Since its first observation in bulk phases in 1945, nuclear magnetic resonance (NMR) spectroscopy, has become one of the foremost methods for molecular identification, for evaluating detailed molecular structures, for understanding conformations and for probing molecular dynamics. If the measurements are carried out under appropriate conditions, NMR spectroscopy can also be used for quantitative analysis. The technique is so powerful because it combines the provision of detailed information at the atomic level with the possibility of understanding whole molecule properties, such as diffusion in solution, and the ability to carry out quantitation. Although powerful in its own right, NMR spectroscopy can be regarded as complementary to other analytical chemical techniques. For example, it can provide information on substances with no ultraviolet (UV) chromophores, such as carbohydrates. It is a universal detector in that if the molecule under study contains NMR-active nuclei, these should be detectable. However, it is only possible to infer details of parts of the molecule that do not contain NMR-active nuclei, such as the presence of a sulfate conjugate of a drug metabolite, in which case the definitive loss of mass 80 seen in mass spectrometry (MS) is complementary.

The vast majority of NMR spectroscopic experiments are carried out in solution to identify the structures of small chemical molecules, including natural products, but a wealth of high–resolution applications is found in other areas, such as determination of the three–dimensional structures of proteins and the analyses of complex biological mixtures (e.g. biofluids). In addition, much effort has been devoted to solid–state NMR spectroscopy in which special techniques have to be used to overcome very broad NMR peaks and to recover useful chemical information. One application in this area of interest to pharmaceutical scientists is the determination of polymorphic forms of drugs in the solid state.

Finally, NMR spectra can be obtained from living humans and animals and *in vivo* NMR or magnetic resonance spectroscopy (MRS) has found a use in disease diagnosis. The same technology and principles lie behind magnetic resonance imaging (MRI), which is now widely available in hospitals for clinical diagnosis.

Sommaire

Theory

Atomic nuclei in a magnetic field

The phenomenon of NMR arises because the positively charged nuclei of certain atoms possess a quantised property called spin, denoted by a spin quantum number I. This spin is associated with a nuclear magnetic moment, also quantised, such that in a magnetic field it is possible for the nuclear magnetic moment to take up various orientations with respect to the field. Each orientation is associated with a discrete energy state and in the presence of the magnetic field these states have different energies. I can have an integer or half–integer value,

including zero. For nuclei with $I = 0$, there is no magnetic moment, which is the case when both the atomic number and the atomic weight are even, such as ^{12}C and ^{16}O. In general, there are $2I + 1$ energy states or levels and so for the simplest magnetic nuclei, with $I = ½$, there are just two levels. As a consequence of the differing energies of the states, and because of the Boltzmann distribution, the populations of spins in the states are not equal and an excess of nuclear spins will occur in the lower level. It is possible to induce transitions of nuclear spins between these levels by applying an oscillating frequency field; for commercially available NMR magnets, these transitions are in the radiofrequency region of the electromagnetic spectrum. There is a linear relationship between the magnitude of the nuclear magnetic moment and the observation frequency of the NMR phenomenon for a given magnetic field strength. There is also a linear relationship for a given nucleus between observation frequency and magnetic field strength. As is shown below, not all nuclei of a given atomic isotope have the same resonance frequency (known as the Larmor frequency), and hence the NMR phenomenon gives rise to a range of resonance frequencies that correspond to peaks in an NMR spectrum.

Many good textbooks describe the theory of NMR spectroscopy in more detail, a selection of which are listed in Further reading at the end of this chapter.

Important NMR-active nuclei

Although all nuclei have at least one isotope that is, in principle, NMR active, most NMR spectra are based on just a few nuclear types. There are several reasons for this. One is that nuclei with $I > ½$ have a property called a nuclear quadrupole moment that, in general, results in short lifetimes in the excited spin states and a rapid return to the low energy state, which gives very broad NMR lines. Secondly, many NMR-responsive nuclei exist at low natural abundances and so are difficult to detect without isotopic enrichment. Thirdly, the strength of the NMR response is related to the size of the nuclear magnetic moment, for which many nuclei have rather small values and so have low detectability. Finally, some nuclei, once excited to the upper level, are slow to relax back to the ground state, which must occur before another scan can be added. This then incurs a time penalty for acquiring the summed scans necessary to improve detection limits. Sometimes these difficulties of low sensitivity, low natural abundance and long relaxation times come together.

The principal nuclei of interest for pharmaceutical and biochemical studies are given in Table 1, together with their NMR properties. The common isotopes of carbon and oxygen, ^{12}C and ^{16}O, do not have magnetic moments and so do not give rise to NMR spectra. The ubiquitous ^{1}H nucleus, or proton, has one of the highest relative sensitivities, surpassed only by its radioactive isotope tritium, ^{3}H. The ^{13}C isotope is useful to characterise the carbon skeleton of organic molecules and, with a natural abundance of about 1.1%, the chance of finding two ^{13}C nuclei in a given molecule is only about 0.01%, which simplifies the spectra considerably because of the absence of coupling between carbon atoms. Many spectroscopic methods have been developed to allow the routine observation of

^{13}C NMR spectra of organic molecules. The ^{19}F nucleus is almost as sensitive as the ^{1}H nucleus (about 83%) in NMR terms, and ^{19}F NMR spectroscopy is used extensively in studies of the metabolism of fluorine–containing drugs. More limited use is made of other nuclei in pharmaceutical and biochemical research, and nuclei such as ^{15}N have been used extensively for protein–structure determination after isotope enrichment. The use of ^{31}P NMR spectroscopy is widespread in biochemistry and medicine as a means of investigating the various phosphorylated molecules important in biology, including many studies *in vivo*. Many other spin-½ nuclei, such as ^{29}Si, ^{119}Sn, ^{129}Xe, ^{195}Pt and ^{199}Hg, have found much use in specialist applications. Nuclei with $I >$½ are quadrupolar and, in general, give broad NMR lines, but in some cases useful information can be gleaned. Examples include ^{2}H NMR in liquid crystals, ^{14}N NMR in heterocyclic chemistry and ^{23}Na NMR studies of intra- and extracellular sodium ions.

Parameters from an NMR spectrum

Chemical shifts

As mentioned earlier, not all nuclei of a given isotope resonate at exactly the same frequency. This is because a given nucleus is surrounded by electrons which are also magnetic and, in the presence of the magnetic field, these provide a fluctuating magnetic field that opposes the main field of the NMR magnet. As a consequence, the nuclei are shielded from the main magnetic field and require a higher field to bring them to resonance and thus they can be considered to have higher Larmor frequencies. The degree of shielding depends on the electron distribution around the nucleus and hence on the chemical environment. Thus, interpretation of chemical shift values allows identification of molecular structural fragments. Chemical shifts are measured relative to that of a reference substance placed into the sample. For ^{1}H and ^{13}C shifts in organic solvents, this is tetramethylsilane (TMS). The chemical shift is then defined as δ(H) = (difference in the resonance frequency in Hz between the analyte and TMS) $\times 10^{6}$/(operating frequency of the spectrometer). Chemical shifts are thus quoted in ppm and are independent of the operating frequency of the spectrometer, which allows comparisons irrespective of magnetic field strength. For aqueous samples, an alternative reference compound is used, of which trimethylsilyl [2,2,3,3–^{2}H$_{4}$]-propionic acid sodium salt (TSP) is the most common example. The chemical shifts for TMS and TSP are set arbitrarily to zero. Typical ^{1}H and ^{13}C NMR chemical shifts of a variety of important molecular fragments are shown in Tables 2 and 3, respectively.

Fragment[a,b]	$\delta(^{13}C)$	Multiplicity[c]
$CH_3.C$; $CH_3,C=C$	0–30	q
$CH_3.S$	7–20	q
$C.CH_2.C$	10–70	t
$CH.(C)_3$	18–68	d
$CH_3.COX$	18–30	q
Acetylenic C	20–100	d, s
$CH_3.N$	25–50	q
$C.CH_2.COX$	25–60	t
$C.(C)_4$	30–80	s
$(C)_2.CH.COX$	35–75	d
$C.CH_2.N$	35–75	t
$(C)_2.CH.N$	40–90	d
$(C)_3.C.COX$	45–100	s
CH_3O	50–62	q
$(C)_3.C.N$	50–100	s
$C.CH_2O$	57–90	t
$(C)_2.CH.O$	65–100	d
$(C)_3.C.OC$	70–110	s
$CH_2=C$	80–135	t
Aromatic CH	80–140	d
C=C	80–160	d, s
OCO	85–110	t, d, s
Aromatic C (not CH)	90–160	s
Nitrile	115–125	s
C.COX	165–180	s
C.COOH	175–185	s
C.CHO	195–205	d
C.CO.C	205–220	s
C.CS.C	220–240	s

Table 3. The range of ^{13}C NMR chemical shifts for carbon atoms in various molecular fragments

[a] Designated carbon in bold face. [b] X = C, O, N. [c] Indicates peak multiplicity in a [1]H-coupled spectrum; s, singlet; d, doublet; t, triplet; q, quartet.

Indirect (J) spin–spin coupling

The resonance lines of individual nuclei can show further splitting because of indirect spin–spin coupling. This is given the symbol J, is measured in hertz and is independent of the observation frequency. Such spin coupling arises from a magnetic interaction between NMR-active nuclei and is transmitted via the intervening electrons, hence the term 'indirect'. Coupling is only observed within a molecule. Thus for two spin-½ nuclei, such as protons, the resonance line for each proton is split into a doublet, the two lines corresponding to the two possible orientations of the adjacent proton relative to the magnetic field. For extended coupling chains, each component of a doublet can be split further into doublets of doublets and so on. If a given proton is adjacent to two equivalent other protons (as in a CH_2 group) then, of the four possible orientations of the two protons, two of them are identical (up/down is the same as down/up) and a 1:2:1 triplet results. For such 'first–order' systems, the multiplicity can be deduced on the basis of Pascal's triangle according to the number of equivalent coupled nuclei. In situations where the chemical shift difference between the protons is large compared to the J-coupling, this simple rule applies. For situations where the chemical shift between coupled partners is not large compared to the magnitude of the coupling constant ($δ/J<10$), or in symmetrical molecules, more complex rules have to be applied and sometimes the only way to interpret a spectrum is via a computer simulation. For [1]H–[1]H interactions, the coupling does not normally extend beyond three bonds, with four–bond couplings being quite small, if resolvable.

Three–bond [1]H–[1]H couplings provide valuable information on the dihedral angles between C–H vectors through an empirical equation known as the Karplus equation. Typically, for CH–CH fragments, if the CH vectors have a dihedral angle of 180°, the coupling is of the order of 10 Hz; for 90° it is close to zero, and for 0° it is about 6 Hz. In olefinic systems, the three–bond coupling across a C=C double bond is about 6 to 10 Hz for a *cis* arrangement and 12 to 16 Hz for a *trans* arrangement. All of these values are modified by the presence of substituents with varying electronegativities. Hence the J-coupling is a valuable parameter for distinguishing between isomers and for measuring molecular conformations. Compilations of coupling constants have been made and empirical models for calculating them in various conformations have been proposed (Pretsch et al. 1989).

Peak areas

If the NMR data are acquired under conditions in which each scan is acquired on a spin system at equilibrium, the areas under the NMR peaks are directly proportional to the number of nuclei contributing to that peak and to the concentration of the molecule in the sample. If an internal standard of known concentration is added to the sample, absolute concentrations can be determined.

Relaxation times

Two times define how fast a nuclear spin interacts with the rest of the sample as a whole (known as the lattice) and how nuclear spins interact with each other in a pair–wise fashion. These are designated T_1 and T_2. T_1 is known as the spin–lattice or longitudinal relaxation time, and is the characteristic time for the exponential process of nuclear spins that reach equilibrium populations in the spin states. For small molecules in mobile solutions, ^{1}H T_1 values are usually in the range of 1 to 10 s. The other relaxation time is known as T_2, the spin–spin or transverse relaxation time, and is related to the rate of spin dephasing caused by spin–spin flips. For small molecules in free solution $T_1 = T_2$. However, macromolecules and exchanging species have short T_2 times, typically in the range 10 to 100 ms, even though T_1 may be much longer. The difference in values of T_2 between small molecules and macromolecules can be used to edit NMR spectra.

Diffusion coefficients

The molecular self–diffusion coefficient is a whole molecule property that does not normally appear in NMR spectra. However, it is a valuable measure of molecular mobility and in free solution is related directly to molecular size. It is possible to measure diffusion coefficients using a specially designed NMR experiment, which includes the application of magnetic field gradients.

Direct (dipolar) spin–spin coupling

Another important interaction in NMR spectroscopy is called the dipolar coupling. This is a direct magnetic interaction between nuclei through space, not through bonds, as for *J*-coupling; it is proportional to the inverse cube of the internuclear distance. This dipolar coupling can be several orders of magnitude larger than *J* couplings. However, it is averaged to zero in isotropic liquids, but in solids is largely responsible for the observed very broad resonance bands. In semi–solids, such as tissues, the dipolar couplings between nuclei are partially averaged out by the considerable molecular freedom and the residual couplings, and hence the line broadening can be removed by the technique of magic–angle–spinning (MAS). However, for molecules tumbling in solution, the fluctuating dipolar interaction is an important relaxation mechanism and, because of the distance dependence involved in its definition, it can be used to interpret nuclear Overhauser enhancements (NOEs) in terms of internuclear distances, and hence provide molecular structural information.

Instrumentation

Practical aspects of 1H NMR spectroscopy

NMR spectroscopy is used mostly to identify molecular structures, usually as pure compounds, although, because it is such a high–resolution technique, it is possible to characterise individual components in chemical and biological mixtures. However, sometimes it is preferable to carry out simple solid–phase extractions before measuring NMR spectra, and it is also possible to couple high–performance liquid chromatography (HPLC) directly to NMR spectrometers (see below). One example of this use has been to characterise components in combinatorial chemistry syntheses. Of course, it is increasingly difficult to identify minor components (e.g. less than 5 mol %) in mixtures, in which case concentration by solid–phase extraction may improve detection. Nevertheless, it is possible to quantify lower levels of impurities given a clearly resolved NMR resonance for the impurity. Each of the two natural–abundance ^{13}C satellite signals of the main component (ca. 0.55 mol %) provides a useful quantification standard. Thus, impurity detection limits depend on the resolution of a suitable NMR resonance, but a detection level of 0.1 mol % is not unreasonable.

Almost all spectrometers used for chemical structural studies are now based on a superconducting magnet with an operating frequency for ^{1}H observation that ranges from 200 to 900 MHz. Both resolution and detection are improved with increased magnetic field strength. For routine chemical studies in which compound availability is not limited, 200, 250, 270 and 300 MHz observation is quite usual. Increasingly, laboratories are using 400 MHz NMR spectroscopy as a standard for chemical studies, but 500 MHz and 600 MHz systems are now relatively common (in 2003). Higher observation frequencies are available (700, 750, 800 and 900 MHz) and are used for specialised research studies, mainly in the field of protein–structure investigation and for complex biological mixtures such as biofluids. Historically, and still generally, a sample for high–resolution NMR spectroscopy is placed in an expensive, precision–machined 5 mm glass tube sealed with a plastic cap, which is inserted into the detector (known as the probe) of an NMR spectrometer. Other tube diameters are also used, including 10 mm for larger samples and 3 mm when sample quantity is limited. More recently, robotic systems to inject samples directly into a flow probe using individual sample vials or 96–well plates have been developed.

Sample preparation

As spectrometers have become more sensitive, it is now possible to provide high–quality ^{1}H NMR spectra for small organic molecules on less than 1 mg of sample using widely available 400 MHz instruments. In addition, although they are still not used routinely (in 2003), a new generation of NMR detectors cooled with liquid helium

(cryoprobes) provides about a four–fold improvement in sensitivity by a corresponding lowering of the level of thermal noise in the electronic circuits in the probe and preamplifier. For organic chemical samples, a dilute solution in a fully deuterated solvent is usually used. Typical solvents are chloroform–d and dimethylsulfoxide–d_6, but occasionally more expensive alternatives have to be used, such as dimethylformamide–d_7, cyclohexane–d_{12} or acetone–d_6. A small amount of the volatile TMS is added as a chemical–shift reference. It is also possible to use D_2O as a solvent for water–soluble materials, in which case TSP is usually used for a chemical–shift reference. For biological samples, such as biofluids, which largely comprise H_2O, it is possible to freeze–dry these and reconstitute them into D_2O, but this risks the loss of volatile components. Alternatively, it is possible simply to add a small amount of D_2O and suppress the huge water NMR resonance by using one of a number of solvent suppression NMR pulse sequences. These either suppress the water signal selectively or have the effect of exciting the whole of the NMR spectrum except the water peak. The deuterated solvent, or the added D_2O, provides a ^2H NMR signal for the spectrometer to use as a field–frequency lock signal, so that all scans are co-registered exactly.

Acquisition of NMR data

The radiofrequency circuits of the spectrometer are tuned and matched to the sample so that the reflected power is minimised and then the magnetic field homogeneity optimised by adjusting currents in small ancillary coils around the sample, either manually or under computer control. This is a process known as 'shimming' and yields the sharpest NMR peaks and hence the best peak heights. The historical method of acquiring an NMR spectrum by slowly sweeping either the magnetic field or the radiofrequency (the continuous wave or CW method) is now obsolete and almost all spectra are obtained using the pulse-Fourier transform (FT) approach. In this method, all nuclei are excited simultaneously using a short, powerful radiofrequency pulse that covers all frequencies of interest. As the nuclei relax back to equilibrium, an oscillating, decaying voltage known as a free–induction decay (FID) is induced in the receiver coil. This signal is converted into a digital form using an analogue–to–digital converter (ADC) and stored in a computer memory as many thousands of data points. The total data–acquisition time for a single scan is typically in the range of 0.1 to 4 s.

Data processing

This amplitude–against–time signal is treated as a series of decaying cosine waves and is converted into an amplitude–against–frequency spectrum by the mathematical process of Fourier transformation. Multiple scans are usually co–added to improve the signal–to–noise ratio (SNR), since this is proportional to the square root of the number of scans. The resultant FID is usually multiplied by some function to improve SNR or reduce the spectral line widths, and hence increase resolution. For example, it is possible to enhance SNR by multiplying the FID by a negative exponential, exp(–At), where t is the acquisition time and A is a positive constant, but this also has the

effect of increasing line widths. Alternatively, it is possible to enhance resolution by multiplying by a positive exponential, exp($+At$), but this has a large deleterious impact on SNR. In addition, if the acquisition time is too short, the FID is truncated before it has decayed into the noise, and artefacts are seen as oscillations on each side of the NMR peaks. A good compromise to overcome truncation artefacts and to improve resolution without too large a compromise on SNR is to carry out a Lorentzian–Gaussian transformation. This is achieved by multiplying

by a positive exponential and a negative exponential squared, exp($+At - Bt^2$). The value of A is chosen to match the experimental line width and B is selected to give a Gaussian line shape of the desired reduced width. Since the delays in the electronics are frequency dependent, not all peaks will appear phased (upright). This has to be corrected either manually or automatically.

System tests

A number of standard system tests are used to define the performance of NMR spectrometers to meet the specifications guaranteed by the manufacturers.

A good line shape is an important requirement of modern NMR spectroscopy. The line–shape test is often called the 'hump' test and comprises an examination of the ^{1}H NMR peak from 10% chloroform in acetone–d_6, degassed and sealed. In addition to an acceptable value for the peak width at half height (50%), a good line shape is also defined by the width at the height of the ^{13}C satellites (0.55%) and at one–fifth of this height (0.11%). NMR peaks should have a Lorentzian line shape and therefore the widths at 0.55% height and 0.11% height should be 13.5 and 30 times the width at 50% height; deviations from these values should be avoided. The spectrum is measured in both spinning and non–spinning conditions.

The standard resolution test for ^{1}H NMR spectroscopy is the measurement of the spectrum of a sample of 10% 1,2–dichlorobenzene in acetone–d_6, degassed and sealed. This molecule gives a complex second–order spectrum, but the eighth line from the left is used and the peak width at half height is measured at a spinning rate of 20 Hz. No weighting function is applied to the FID before Fourier transformation and a typical value for a 200 MHz spectrometer is around 0.07 Hz.

The sensitivity test for ^{1}H NMR spectroscopy is based on a degassed and sealed sample of 0.1% ethylbenzene in CDCl$_3$. A single 90° pulse is used with a spectral width of 10 ppm and a 1 Hz line–broadening function is applied to the FID. The region of the spectrum between $\delta 3$ and $\delta 5$ is expanded vertically, typically by a factor of 16, to show the full extent of the noise band, and this is compared to the height of the peak from the methylene group. The SNR is defined as the peak height multiplied by 2.5 divided by the peak–to–peak noise value in that specified region.

For quantitative measurements, pharmacopoeial requirements are to verify the repeatability of the integrator response using a 5% solution of ethylbenzene in $CDCl_3$. Five successive scans of the protons of the phenyl and ethyl groups are carried out and the mean of the values obtained. None of the values should differ by more than 2.5% from the mean.

There are two main tests of ^{13}C NMR performance. The American Society for Testing and Materials (ASTM) approach uses a solution of 60% benzene–d_6 in 1,4–dioxane, degassed and sealed, and tests only the ^{13}C performance. The other test uses a solution of 10% ethylbenzene in $CDCl_3$, degassed and sealed, and tests both the ^{13}C sensitivity and the ^{1}H decoupling efficiency. For the ASTM test, a single 90° pulse is applied to the sample with the ^{1}H decoupler switched off. The FID is zero–filled to 65 536 (64k) points, weighted by a line broadening function of 3.5 Hz, and the full spectrum is plotted. The region between δ_c80 and δ_c120 is expanded vertically by a factor of four to allow good visualisation of the noise band, and this height is compared with that from the triplet signal from the C_6D_6. When using the ethylbenzene test sample, a single 90° pulse is applied to the sample, the FID is zero–filled to 65 536 (64k) points, a line broadening of 0.3 Hz is applied and the full spectrum is plotted. The noise height is taken, as before, from the region between δ_c80 and δ_c120 and compared to the peak height of the aromatic CH ^{13}C NMR signals.

Two–dimensional NMR spectroscopy

One–dimensional NMR spectroscopy gives rise to one frequency axis. As mentioned above, this is achieved by Fourier transformation of a FID obtained as the response after a radiofrequency pulse or pulse sequence. By incorporating a second incremented time period into an NMR pulse sequence, it is possible to carry out Fourier transformation with respect to both time periods to give two frequency axes that result in two–dimensional NMR spectra. This spreads out the NMR spectrum in a variety of ways, and so makes interpretation simpler. Extension to the use of three or more time periods leads to the concept of three–dimensional (or more) FTNMR spectroscopy.

All two–dimensional NMR experiments can be considered to have a preparation period, which might simply be a relaxation delay or could be some way of preparing the spin system for further manipulation. Next, the evolution period contains the first variable time period, t_1, which is incremented in equal time steps. Then it is possible to include a mixing period to allow the various spins to interact in some non–equilibrium condition. However, not all pulse sequences use this period. This is followed by the detection of the FID during the second time period, t_2.

To see how this works in practice, it is possible to consider one of the simplest two–dimensional pulse sequences. This is correlation spectroscopy (or COSY), which uses two radiofrequency pulses with an incrementable time delay

between them. The time–domain data comprise a series of FIDs for a range of inter–pulse delays. After the first Fourier transformation, the data consist of an array of spectra on a frequency axis, each of which corresponds to a given inter–pulse delay, but with the phase of the peaks modulated as a result of the various values of t_1. After a second Fourier transformation, the data array comprises NMR intensity as a function of two frequency axes and is shown as a three–dimensional plot. This is quite difficult to interpret, especially if it is complicated by many signals. The data are visualised more easily as a contour plot 'looking down' on the peaks. In the COSY spectrum the peaks of the normal spectrum appear along the diagonal and the presence of off–diagonal cross–peaks indicates two nuclei that have a common indirect spin–spin (J) coupling. Hence, this experiment correlates nuclei that are spin–coupled together and are thus close together in bond terms. The frequency axes are labelled F1 and F2, deriving from t_1 and t_2, respectively.

There are many two–dimensional NMR pulse sequences in the literature, most of which are for very specialised applications and are used only by their inventors. However, a few very useful experiments are used widely for molecular structural studies. Two–dimensional NMR experiments can be classified into two types – resolved and correlation. The former resolve a particular type of NMR parameter into the second dimension. For high–resolution solution–state studies, this type is restricted largely to the J-resolved experiment (JRES), which separates the chemical shifts and spin–coupling multiplets into the two dimensions. This can be carried out in a homonuclear mode for ^{1}H chemical shifts resolved from ^{1}H-^{1}H couplings and in a heteronuclear mode in which, usually, ^{13}C chemical shifts are resolved from ^{1}H-^{13}C couplings.

Most two-dimensional NMR experiments are of the correlation type, which includes the COSY experiment that provides a correlation between ^{1}H chemical shifts that have a direct J coupling. The total correlation spectroscopy (TOCSY) pulse sequence provides a correlation as off–diagonal peaks, also based on the J coupling, but now along an unbroken chain of couplings. The nuclear Overhauser enhancement spectroscopy (NOESY) experiment provides a correlation between ^{1}H chemical shifts connected by an NOE. This experiment can also be used to investigate spin systems of molecules undergoing chemical exchange.

It is possible to obtain heteronuclear correlation two-dimensional NMR experiments, usually the correlation of ^{13}C and ^{1}H chemical shifts, although the use of other nuclei such as ^{15}N or ^{31}P is possible. Direct detection of the heteronucleus is possible and historically was the main approach. However, this is usually a low–sensitivity experiment because of the lower inherent sensitivity, or natural abundance, or both, of most heteronuclei compared to ^{1}H. To overcome this problem, experiments have been devised that use indirect detection of the heteronuclei attached to the protons (see the next section on ^{13}C NMR spectroscopy).

A summary of the principal useful two-dimensional NMR pulse sequences is given in Table 4. This also lists the typical length of time required for data acquisition and the quantity of material required for each experiment.

Experiment	F2	F1	NMR interaction	Information	Time/quantity[a]	Comments
COSY	δ_H	δ_H	$^2J_{HH}$ and $^3J_{HH}$	H–H coupling connectivity	0.25/1	Easy
DQF-COSY	δ_H	δ_H	$^2J_{HH}$ and $^3J_{HH}$	H–H coupling connectivity	0.25/1	Easy
TOCSY	δ_H	δ_H	$^2J_{HH}$ and $^3J_{HH}$	All H in a spin system	0.25/1	Easy
NOESY	δ_H	δ_H	H–H dipolar	r_{HH}, conformation	5/5	10–100 times weaker than COSY
ROESY	δ_H	δ_H	H–H dipolar	r_{HH}, conformation	5/5	10–100 times weaker than COSY
HETCOR	δ_C	δ_H	$^1J_{CH}$	C–H	5/10	^{13}C detected
HMQC	δ_H	δ_C	$^1J_{CH}$	C–H	1/5	^{1}H detected
HSQC	δ_H	δ_C	$^1J_{CH}$	C–H	1/5	^{1}H detected
HMBC	δ_H	δ_C	$^2J_{CH}$ and $^3J_{CH}$	C–C–H and C–C–C–H	3/5	^{1}H detected
JRES (homo)	δ_H	J_{HH}	J_{HH}	Measurement of J_{HH}	0.25/1	Easy
JRES (hetero)	δ_C	J_{CH}	J_{CH}	Measurement of J_{CH} and number of attached Hs	5/10	Moderate
INADEQUATE	δ_C	δ_{Ca} + δ_{Cb}	$^1J_{CC}$	C–C connectivity	15/100	1 in 10^4 molecules, difficult

Table 4. Two–dimensional NMR experiments useful for small molecules

[a] Time in hours and quantity in mg for a molecule of M_r 500 on a 400 MHz NMR spectrometer.

^{13}C NMR spectroscopy

In addition, much molecular structural information is available by measuring the ^{13}C NMR spectrum of organic molecules at natural abundance. As this nucleus is less abundant and inherently less sensitive than ^{1}H, larger amounts of sample are needed (typically 10 mg for a molecular weight 250 material at a ^{13}C observation frequency of 125 MHz, which corresponds to 400 MHz for ^{1}H measurement) and usually more FIDs need to be accumulated to build up the SNR. However, some factors alleviate these difficulties. These include the broadband decoupling of all ^{1}H–^{13}C coupling patterns using special pulse sequences, so that each carbon signal is a singlet. This also has the effect of boosting the SNR through the NOE from the ^{1}H spins, this being most effective for CH, CH_2 and CH_3 carbons. One disadvantage of ^{13}C NMR spectroscopy is that some carbons, particularly quaternary carbons (which have only at best weak NOE) also have long relaxation times and are saturated easily and thus give weak peaks. For these reasons, quantitative studies based on ^{13}C resonances are more prone to error than those based on ^{1}H resonances. The undesirable long ^{13}C relaxation times can be shortened by the addition of a paramagnetic agent, such as the acetylacetonate complex of chromium, $Cr(acac)_3$. The identification of the carbon resonances as C, CH, CH_2 or CH_3 can be achieved through off–resonance decoupling of the protons, which leaves a residual characteristic coupling pattern of singlet, doublet, triplet and quartet, respectively (see Table 3). Alternatively, a number of NMR pulse sequences, such as the attached proton test (APT) or the distortionless enhancement by polarisation transfer (DEPT) approach can be applied to give the same information.

Directly observed ^{13}C NMR spectra have now been surpassed largely by a technique that involves indirect ^{13}C detection through ^{1}H NMR spectroscopy. This has come about through the advent of inverse–geometry NMR probes and the development of NMR pulse sequences that allow transfer of magnetisation between ^{13}C and ^{1}H. In an inverse experiment, usually conducted in the two–dimensional mode, ^{1}H magnetisation is generated, and then transferred to ^{13}C , the ^{1}H and ^{13}C spins are allowed to interact for a variable time based upon the one–bond ^{1}H–^{13}C coupling constant of about 125 to 160 Hz, and then the magnetisation is transferred back to ^{1}H for detection. The resultant two–dimensional NMR spectrum has ^{1}H chemical shifts on the detection axis (F2), ^{13}C chemical shifts on the indirect axis (F1) and a peak at the intersection of the ^{1}H and ^{13}C chemical shifts for CH, CH_2 and CH_3 groups. A number of pulse sequences achieve this correlation, the two best known being heteronuclear multiple quantum

coherence (HMQC) and heteronuclear single quantum coherence (HSQC). Quaternary carbons do not appear in such a spectrum and have to be detected in a separate experiment based on a longer range ^{1}H–^{13}C spin coupling, usually that over three bonds with a typical value of 6 to 10 Hz. For long–range coupling correlation, the experiment usually used is heteronuclear multiple bond correlation (HMBC). Full details of these methods are given Croasmun and Carlson (1994) especially, and in the books listed in Further reading. Such inverse–detected correlations have rendered the conventional ^{13}C-detected two–dimensional methods such as HETCOR obsolete.

^{13}C has a natural abundance of only around 1.1%, so ^{13}C–^{13}C correlation spectroscopy has to rely on the observation of only 0.01% of the sample, and at the same time suppress the ^{13}C resonances of all singly labelled molecules that are 100 times larger. However, such a method does exist, given the rather ridiculous acronym INADEQUATE, and provides a correlation between carbons separated by one bond, the ^{13}C–^{13}C coupling constant in this case being around 40 Hz. Typical sample requirements for two–dimensional NMR studies that involve ^{13}C are given in Table 4.

An example of molecular structure determination –ibuprofen

The various types of NMR spectra often used to confirm a molecular structure are illustrated by the example of the non–steroidal anti–inflammatory drug ibuprofen dissolved in dimethyl sulfoxide–d_6 (DMSO–d_6). Fig 1 shows the 600 MHz ^{1}H NMR spectrum of this substance; the molecular structure and numbering system is also shown, along with the assignments of the various NMR signals.

Figure 1. 600 MHz 1H NMR spectrum and molecular structure with numbering system of the non–steroidal anti–inflammatory drug ibuprofen dissolved in DMSO–d6. Assignments are as marked.

Thus the two doublets at δ0.85 and δ1.33 arise from the methyl groups with typical chemical shifts, that from CH$_3$(12,13) being of double intensity, with both showing spin coupling to a single proton. The resonance at δ1.80 comprises nine lines in a binomial intensity pattern because it arises from CH(11), which is coupled to eight adjacent hydrogens. The 1:3:3:1 quartet at δ3.61 is from CH(7). The doublet at δ2.40 is the resonance from CH$_2$(10) and arises from coupling to CH(11) exclusively. The small multiplet at δ2.49 is from the residual DMSO–d_5 in the DMSO–d_6 solvent and comprises a 1:2:3:2:1 quintet because the residual CHD$_2$ proton is coupled to two deuterium nuclei that have a spin quantum number of unity, each giving rise to a 1:1:1 splitting, and the two splittings overlap to give the resultant multiplet. Finally, the aromatic protons give rise to the multiplets at δ7.09 and δ7.17. These are not simple doublets, as can be seen by the partially resolved lines in the wings of the main peaks. These patterns arise as a consequence of the symmetry of the aromatic ring, with H2/H6 being chemically equivalent, but not magnetically equivalent (e.g. H2 will have a different spin coupling to H3 and H5). The broad

peak near $\delta 3.3$ arises from the residual water in the sample and solvent.

Figure 2. Two–dimensional 600 MHz 1H–1H COSY NMR spectrum and molecular structure of the non–steroidal anti–inflammatory drug ibuprofen dissolved in DMSO–d6. The spin–spin coupling connectivities for the CH2(10)–CH(11)–CH3(12,13) and CH(7)–CH3(8) spin systems are shown as dotted lines. Assignments are as marked.

The 600 MHz ^{1}H–^{1}H COSY spectrum of ibuprofen is shown in Fig 2. Dotted lines are used to indicate the off–diagonal peaks that connect two chemical shifts arising from spin–spin coupled protons. For example, the peak at $\delta 0.85$ from $CH_3(12,13)$ is connected by an off–diagonal peak to the chemical shift at $\delta 1.80$ arising from CH(11), which in turn is connected to the peak at the chemical shift of $\delta 2.40$ corresponding to $CH_2(10)$.

The 125 MHz ^{13}C NMR spectrum of ibuprofen in DMSO–d$_6$ is given in Fig 3. Since all spin–spin couplings between protons and ^{13}C are removed by decoupling, each non–equivalent carbon gives a single resonance peak. The assignments of the peaks are also given on Fig 3. Thus, the carboxyl carbon appears at 175.3 ppm and the two quaternary aromatic carbons are at 139.6 and 138.4 ppm; the assignments to C1 and C4, respectively, are based on a two–dimensional experiment, known as HMBC. This correlates ^{13}C and ^{1}H peaks via long–range spin–spin couplings (see the section on ^{13}C NMR spectroscopy). Next come the other aromatic carbons, C2/6 at 129.0 ppm and C3/C5 at 127.0 ppm. These peaks are more intense because each arises from two equivalent carbons, and also, because of the decoupling of the ^{1}H–^{13}C spin–spin interactions, each carbon experiences an NOE that boosts the intensity of its NMR resonance. The resonances from CH(7) and $CH_2(10)$ appear very close together, the assignments again being based on an HMBC spectrum. The 1:3:6:7:6:3:1 septet at 39.5 ppm arises from the carbons in the solvent DMSO–d$_6$ and the splitting is caused by ^{13}C–^{2}H spin–spin coupling, which is not affected by decoupling the ^{13}C–^{1}H interactions. The resonance for CH(11) is at 29.6 ppm and the methyl carbons appear at 22.1 ppm [$CH_3(12,13)$] and 18.4 ppm [$CH_3(8)$], one being twice as large as the other, as expected.

Figure 3. 125 MHz 13C NMR spectrum with broadband 1H decoupling, and molecular structure of the non–steroidal anti–inflammatory drug ibuprofen dissolved in DMSO–d6. Assignments are as marked.

To illustrate the use of indirectly detected ^{1}H–^{13}C two–dimensional NMR spectroscopy, the ^{1}H–^{13}C HSQC spectrum of ibuprofen is shown in Fig 4. This provides the connectivity between ^{13}C nuclei and directly attached ^{1}H nuclei. Hence, given the partial assignment of both the ^{1}H and ^{13}C NMR spectra based on inspection, it is often possible to complete the assignment of both spectra using this type of data. The two–dimensional contour plot shows a

number of peaks, each of which occurs at the intersection of the ^{1}H and ^{13}C NMR chemical shifts of a given CH, CH_2 or CH_3 group. Quaternary carbons do not appear in this spectrum because they have no directly attached hydrogens to provide the one–bond ^{1}H–^{13}C spin–spin coupling used for the correlation. For ease of interpretation, the ^{1}H and ^{13}C one–dimensional NMR spectra, assigned earlier in Figs 1 and 3, are plotted along the appropriate axes.

Figure 4. Two–dimensional 1H–13C HSQCNMR spectrum and molecular structure of the non–steroidal anti–inflammatory drug ibuprofen dissolved in DMSO–d6. A peak is seen at the intersection of the 1H and 13C NMR chemical shifts of each CH, CH2 and CH3 group. Assignments are as marked by the peaks on the axes.

Thus, in summary, all of the acquired NMR data are consistent with the known chemical structure of ibuprofen. To assign the proportions of the two enantiomers that may be present, it would be necessary to add a chiral chemical shift reagent, as described earlier. Even then, the absolute stereochemistry of the two forms could not be determined from such a spectrum, and recourse would need to be made to other techniques, such as circular dichroism.

Quantitative analysis

Quantification is possible using NMR spectroscopy. In this case, it is necessary to ensure that, for the summation of successive scans, the spin system is in equilibrium at the start of each scan. This is achieved by waiting five times the longest T_1 for the signals of interest between scans. This ensures that 99.3% of the signal is recovered. This is reasonable for ^{1}H NMR spectroscopy, in which T_1 values are usually less than 4 to 5 s, but for ^{13}C NMR spectroscopy (see above), T_1 can be much longer. As shown for ^{13}C NMR spectroscopy, the peak areas are, in general, not related quantitatively to concentration, and steps have to be taken to overcome this. This can be achieved by adding a relaxation agent and/or by modifying the pulse sequence to remove the NOE, which can affect signal intensity.

Quantification is carried out by measuring peak areas in the usual fashion, either by adding a known amount of a pure internal standard or by the method of standard addition. For a peak from an analyte with relative molecular mass M_{rA}, arising from N_A equivalent nuclei having an area I_A, and for a corresponding peak from an internal standard with relative molecular mass M_{rS}, from N_S nuclei with area I_S, the mass of the analyte in the sample is given by:

where m_s is the mass of the standard in the sample. Peak areas can be measured by integration, that is by summing the intensity of data points over a defined region of a peak or by curve fitting to the peak shape.

The two enantiomers of a chiral compound give identical NMR spectra in a non–chiral solvent because the magnetic environments of the nuclei are identical in both isomers. However, addition of a chiral paramagnetic shift reagent or use of a chiral solvent allows resolution of the NMR spectra of the two forms and hence integration of peak areas and measurement of the enantiomer proportions.

Directly–coupled HPLC–NMR–MS

The identification of components in complex mixtures has been rendered much more efficient by the ability to couple HPLC directly with NMR spectroscopy. The on–line coupling of HPLC with MS has already proved to be of immense value, and the complementary structural capability of NMR spectroscopy has meant that complete molecular identification can now be very fast. Most applications to date have focused on the identification of drug metabolites in biofluids and similar matrices, but there are also applications in drug impurity analysis and drug degradation profiles. The general scheme for HPLC–NMR–MS is shown in Fig 5, which indicates that the typical configuration is to have the HPLC apparatus situated on the bench as normal. After the column and UV (possibly diode–array) detection the eluate is split, usually in the ratio of 95:5, to the NMR and mass spectrometers, respectively. The length of the capillary lines to these spectrometers can be adjusted so that either the NMR spectrometer or MS detects the eluate first. For example, the MS could be put first to identify characteristic ions, such as those relating to the isotope patterns of chlorine or bromine. Alternatively, the NMR spectrometer could be put first, for example set up to pick up well–characterised fingerprints, such as the presence of a ^{19}F NMR signal.

Figure 5. The experimental arrangement for directly coupled HPLC–NMR–MS. The bold lines indicate sample flow and the thinner lines are the electronic control and data signals.

Currently, five main options can be employed for HPLC–NMR, using either isocratic or gradient elution. The simplest of these is continuous–flow detection, in which NMR spectra are collected in real time as chromatographic peaks elute, but this is only practical when using ^{1}H or ^{19}F NMR for detection, unless isotopically enriched compounds are available. Where continuous–flow NMR detection is used for gradient elution, the NMR resonance positions of the solvent peaks shift with the changing solvent composition and, for effective solvent suppression, these solvent resonance frequencies must be determined as the chromatographic run proceeds.

If there is a method to detect retention time on–line (UV, MS or radioactivity), stop–flow HPLC–NMR can be carried out using all the usual techniques available for high–resolution NMR spectroscopy. In particular, these include

two–dimensional NMR experiments. In practice, it is possible to acquire NMR data on a number of peaks in a chromatogram using a series of stops during elution without on–column diffusion causing an unacceptable loss of chromatographic resolution.

There are two further categories of stop–flow experiment. First, and an increasingly popular approach, fractions eluted from the column can be stored in capillary loops for later off–line NMR study ('peak picking'). Second, the flow can be halted at short intervals during the passage of the eluting peak through the NMR flow cell ('time–slicing'), in a manner analogous to the use of a diode–array UV detector, to obtain spectra from various portions of the peak. This allows chromatographic peak purity to be estimated. Time slicing is most useful if the separation is poor, or if the compounds under study have weak or no UV chromophores, which makes it difficult to determine the retention times.

Fully automated analysis is now possible and, in this mode, automatic detection of UV peaks in the chromatogram based on predetermined time–windows or peak intensities is allowed under software control. The successful detection of each UV peak triggers the system to stop the flow at an appropriate time to isolate the peak in the NMR flow probe. This automatic NMR operation includes field homogeneity optimisation, setting and optimisation of all NMR acquisition parameters and the predefinition of the resultant SNR required in the spectrum. The measurement of two–dimensional NMR spectra can also be performed. With currently available commercial software, the automated run can be halted at any time with reversion to manual control if desired.

Usually, for reversed–phase HPLC separations, a gradient eluent system of acetonitrile and water is used. To reduce the solvent suppression demands, the aqueous phase is made up of D_2O, but to save on cost any organic solvents are used in their conventional protio forms. In some cases, it is preferable to use deuterated acetonitrile, for example, as this aids the detection of NMR peaks near the methyl peak of the solvent. The increase in cost is outweighed easily by the time saving in carrying out the study and the improvement in the chance of success. Apart from this, it is not usually necessary to make any compromises in a desired chromatographic procedure to accommodate the various types of HPLC–NMR experiment. The powerful structural elucidation capabilities of NMR spectroscopy often ensure that complete chromatographic separation is not necessary for full characterisation of the peak.

In the case of HPLC–NMR–MS experiments, there are some additional considerations. So far, the principal MS ionisation method used has been electrospray in either positive or negative ion mode (using either single quadrupole or ion–trap mass spectrometers), which places further constraints on the chromatographic solvent systems. When using HPLC–NMR the chromatography is often developed off–line from the NMR using non–deuterated solvents. It is not always simply a matter of replacing non–deuterated solvents with deuterated solvents to reproduce the chromatography for HPLC–NMR or HPLC–NMR–MS, as this can give rise to changes in retention times. For this reason it is standard practice to carry out an initial chromatographic run with a small injection volume (e.g. 10 µL) and then scale–up (e.g. 50 µL) for stop–flow NMR, once optimum conditions have

been established. By this means it is often possible to acquire valuable MS data that can be used to guide the selection of peaks for study by NMR spectroscopy. It is also possible to mix the eluent just prior to the mass spectrometer with a non–deuterated solvent to back–exchange any deuterium atoms in exchangeable situations (e.g. NH and OH groups) for hydrogens. Comparison of the MS data with and without D_2O present enables the number of exchangeable hydrogens in any compound to be determined. In addition, MS can be used to search for particular diagnostic groups or fragments, such as an increase in m/z 16 for Phase I hydroxylated drug metabolites or an increase of m/z 196 for a glucuronide.

At present, a number of different mass spectrometers have been used in HPLC–NMR–MS studies. The simplest approach is to use a quadrupole mass spectrometer that gives integer mass values for ions. Employment of an ion–trap design allows fragmentation of the initial ions to be carried out in tandem experiments, such as MS–MS, MS–MS–MS (i.e. MS^n), etc. The use of time–of–flight or FT mass spectrometers provides very high resolution that allows good predictions of empirical molecular formulae from accurate ion masses, including those from MS^n experiments. Finally, some studies using inductively coupled plasma MS have been carried out. This technique provides no molecular information, as the sample is reduced to atomic form in the plasma, but it yields atom–specific chromatographic detection, which can be valuable for identifying the retention times of molecules of interest, such as those that contain heavy atoms or halogens such as chlorine and bromine.

By far the largest body of work to date using HPLC–NMR and HPLC–NMR–MS is in the field of drug metabolism, in which the methodology has been used extensively to identify metabolites in studies from clinical trials that involve human subjects, in the investigation of model drugs in animals *in vivo* and also in *in vitro* systems, such as liver microsome incubations. Applications to compounds used clinically include paracetamol, ibuprofen, flurbiprofen, naproxen, antipyrine, tolfenamic acid, and studies on many compounds currently in development in which early information on metabolic products can be used to aid drug design.

Many marketed drugs are either natural products or are modifications of such substances. Hence considerable effort has been spent in isolating and characterising chemicals from natural sources so that they can be tested in a variety of biological screens. Often it is necessary to carry out laborious extraction and purification steps, so the advent of directly coupled HPLC–NMR–MS has been explored as an alternative technique for natural product identification. The use of HPLC–NMR, and other combined techniques such as HPLC–MS–MS, to identify natural products from plant sources has been reviewed by Wolfender et al. (1998; also see Lindon et al. 2000).

Examples of the use of NMR spectroscopy in the British

Pharmacopoeia

NMR spectroscopy has been incorporated into a number of tests in the 2002 edition of the British Pharmacopoeia. These include the assay for buserelin, a peptide analogue [5–oxoPro-His-Trp-Ser-Tyr-D-Ser- (*tert*-Bu)-Leu-Arg-Pro-NHEt]. In this case the ^{1}H NMR spectrum of the sample is compared visually to that of a reference NMR spectrum from the European Pharmacopoeia and must be 'qualitatively similar'. In another approach, the test for low molecular weight heparins relies on the use of a 75 MHz ^{13}C NMR spectrum, which is compared with that from a reference sample of the material. A similar approach is used to identify the antibiotic tobramycin, but the nucleus of observation is not specified.

Poloxamer is a synthetic block copolymer of ethylene oxide and propylene oxide and thus contains oxyethylene units ($-O-CH_2-CH_2-O-$) and oxypropylene units ($-O-CH.CH_3-CH_2-O-$) units. The characterisation of poloxamer includes the use of ^{1}H NMR spectroscopy to determine the ratio of the two types of sub–unit, based on integration of the ^{1}H NMR resonances at δ1.08 from the methyl group of the oxypropylene units and at δ3.2 from methylene groups from both types of sub–unit.

NMR spectroscopy of intact biofluids and tissues

High resolution one- and multidimensional NMR spectra measured on biological samples, such as biofluids, can be extremely complex, and contain many thousands of resonances. In many cases, visual inspection of such spectra releases only a small percentage of the information available in the data. For this reason there has been a move towards the use of computer–based methods to extract the maximum information from such complex spectra. Although statistical methods have been applied for many years to NMR data, the first truly concerted application of multivariate statistics to high–resolution NMR spectra in the biomedical area was the classification of ^{1}H NMR spectra from rat urine samples according to the type of organ toxin administered to the animals. This provided a fast non–invasive way to assess candidate drug toxicity (see Lindon et al. 2000).

In the post–genomic era, there is much emphasis on the study of the consequences of changes in gene expression (functional genomics), by examination of either the protein complement (proteomics) or the small molecule metabolite composition of a biological system (metabonomics) (Nicholson et al. 2002). The term metabonomics has been defined as 'the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification', to describe the study of the dynamic and time–dependent profile of metabolites within an organism and how they are altered by some biological process. Currently, the only effective way to extract meaningful metabonomic information is to combine high–resolution

NMR spectroscopy of biofluids with multivariate statistics methods, often known as pattern recognition (PR) Lindon et al. 2001.

Genetic, disease or foreign compound effects can result, in some cases, in changes in gene expression and/or changes in protein production, but all will, in principle, cause alterations to biochemical regulation and control. Thus, all drug–induced and other pathophysiological perturbations result at least in disturbances in the ratios and concentrations, binding or fluxes of endogenous biochemicals. In body fluids, metabolites are in dynamic equilibrium with those inside cells and tissues, and consequently abnormal cellular processes in tissues of the whole organism after a toxic insult or during a disease process are reflected in altered biofluid compositions.

In all cases the problem involves the detection and identification of low amounts of analytes in a very complex matrix with many potential interferences, all obscured by inter–subject biological variation. High–resolution ^{1}H NMR spectroscopy is particularly appropriate for investigating abnormal body fluid compositions (such as blood plasma, urine or bile), as a wide range of metabolites can be quantified simultaneously with no sample preparation and without preselection.

To illustrate the complexity of biofluid NMR spectra, Fig 6 shows the 800 MHz ^{1}H NMR spectrum of a control human urine sample, with successive horizontal and vertical expansions. It has been shown that each biofluid has its own characteristic fingerprint. NMR spectroscopy may also be used effectively to screen for abnormal metabolite profiles in tissue extracts or cell suspensions. More recently, it has been shown that the same approach can be used to investigate the metabolic composition of intact tissues using a technique called high–resolution magic angle spinning (MAS) ^{1}H NMR spectroscopy.

Figure 6. 800 MHz 1H NMR spectra of a typical control human urine sample showing successive horizontal and vertical expansions, and illustrating the complexity of the biochemical profile. The vertical arrow points to a very minor peak from the axial methyl group of a bile acid, a substance known to be elevated in certain types of hepatotoxicity. This demonstrates the ability of 1H NMR spectroscopy of biofluids to detect very subtle changes in minor components.

Computer–based PR and expert system approaches have to be used to interpret the NMR data on biofluids and tissues obtained in various experimental toxicity or disease states. Often these include the use of principal components analysis (PCA) to classify the samples according to their NMR spectral profiles. Each spectrum can be reduced to a set of intensity–based numbers that are characteristic of that spectrum, and these descriptors are used as input for the PCA. Principal components (PCs) are new variables created from linear combinations of the starting spectral descriptors, such that each PC is uncorrelated with all other PCs, and the first PC contains the largest part of the variance of the data set, with subsequent PCs containing correspondingly smaller amounts of variance. Thus, a plot of the first two or three PCs gives the 'best' representation in terms of variation in the data set in two or three dimensions. Such PC maps can be used to visualise inherent clustering behaviour of samples,

for example marked according to toxic mechanism. The position of a NMR spectrum from a biofluid or tissue sample on a map is determined purely by its metabolic profile, as opposed to any other independent knowledge of the class of the sample. In this simple metabonomic approach, a urine sample from an animal treated with, for example, a compound of toxic effect is mapped against a database of NMR spectra of samples from animals in a range of toxic classes and the proximity of the sample to other types of sample is determined.

However, in the real world, biochemical data are more complex as effects caused by genetic modification, disease, nutrition, lifestyle, drug therapy or drug toxicity develop and recover in real time, and hence there can be complex, time–related changes in the NMR spectra of biofluids. Also, it is more rigorous to compare the NMR-derived metabolic profile in the original multidimensional NMR metabonomic descriptor space. Hence, to develop automatic classification methods, it is necessary to develop more complex expert systems. Here, a 'training set' of NMR metabonomic data is used to construct a mathematical model that predicts correctly the class of each sample. This training set is tested with independent data (a 'validation set') to determine the robustness of the computer–based model. These models can be based on a range of different mathematical procedures, such as PCs, artificial neural networks and rule induction, and several types can be embedded in a single expert system. In general, the methods allow the quantitative description of the statistical boundaries that characterise each class of sample in terms of their metabolic profiles. Some methods also allow a level of probability to be placed on the classification and can distinguish samples that belong to none of the classes in the training set. Using such systems, a sample can be classified as belonging to a single class, to multiple classes or to no class. By building an exhaustive series of models it is possible to use such methods to provide classification probabilities for a wide range of sample types.