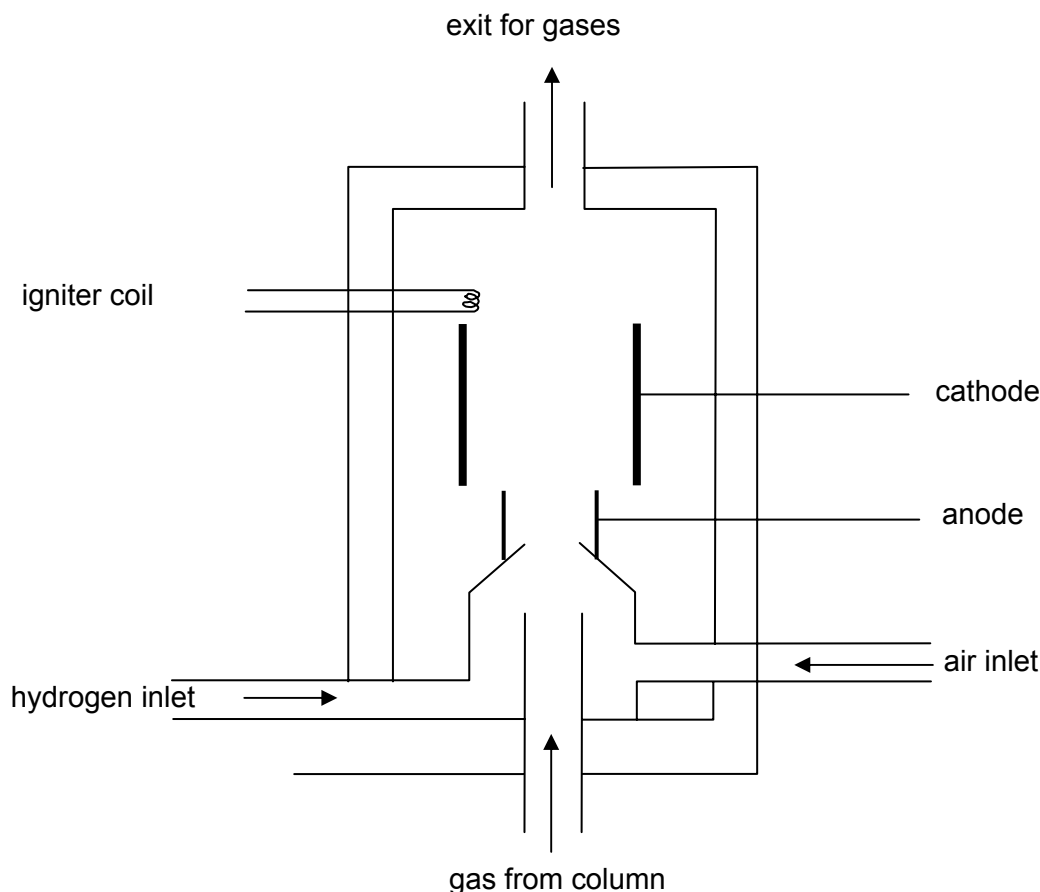


Figure 2.34 – flame ionisation detector

Once a mixture has been separated by GLC its components need to be identified. For known substances this can be done from a knowledge of the time it takes for the components to reach the

detector once they have been injected into the column. These are known as **retention times** and will vary depending on each of the following:

- the flow rate of the carrier gas;
- the temperature of the column;
- the length and diameter of the column;
- the nature of and interactions between the component and the stationary and mobile phases; and
- the volatility of the solute.

Each substance to be identified by GLC is run through the column so that its retention time can be determined. For compounds of completely unknown structure or composition the components must be collected individually and then analysed by using another method – e.g. mass spectrometry.

Figure 2.35 shows a GLC chromatogram of a mixture of organic chemicals.

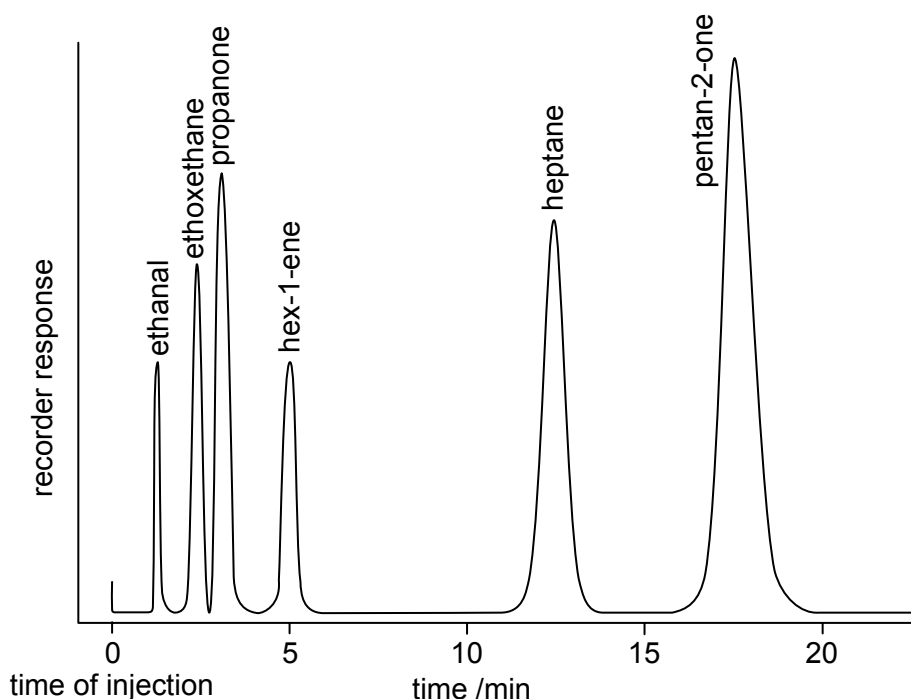


Figure 2.35 – a GLC chromatogram of a mixture of organic chemicals

SAQ 11. (a) Referring to Figure 2.35, which compound is present in the largest amount?

(b) Which compound has a retention time of 2.5 minutes?

High performance liquid chromatography (HPLC)

The principle behind HPLC, and the equipment used, is very similar to GLC. The main differences are that the moving phase is a high-purity solvent rather than a gas, the columns are shorter (10 – 30 cm) and the components are usually detected by measuring the absorbance of UV radiation through a microcell at the end of the column (see Figure 2.36). The efficiency of a separation increases if the particles in the stationary phase are made smaller. This is because the solute can equilibrate more rapidly between the two phases. The stationary phase in an HPLC column normally consists of uniform porous silica particles of diameter 1×10^{-6} m, with surface pores of 1×10^{-9} m diameter to increase the surface area. Sometimes the particles are coated with molecular fragments of a solvent having particular polar groups, joined to the silica by covalent bonds. This ensures they do not dissolve in the mobile phase. Thus it is debatable whether this is an example of the solid adsorption-liquid partition or the liquid-liquid partition of the solute.

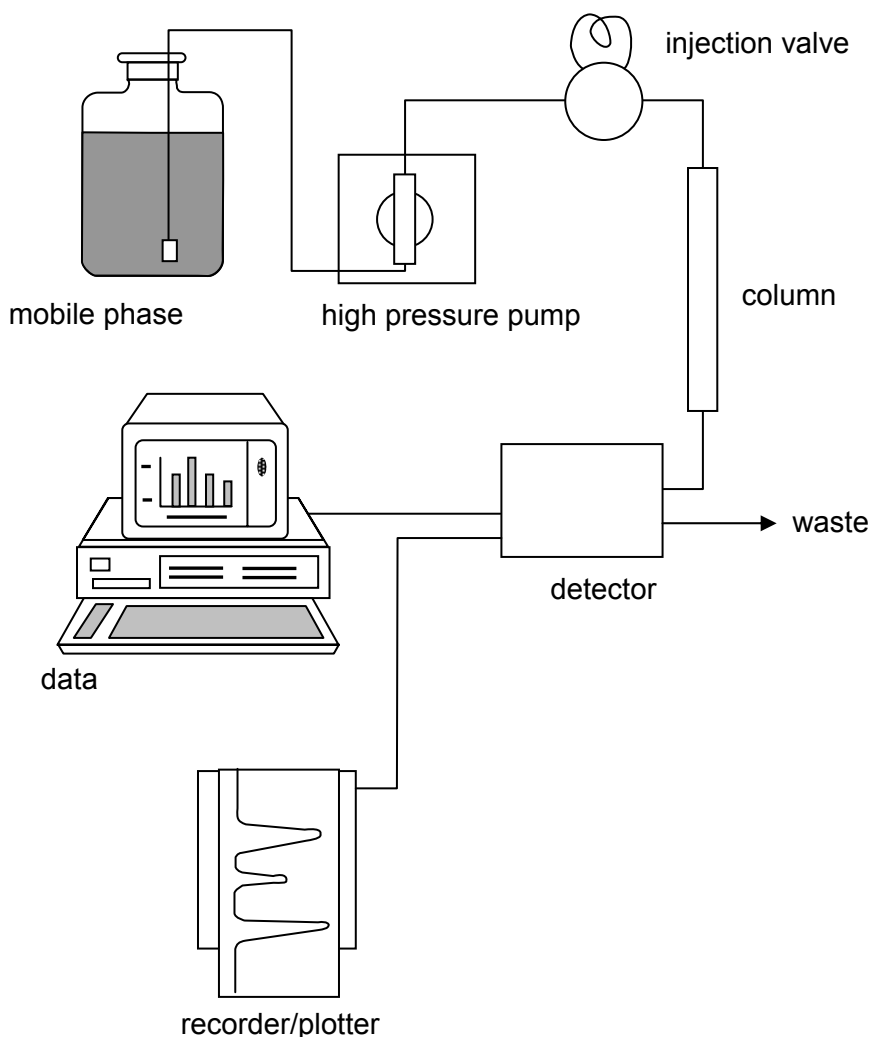


Figure 2.36 – HPLC apparatus

To ensure reproducibility of retention times, various factors have to be controlled exactly. A constant flow rate is maintained by special twin cylinder reciprocating pumps, generating steady precise pressures of up to 100 atmospheres. Even with such high pressures, the flow rate through the column is small (about $2 \text{ cm}^3 \text{ min}^{-1}$), due to the tightly packed small particles of the stationary phase. Nevertheless, because the column volume is small the timing of the injection of the sample must be precise if an accurate retention time is to be measured. For example, a 20 cm^3 column of diameter 0.4 cm has a volume of about 2.5 cm^3 . The mobile phase will take only 75 sec to travel through the whole of this column.

Applications of Chromatography in Analysis

The relative use of HPLC and GLC varies from industry to industry and very much depends on the compounds to be separated. Many compounds decompose at the temperatures required for efficient GLC separation, while HPLC separation can be achieved readily. However, GLC is particularly useful in detecting residual solvents in formulations and is also invaluable in looking for degradation products. Amines and acids are not separated well by GLC because they tend to be too polar, but can be efficiently separated using HPLC.

Other methods are often used in conjunction with HPLC for determining the presence of components in the output stream, and/or identifying the compounds in those components. These include:

- mass spectrometry
- infrared spectroscopy
- visible spectroscopy and
- ultraviolet spectroscopy.

Some examples are:

- The residual caffeine content in “decaffeinated” coffee and tea can be analysed by extraction with hot water and analysis of the solution by HPLC, comparing the retention times and the peak heights with those of standard solutions of caffeine.
- Dope testing of racehorses. Each year about 2/3 of all race winning horses are tested for the presence of performance-enhancing drugs (or even sedatives, that can cause the horse to knowingly *under*-perform). The procedure is as follows.
 1. A sample of the horse’s urine is first separated into acidic components, basic components and neutral components by passing through various silica chromatography columns.
 2. After elution from the columns, the various components are analysed as follows:
 - strong acidic components by TLC on silica or alumina
 - neutral components by HPLC
 - basic components by GLC followed by MSTo each component various standards are added for comparison of retention times etc.
- Detection of explosive residues on skin, clothing and in vehicles is often achieved by dissolving the residue in an organic solvent and analysing the solution by GLC or HPLC.
- Detection of pesticide residues in fruit and vegetable is routinely carries out in much the same way as the detection of explosive residues: the food is macerated in a buffer solution and extracted with an organic solvent (most pesticides are more soluble in organic solvents than in water). The solution is then analysed by HPLC, using standard samples of pesticides to compare retention times.

Electrophoresis

Although the appearance of a gel electrophoresis plate after development is similar to that of a thin layer chromatography plate, the principles behind the two methods of separation are very different. In electrophoresis there is no mobile phase, and only ions, not neutral molecules, move through the buffer solution along the plate.

The principle can be simply demonstrated as follows:

- soak a strip of filter paper with pH 7 buffer and clip it to a microscope slide with two crocodile clips
- place a drop of a dilute solution containing the three amino acids glycine, glutamic acid and lysine (see Figure 2.37) in the middle of the paper
- apply a voltage of 50-100 V DC across the crocodile clips, and leave for an hour or so
- disconnect the clips, and develop the spots with ninhydrin spray

The original spot will have been separated into three spots, as in Figure 2.38.

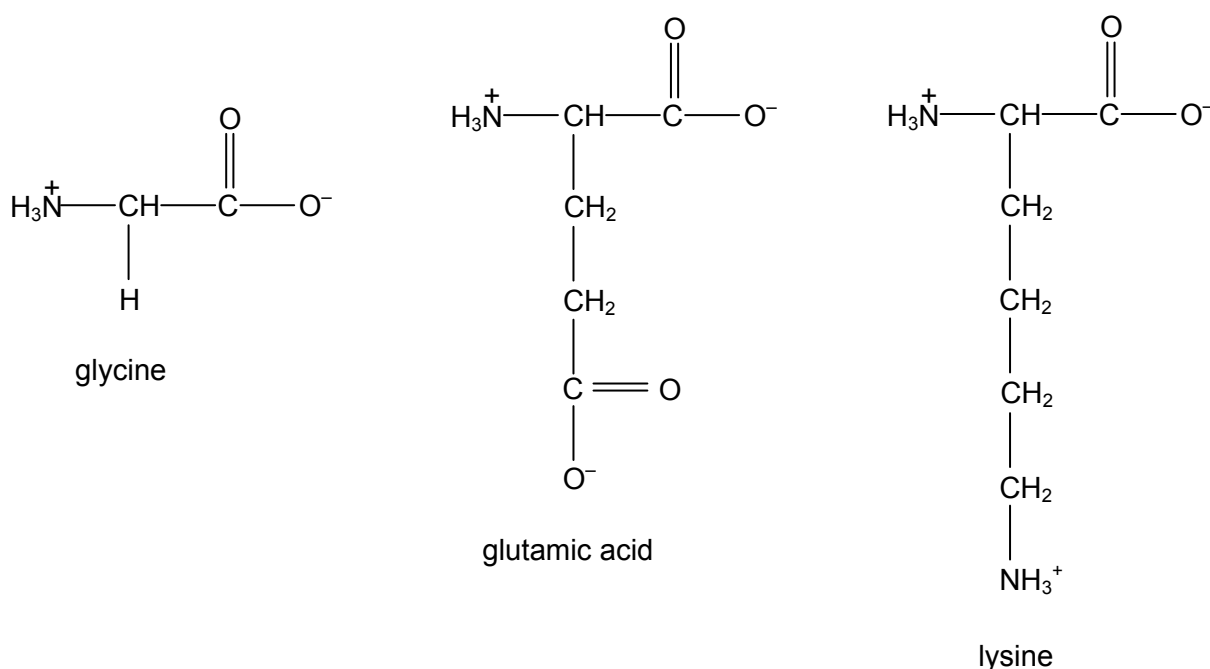


Figure 2.37 – structures of glycine, glutamic acid and lysine at pH 7

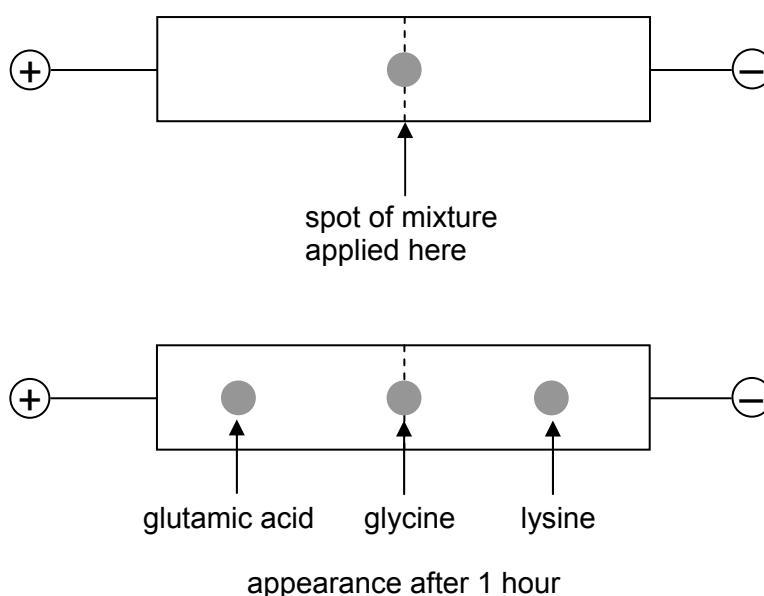


Figure 2.38 – results of electrophoresis of glycine, glutamic acid and lysine at pH 7

SAQ 12. Suggest and explain how the result of electrophoresis might differ if the experiment was carried out at pH 12 instead of pH 7.

In industrial, analytical or research environments, paper electrophoresis as described above is too imprecise a method to use. Either gel or capillary electrophoresis is employed. Gel electrophoresis is more common, and is the one described here.

The term **gel electrophoresis** covers a range of techniques that is used to separate, analyse and purify mixtures of biological molecules such as proteins and nucleic acids. These techniques can be adapted to:

- measure the relative masses of macromolecules;
- prepare nucleic acids and polypeptides for sequencing the component monomers e.g. purine and pyrimidine bases and amino acids; and
- separate proteins, so antibodies can be raised.

The sample mixture is placed in a gel and is separated into its constituents by applying an electric field to the gel, which is soaked in a liquid buffer. The gel is a sponge-like structure based on a three-dimensional polymeric network. It has the texture of a jelly.

Gels are used because their properties can be precisely controlled during their preparation, and they are more chemically stable as a support medium.

The components in the sample mixture have an electrical charge because proteins, like amino acids, carry either an overall positive or an overall negative charge, depending on the pH of their environment. Nucleic acids such as deoxyribonucleic acid (DNA) are usually negatively charged (at the phosphate groups in the chain) at the pH used for their separation using electrophoresis. The molecules move in response to an electrical field applied across the mixture. The rate of progress of the molecules depends on their size, charge and shape. On separation, the components are concentrated into bands or zones.

Each band or zone can be quantified using a variety of methods. Using sensitive techniques, such as bioassays or silver staining, amounts as small as 10^{-18} g can be detected.

Gel electrophoresis has a variety of applications such as checking the adulteration of foods, chromosome sequencing, DNA fingerprinting and characterising the chemicals responsible for allergic symptoms. It is used to screen infants' milk for α -lactoglobulin, a protein which is lethal for small babies.

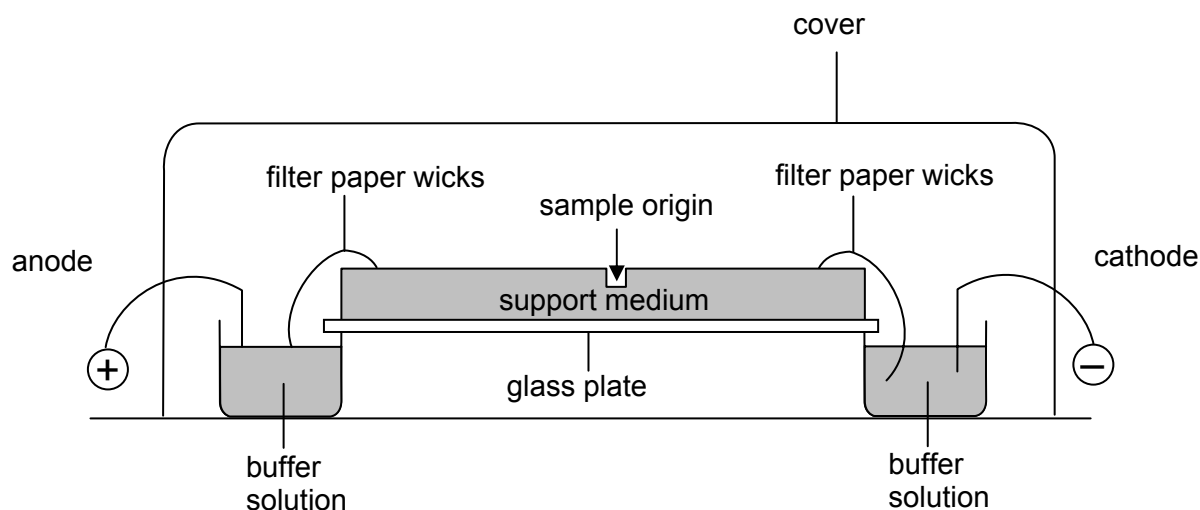


Figure 2.39 – gel electrophoresis apparatus

The main components used in a gel electrophoresis analysis (see Figure 2.40) are:

- an electrophoresis chamber;
- a gel support medium soaked in conducting buffer;
- a means of generating an electric field –e.g. a power pack;
- probes for detecting and/or measuring the separated molecules; and, if necessary,
- a means of extracting the individual products.

The mixtures are placed in small wells in the gel medium and when the electric field is applied the components of the mixture separate out. Their rate of migration across the gel becomes constant when the force of attraction between the electrode and the oppositely charged component is equal to the frictional force of the gel medium resisting the motion for that particular species. The relative rates at which components move and therefore the extent to which they separate are influenced by the strength of the field, the nature of the gel, and the surrounding buffer. Each species responds to these factors according to what is called its electrophoretic mobility (μ). This is a constant for each species.

Factors affecting mobility

1. Voltage: the velocity of a molecule is directly proportional to the voltage gradient across the gel.
2. Size: smaller molecules migrate quicker than larger molecules carrying the same charge.

3. Shape: a molecule with lots of side-chains experiences more frictional resistance than a linear molecule of the same mass and charge, and will therefore move more slowly.
4. Buffer pH: proteins exist as zwitterions like amino acids, and can be either positively or negatively charged because they contain both acidic and basic groups. The extent, and direction, of ionisation depends on the pH of the buffer.
5. Temperature: a rise in temperature can speed up electrophoresis, but can also denature the proteins. The temperature is normally controlled by cooling the plate with a flow of water underneath.

If a number of components in a sample have similar electrophoretic mobilities, a complete separation may not be achieved. In this case, either the buffer medium or the gel support can be altered. In addition, just as in paper chromatography, two-way electrophoresis in different buffers can often separate components.

Some medical applications of gel electrophoresis

Defects in newborn babies. Neural tube malformations can be detected from proteins leaking from the central nervous system of a foetus into the amniotic fluid (the fluid enveloping the foetus). Analysing these proteins can indicate spinal problems in new-born babies.

Sweating polypeptides. Over 400 polypeptide spots show up in a two-dimensional electropherogram of human sweat. Many of these have been previously unidentified.

Alcohol abuse. The extent of alcohol abuse can be investigated by analysing blood. This is because excess alcohol is associated with changes in acidic proteins and glycoproteins in blood plasma.

'Fish eye' disease. This is an inherited condition in which lipid is laid down in the eyes making them appear opaque like fish's eyes. Its medical name is dyslipoproteinanaemia. Protein samples from people with this condition have been analysed, suggesting that there is a deficiency in an enzyme system associated with lipid metabolism.

Heart attacks. Blood samples are taken at regular intervals (once a day for three days after initial chest pain). Studies are being done to identify more sensitive 'marker' proteins which will indicate early blockage in blood vessels.

Assessing fitness. When a person is very much out of condition there can be a serious increase in the amount of protein in urine, which can be detected by two-dimensional electrophoresis. As well as being a pointer to lack of physical fitness, increased protein content in the urine can also be a possible indicator of the onset of diabetes.

Other types of electrophoresis

Different techniques of electrophoresis include *capillary gel electrophoresis* and *capillary zone electrophoresis*. These have found useful applications in

- the analysis of haemoglobins to identify congenital diseases such as sickle-cell anaemia, thalassemia, and conditions associated with diabetes
- the detection of chemical weapons products – for example, the alkylphosphonates arises from the breakdown of sarin, one of the nerve gases
- detecting gunshot residues on skin and clothing
- detecting drugs by analysis hair samples
- separating milk proteins (see Figure 2.40)

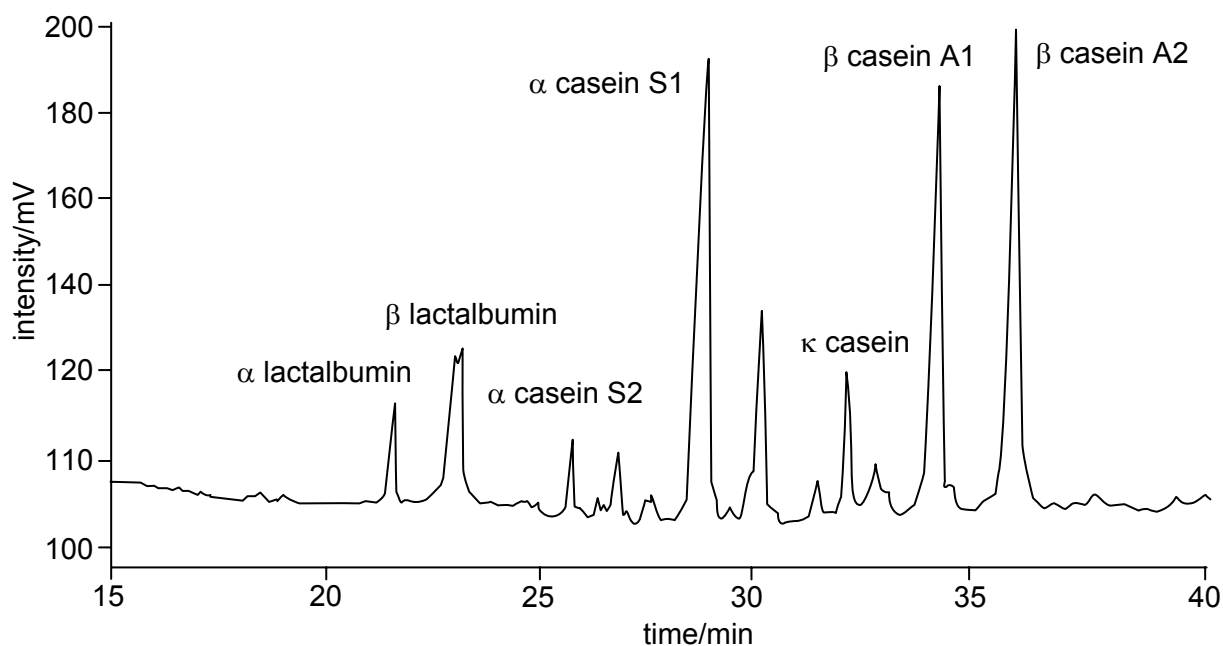


Figure 2.40 – separation of milk proteins

Genetic fingerprinting

The principle and method

Genetic fingerprinting is perhaps the most high profile and widely used application of the electrophoresis of nucleic acids.

DNA, or deoxyribonucleic acid, is the biological macromolecule that is contained within the chromosomes of cells. It is a two-stranded polymer and each strand is composed of a sugar (deoxyribose)-phosphate backbone, with a cyclic organic base attached to each sugar residue. The strands are held together by hydrogen bonds between the organic bases and twine around each other in a helical fashion (see Figure 2.41). All DNA molecules contain the same molecular backbones, and the same four bases. They differ only in their overall length (even the shortest has many thousand units in its chain) and in the order in which the bases are arranged along the backbone.

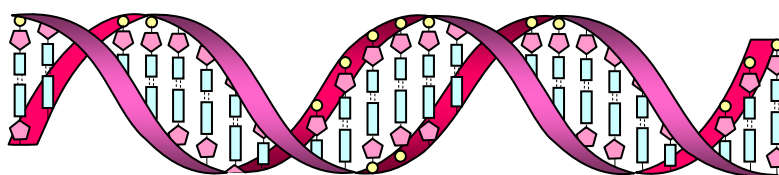


Figure 2.41 – structure of DNA

Genes are sections of a DNA molecule that become templates for the synthesis of strands of RNA (ribonucleic acid) which in turn form the code that translates into the arrangement of amino acids along the chains of the enzymes and other protein molecules in the organism. Of the DNA contained within the chromosomes, only about 10% comprises the genes. This small portion of “useful” DNA will be almost identical in most members of a particular species (and much will be identical for different species too). This is because any large change in this genetic material would be mirrored by a change in the amino acid sequence of the final protein, which would affect their efficiency as enzymes and could prove fatal to the organism.

The other 90% of chromosomal DNA (the “useless” DNA) lies in between the genes, and is highly variable in its base sequences. These portions often contain sequences of bases (about 10-50 base-pairs in length) that are repeated several times. All members of the same species have these

repeats, but individuals vary in the number of times each sequence is repeated. The areas are termed minisatellites, or VNTRs (variable number of tandem repeats). There are also quite short sequences (2-5 base-pairs) that are repeated many times. These are termed short tandem repeats (STRs), or microsatellites (see Figure 2.42). The key to genetic fingerprinting is that an individual's pattern of VNTRs or STRs is entirely unique – no one else (except an identical twin) will have the same pattern (see Figure 2.43). Furthermore, each individual inherits half their pattern from their mother, and half from their father. Members of a family throughout many generations will have similar minisatellite patterns to each other.

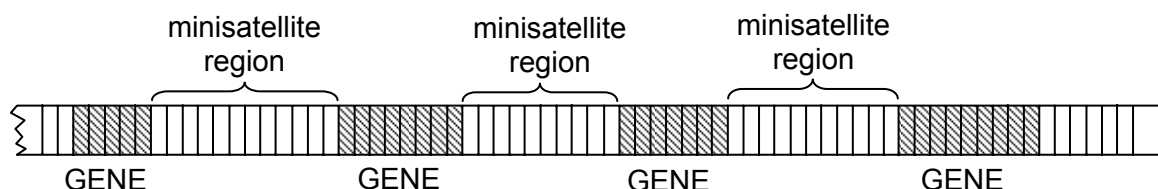


Figure 2.42 – strand of DNA showing genes and minisatellites

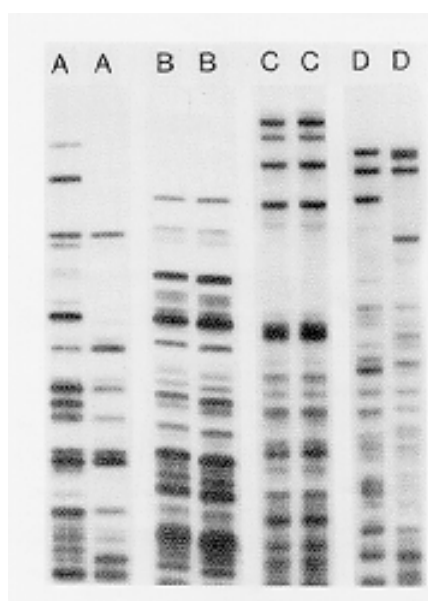


Figure 2.43 – DNA from twins

The technique of genetic fingerprinting starts with the extraction of the DNA from a sample of chromosomal material, such as blood (only the white cells contain DNA), hair, inner cheek cells, semen or skin. About 25 mg of DNA is required. The DNA is then broken into fragments by a restriction enzyme. There is a range of these enzymes available, each one breaking the DNA at several different, but known and specific, places. This produces fragments that can now be subjected to electrophoresis.

Modern techniques using the polymerase chain reaction (PCR) can be used when the amount of available DNA is smaller, in which case an analysis can start with as little as 0.2 nanograms (2×10^{-10} g). In this technique, after the DNA has been broken up into pieces in the usual way, an individual segment is extracted. Heating a solution of the segment separates the two DNA strands from each other. The sample is then cooled and a short “primer” length of DNA which matches the end of the segment is added, together with the enzyme DNA polymerase and the four monomer nucleotides A, G, C, T. On warming the solution the four nucleotide bases bond to the two single strands in the precise order as determined by the original base sequence, to form two new doubly-stranded DNA molecules. The process is repeated many times (it can be automated readily) until a sufficient amount of DNA has been made - 1 billion (1×10^9) copies would be produced after only 30 repeats.

The samples of DNA to be analysed are placed in wells near the cathode of an agarose gel plate. Since the phosphate groups along the chains of DNA are negatively charged, all DNA samples will move towards the anode, but the smaller fragments will move faster than the larger ones.

Discovering the positions of the separated fragments on the developed electropherogram requires a special technique, since the amounts of DNA are so small. Spraying and staining with a dye would give only a faint image. The DNA fragments are transferred from the gel onto a nylon membrane and are broken into single strands by soaking in dilute alkali. The membrane is then soaked in a solution containing short lengths of DNA identical to the minisatellite sequence, but labelled with radioactive ^{32}P . The ^{32}P DNA thus sticks to the membrane where the DNA fragments are situated, and if an X-ray sensitive film is placed over the plate, it will record the pattern originally on the gel plate (see Figure 2.44).

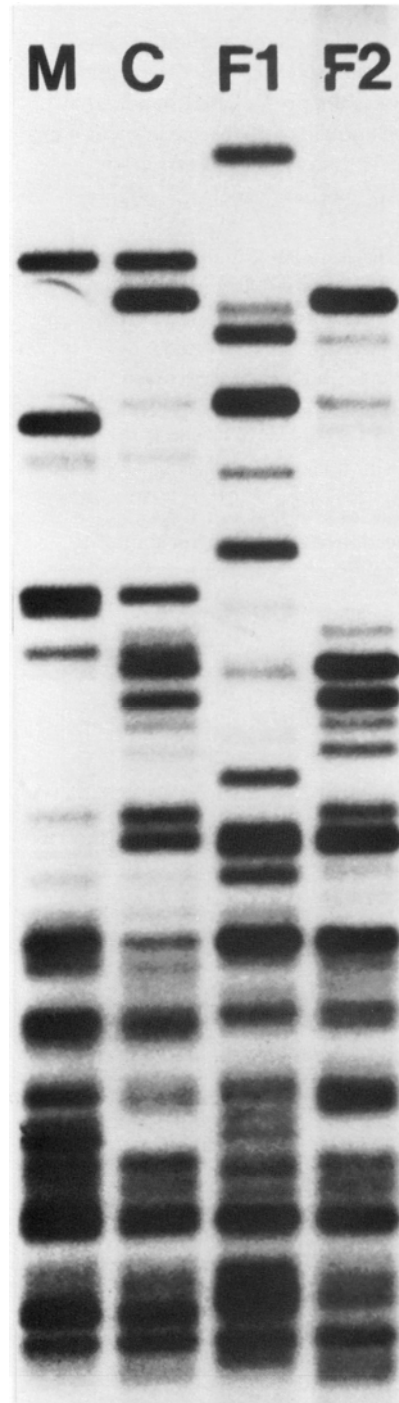


Figure 2.44 – genetic fingerprints of a child, C, and its mother, M, and two possible fathers, F1 and F2

The uses of genetic fingerprinting

Generic fingerprinting has been used in

- paternity testing
- establishing other familial relationships – between both the living and the dead
- establishing the relationship between archaeological artefacts
- forensic testing
- medicine

One example of each of these will suffice to illustrate the application.

Figure 2.44 demonstrated how genetic fingerprinting could be used to establish paternity. A child inherits one set of chromosomes from its mother, and the other set from its father. In the figure it can be seen that the child's fragments correspond with those of its mother, together with those of possible father F2, but show little correspondence to those of possible father F1.

The same technique can be used to establish familial links of children to parents – for example in immigration disputes. An interesting historical example is the investigation by STR analysis of the DNA from the bones of five bodies dug up in Russia in 1991. It was suggested that this was the family of the last Tsar of Russia, assassinated by a Bolshevik firing squad in 1918. The DNA of the three female children had clear correspondences to that of the two adults (who were not genetically related to each other), clearly showing the presence of a family group. Further work linked the mitochondrial DNA to that of the present Duke of Edinburgh, who was known to be related to the late Tsarina through a maternal line (mitochondrial DNA is only inherited from the mother, not the father). Thus the identities of the exhumed bodies were established.

The archaeological investigation of the Dead Sea Scrolls, written on goat skins two thousand years ago, has been made easier by the genetic fingerprinting of the DNA from small scrapings on the old skin cells. The various scrolls were written on the skins of different animals, but had over the years been broken into many fragments. Being able to piece together the original skins by matching the DNA has helped archaeologists make sense of the texts.

A high profile use of genetic fingerprinting was in the trial of the basketball player OJ Simpson in the USA in 1995. A sample of DNA from his cheek cells was identical to samples of blood found at the scenes of two murders he was accused of committing.

Another useful application of genetic fingerprinting in forensic science is its ability to *exclude* suspects, as well as give strong evidence to *include* one particular suspect. Figure 2.45 shows the DNA fingerprints of seven suspects, and of a bloodstain at the scene of a crime.

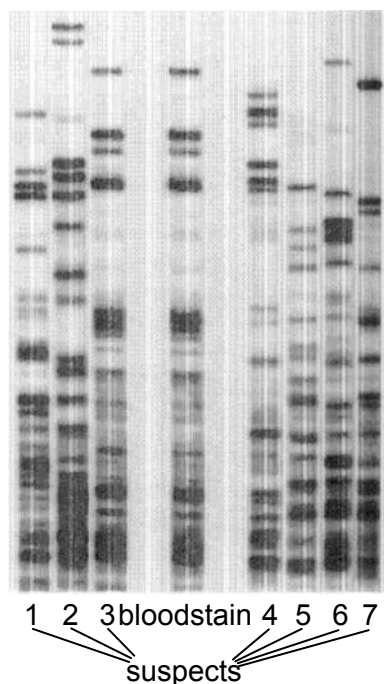


Figure 2.45 – DNA fingerprints of seven suspects and blood taken from the scene of a crime

SAQ 13.(a) Which suspect's DNA best matches the DNA of the bloodstain?

(b) Are there DNA samples from any of the other suspects that show a correspondence?

Medical applications have included the investigations of whether cases of tuberculosis or cancers are re-infections from previous illnesses, or brand new infections. This can help in their treatment. Recently a genetic fingerprinting analysis of various samples of the MRSA “superbug” *Staphylococcus aureus* from around the world has shown that the strain can be traced back to the 1950s, and even early examples were resistant to penicillin, streptomycin and tetracyclin antibiotics. This should help the development of new types of antibiotic.

Summary

- The ratio of concentrations of a solute in equilibrium with two immiscible solvent is given by the partition coefficient: $K_{pc} = \frac{[\text{solute}(\text{solvent A})]}{[\text{solute}(\text{solvent B})]}$.
- Partition of different components between a mobile phase and a stationary phase is the basis of chromatography, of which the four main type are paper chromatography (PC), thin layer chromatography (TLC), gas-liquid chromatography and high performance liquid chromatography (HPLC).
- These four techniques are used to separate and identify compounds in a number of applications in medicine, forensics, environmental monitoring.
- Electrophoresis is a useful technique for separating amino acids, proteins and nucleic acids
- A major application of electrophoresis is for genetic fingerprinting.

2.4 – Examples of applications of analytical chemistry

By the end of this section, you should be able to:

- appreciate the use of the techniques of gas liquid chromatography, mass spectrometry, solvent extraction in the analysis of environmental, archaeological and geological samples,
- understand the conclusions that can be drawn from such analyses,
- predict techniques suitable for the separation and analysis of samples taken from environmental sources.

Three methods of dating using mass spectrometry

Carbon dating

The use of the radioactive isotope ^{14}C in the dating of once-living artefacts is well known, but because of its fairly short half-life (5730 years) it becomes less accurate for ages greater than about 10 000 years. Using a “tandem accelerator” to purge samples of ^{14}N and $^{12}\text{CH}_2$ (both of which have an m/e value of 14), and then analysing the $^{14}\text{C}:^{12}\text{C}$ ratio in a conventional mass spectrometer, it is possible to extend the range to 100 000 years.

Potassium-argon dating

Naturally occurring potassium contains (now) 0.01167 per cent of the radioactive isotope ^{40}K with a half-life of 1.3×10^9 years. When it undergoes radioactive decay, one of the products is ^{40}Ar . If an igneous rock sample contains potassium, and the argon formed by the decay of ^{40}K is unable to diffuse out of the crystal lattice, it is possible to determine the age of the rock sample – that is, the time when the rock crystallised from the magma.

The argon content of a sample is determined by using a mass spectrometer to analyse the gas released when the sample is heated and melted in a vacuum. (It is assumed that none of the argon produced within the crystal diffuses out of the crystal lattice of the rock, and that there has been no loss or gain of potassium since it was formed.) A known amount of ^{38}Ar is added, and the ^{40}Ar content is measured as a proportion of the ^{38}Ar , in a method known as *isotope dilution*. The total remaining potassium can be measured by using several methods, including flame photometry, atomic absorption spectrometry and isotope dilution. By using the half life of ^{40}K and the $^{40}\text{K}:^{40}\text{Ar}$ ratio, the age of the rock can be determined.

Oxygen isotopic ratios in the dating of ice cores

Naturally-occurring oxygen consists of three isotopes, ^{16}O , ^{17}O and ^{18}O , in the percentage abundances of 99.76%, 0.04% and 0.20% respectively. Most of the ice in the ice sheets at the poles has been formed over many thousands of years by water molecules evaporating from the oceans and travelling in the vapour state towards the polar regions, where they precipitate as snow or hail. H_2^{16}O evaporates, and travels through the atmosphere, more quickly than H_2^{18}O . Thus the snow precipitating in the polar regions is less rich in ^{18}O -water than is the water in the oceans. This is a temperature-dependent phenomenon, however. Snow that falls in the winter months is even more depleted than that which falls in the summer. Thus by measuring the $^{16}\text{O}/^{18}\text{O}$ ratio continuously throughout its depth, and counting the annual changes (in a similar way to aging a tree by counting its annual growth rings), the age of an ice field can be determined.

For example, the Vostock Ice-Core was collected in east Antarctica by a Russian expedition over the years 1970 – 1983. It measured over 2000m long, representing just over half the thickness of the ice sheet from which it was collected. The age of the bottom of the ice core was calculated at about 150 000 years – the origins of the ice sheet itself must be considerably older.

The use of combined GLC-MS in environmental monitoring

Biomarkers in the petrochemical industry

Biomarkers are organic compounds whose carbon skeletons provide an unambiguous link with a known natural product, and these are particularly useful in the petrochemical industry.

When the marine organisms that formed crude oil died, their bodies were subjected to extreme heat and pressure, resulting in the formation of the oil. In the chemically reducing environment present under the sea many of the natural products – e.g. steroids and terpenoids – became saturated (i.e. highly reduced to form cycloalkanes).

These saturated products are now still present in crude oil, and serve as biomarkers. It is thus possible to identify the organisms from which the oil was formed, enabling the terrain at the time of formation to be deduced. Once the biomarkers from a particular oilfield have been identified it is also possible to determine where an oil sample originated. Thus, if an oil spillage occurs at sea and nobody admits responsibility, the biomarkers could be used to identify its source. It may then be a simple matter to find out which tankers were carrying oil from that field at the time of the incident, and so identify the culprit.

Identifying the biomarkers is done by separating the components of the crude oil using GLC and feeding them directly into the chamber of a mass spectrometer. The components can be identified by their retention times on the GLC column and the fragmentation patterns of their mass spectra.

Figure 2.46 shows GLC traces of the steroidal cycloalkanes fractions of two samples of crude oil.

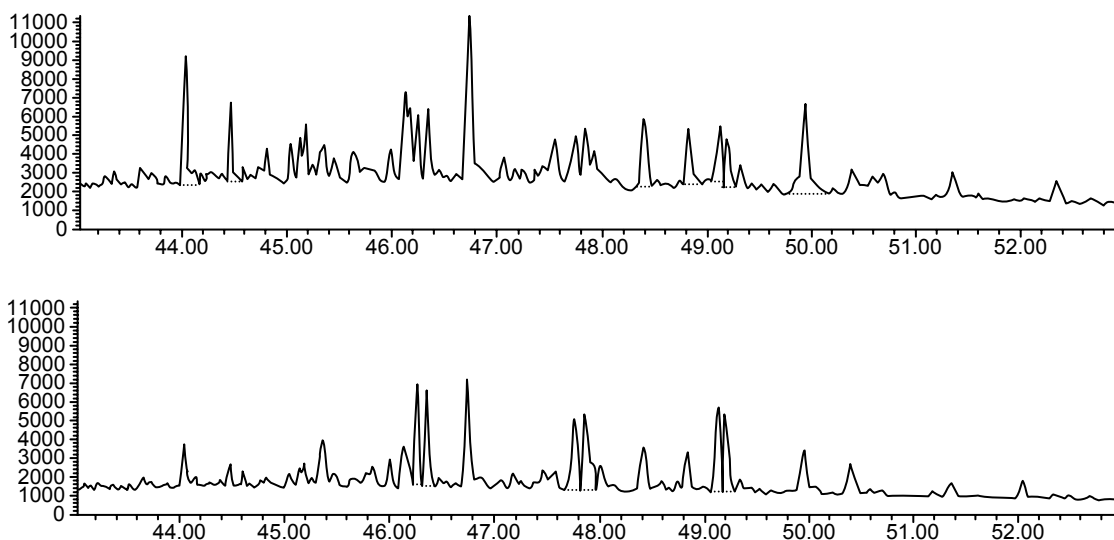


Figure 2.46 – GLC traces of the steroidal cycloalkane fractions of two samples of crude oil

SAQ 14. Are the two samples identical or not? Choose a number of peaks from the two traces to support your choice.

Detecting PCBs and dioxins

Although concentration methods are sometimes required to supply the spectrometer with sufficient sample to analyse, GLC-MS is sensitive enough to detect pollutants such as polychlorinated biphenyls (PCBs), which are found in surface water in concentrations on the nanogram per dm^3 ($1 \times 10^{-9} \text{ g dm}^{-3}$) scale. Monitoring for dioxin can also be done by using mass spectrometry. Dioxin (more correctly, tetrachlorodibenzodioxin) can be found in the effluent stacks of incinerators where PVC and other chlorinated polymers have been burned. Both PCBs and dioxin are stable, fat-soluble compounds that do not degrade in nature, but build up in food chains. Some seals, for example, have been found to have several per cent of PCBs in their total fat content.

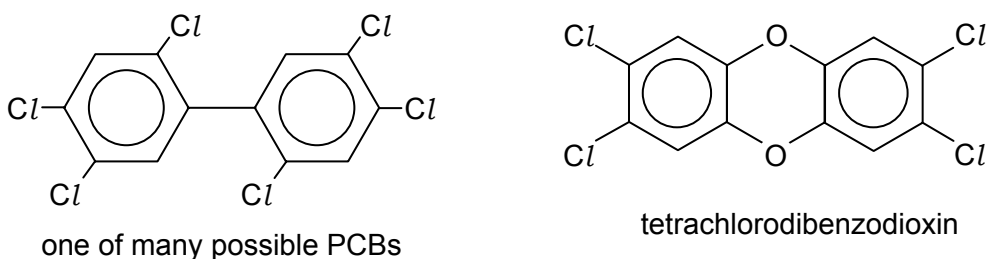


Figure 2.47 – a PCB and a dioxin

The Mars Viking lander

Mass spectrometers have been adapted for use on board spaceships, where less attention has to be paid to the inclusion of vacuum pumps because the atmosphere is already at low pressure. The Viking spacecraft that landed on Mars was equipped with GLC-MS, and both the Martian atmosphere and the Martian soil were analysed. It found that the partial pressures of water, nitrogen and carbon dioxide in the atmosphere are very low, and that the soil contained a lot of chlorine and sulphur.

Two stories of detection by mass spectrometry**The solving of a historical mystery**

In the spring of 1845 Sir John Franklin and a crew of 128 men set sail to try to complete the North West Passage – a hoped-for route from Europe to Asia around the north of Canada. They never returned, and despite several search parties being sent out, the first evidence of their demise did not come to light until 1850. It was not until 1981, and later in 1986, that an investigation by physical anthropologist Owen Beattie was carried out on the still-frozen remains of crew members. An analysis of a sample of hair showed it contained 600 ppm (parts per million) of lead, which is 120 times the amount one might expect in hair. Nearby, Beattie found a pile of old empty tin food cans, which had been the expedition's provisions during their last winter. He observed that there were lumps of solder in the inside of the seams. Until the late 19th century the solder used to seal food tins consisted of 50% lead and 50% tin.

The isotopic composition of lead varies from one ore to another, due to lead being the final stable element from various different radioactive decay chains. No two samples of lead will contain exactly the same proportions of the four main isotopes ^{204}Pb , ^{206}Pb , ^{207}Pb and ^{208}Pb . By analysing the lead from the hair and that from the solder on the tins Beattie found them to be identical.

The crew had been poisoned by the lead that had dissolved in the weak acids in the food contained in the cans.

An ecological detective story

This account illustrates the use of both solvent extraction and GLC-MS in the separation and identification of drug samples.

In the spring of 1991, various incidents were reported of eagles being found dead or dying near the small town of Kodiak in the USA. There was no clear evidence of any disease, and nothing to suggest violent injury. Poisoning was suspected. Inside the eagles' stomachs, there were still pieces of undigested meat from their last meal. They had clearly eaten well, but died soon afterwards. This suggested that any poison that they may have eaten was not slow-acting, but had worked quickly.

A sample of the meat was macerated, acidified with $\text{HCl}(\text{aq})$ and filtered. The aqueous extract was then shaken with ether and the layers separated. Any substances that were basic would remain in the $\text{HCl}(\text{aq})$ as ionic salts (many drugs and poisons are amines, which are bases), whereas neutral or acidic covalent compound would dissolve in the ether layer. The aqueous layer was then made alkaline with $\text{NaOH}(\text{aq})$ to liberate the free un-ionised bases, and extracted with trichloromethane.

The ether and the chloroform layers were separately analysed by GLC-MS using a capillary column 30 metres long. The retention times and the fragmentation patterns of the various components were compared to those of known poisons, pesticides etc.

No identifiable drugs or poisons were found in the chloroform extract, but the mass spectrum of one of the components in the ether extract contained fragments at m/e values of 71, 155, 197 and 226. On searching its computerised memory, the mass spectrometer identified this component as **pentobarbital**, a popular short-acting barbiturate derivative used primarily as a sedative.

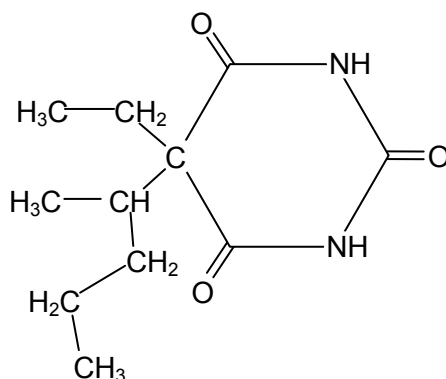


Figure 2.48 – pentobarbital

SAQ 15. Identify the molecular formulae and possible structures of the 4 peaks in the mass spectrum of pentobarbital

The next stage was to identify the source of the meat. Extracting the proteins from the meat and testing them with samples of serum from various species of animal produced a positive reaction with cat serum. Further investigations found that the local animal shelter had been inundated with stray cats for which they could find no homes, so they had painlessly “put them to sleep” with large doses of pentobarbital. Unfortunately they had not buried the cats’ bodies deep enough for the eagles not to find them. Once the source of the poison had been identified, the animal shelter was much more careful about the disposal of the dead bodies, and the eagle population is now back to its normal level.

Summary

- Partition between solvents, various types of chromatography and electrophoresis find many applications in environmental monitoring.
- Mass spectrometry – either on its own or coupled to GLC or HPLC, is of great use in archaeological and geological dating, and in the analysis of pesticides and other pollutants.

2.5 – Key definitions

absorption: the reduction in the intensity of radiation when passed through a sample

adsorption: the attraction of a molecule to the surface of a solid

base peak: the peak in a mass spectrum that corresponds to the most abundant fragment

capillary: a glass tube with an internal diameter of about 1 mm, used to make columns for GLC, HPLC and capillary electrophoresis

carbon dating: a method of dating once-living artefacts containing carbon or its compounds by measuring the $^{12}\text{C}/^{14}\text{C}$ ratio

chemical shift: (symbol δ) the frequency of absorption of an NMR peak compared to the frequency of the TMS peak, expressed as parts per million

chromatography: the separation of the components of a mixture by their different speeds of movement through/over the stationary phase

δ scale: see chemical shift

diamagnetism: the phenomenon of the electrons in a molecule moving round their orbits so as to produce an induced magnetic field that opposes an applied field

dioxins: stable, poisonous compounds with the formula $\text{C}_{12}\text{H}_4\text{O}_2\text{Cl}_4$, produced when chlorinated plastics are incinerated or as a by-product in the preparation of some pesticides

DNA: deoxyribonucleic acid. The double helix polymer that contains the hereditary information of the genes

electropherogram: the visual representation of the results of electrophoresis

electrophoresis: the separation of ions/charged molecules by their different rates of movement in an electric field

electrophoretic mobility: a measure of the extent of movement of a particular charged molecule under standard electrophoresis conditions

fragmentation: the breaking up of a molecule into smaller parts by the cleavage of covalent bonds
gel electrophoresis: electrophoresis using a thin layer of gel supported on a glass plate.

genetic fingerprinting: the technique of using gel electrophoresis to create a profile of an individual's pattern of STRs or VNTRs

GLC: gas-liquid chromatography

GLC-MS: a technique where a mass spectrometer is coupled directly to the outlet of a gas chromatograph, to allow the mass spectrum of each peak to be obtained as soon as it is eluted.

heavy water (D_2O): Water containing the isotope ^2H (given the symbol D) instead of ^1H .

HPLC: high performance liquid chromatography

hydrogen bonding: a relatively strong intermolecular force between a δ^+ hydrogen atom of one molecule and a δ^- nitrogen, oxygen or fluorine atom of another molecule

immiscible: two liquids that do not mix/dissolve in each other

induced dipole forces: see van der Waals forces

integration trace: the trace on an NMR spectrum that indicates the relative number of hydrogen atoms responsible for each peak in the spectrum

isotopes: atoms of the same element with different numbers of neutrons, and hence different masses

isotopic dilution: a method of estimation of the amount of an isotope present in a sample by adding a known amount of a different isotope of the same element and comparing their relative abundances in a mass spectrum

K_{pc}: see partition coefficient

lattice: a regular arrangement of particles (atoms, molecules or ions) in a crystalline solid

M+1 peak: an ion that has a mass one unit greater than the molecular ion in a mass spectrum, due to the presence of the ¹³C isotope in naturally occurring carbon

M+2 peak: an ion that has a mass two units greater than the molecular ion in a mass spectrum, due to the presence of the ³⁷Cl isotope in chlorine, or the ⁸¹Br isotope in bromine

mass/charge ratio: the ratio of the mass of an ion (in atomic mass units) to its charge (in units of the charge on the electron)

MHz: megahertz (1×10^6 Hz). A unit of frequency of electromagnetic radiation

mobile phase: the solvent in the chromatography process, which moves through the column or over the paper or thin layer

molecular ion: the ion formed by the loss of an electron from the original complete molecule during mass spectrometry

monochromatic: radiation that consists of a single wavelength, or a very small range of wavelengths

MS: mass spectrometry

NMR: nuclear magnetic resonance spectroscopy

partition: the division of a solute between two phases

partition coefficient (K_{pc}): the ratio of the concentrations of a solute in two different immiscible solvent when an equilibrium has been established

PC: paper chromatography

PCBs: polychlorobiphenyls – inert compounds used as insulators in transformers, and plasticisers in PVC, paints and adhesives. They persist in the environment and can cause cancer and other illnesses

PCR: polymerase chain reaction – a method of making many thousands of copies of specific fragments of DNA using the enzyme DNA polymerase

pesticide: a chemical applied to crops to kill weeds, insects or fungi

polarity of a solvent: the extent to which a solvent dissolves polar molecules (i.e. those with dipole moments)

potassium-argon dating: a method of dating rocks by measuring the ⁴⁰Ar/⁴⁰K ratio

relative abundance: the height of a peak in a mass spectrum compared to the height of the base peak, expressed as a percentage

restriction enzyme: an enzyme that splits the DNA molecule in specific places.

retardation factor: see retention ratio

retention ratio: a number between 0 and 1 which is a measure of how far a component has travelled compared to the distance travelled by the solvent front during PC or TLC

retention time: the time taken for a component in a mixture to travel through the column in GLC or HPLC

R_f: see retention ratio

shielding of protons: the screening of a proton from the applied external field by the induced diamagnetic field of the electrons surrounding it

solute: the substance that dissolved in a solution

solvent: the liquid in which the solute dissolves

solvent extraction: the process whereby a solvent is used to dissolve solutes from a mixture (e.g. leaves or seeds), or whereby a second (immiscible) solvent is used to extract a solute out of another solvent

spin-spin coupling: the interaction between the spins of protons on adjacent atoms, causing their respective NMR absorbances to be split into a number of lines

splitting pattern: the pattern or lines produced when spin-spin coupling occurs

stationary phase: the immobile phase in chromatography, e.g. the paper surface in PC, the surface of the thin-layer particles in TLC, or the involatile liquid adsorbed onto the column in GLC or HPLC

STR: short tandem repeats – very short stable segments of DNA used in genetic fingerprinting

tesla: a unit of magnetic field strength, about 10,000 times as strong as the Earth's magnetic field

TLC: thin layer chromatography

TMS: tetramethylsilane. An inert volatile liquid used as an internal reference in NMR

two-way chromatography: a technique used in PC or TLC wherein one spot of a mixture is placed at the corner of a square sheet and is developed in the first solvent as usual. The sheet is then turned through 90 and developed in the second solvent. This allows a better separation of components having similar R_f values

unit cell: the smallest 3-dimensional repeating unit in a crystal lattice

van der Waals' forces: weak intermolecular forces arising from the oscillation of electrons around the nuclei in the atoms in a molecule, causing an instantaneous dipole in one molecule which then induces an attractive dipole on an adjacent molecule

VNTR: variable number of tandem repeats – 5-10-base lengths of DNA that repeat themselves along the chain, used in genetic fingerprinting

X-rays: electromagnetic waves of very short wavelength, about 0.1 nm (visible light has a wavelength about 500 nm)

XRC: X-ray crystallography

2.6 – Resource list

Modern Chemical Techniques; An Essential Reference for Students and Teachers, by Ben Faust, Royal Society of Chemistry, ISBN 0 85404 286 5

More Modern Chemical Techniques by R Levinson, Royal Society of Chemistry, ISBN 0 85404 929 0

Methods of Analysis and Detection (Cambridge Advanced Sciences) by Anne McCarthy, Cambridge University Press, ISBN 0 52178 724 6

Cutting Edge Chemistry by Ted Lister, Royal Society of Chemistry, ISBN 0 85404 914 2

For an online list of resources in analytical chemistry, see LearnNet, Royal Society of Chemistry, <http://www.chemsoc.org/CFLEARNNET/dets.cfm?subj=a>