

Instrumental Platforms for NMR-based Metabolomics

ANTHONY C. DONA[†]

MRC-NIHR National Phenome Centre, Department of Surgery & Cancer,
Imperial College of London, South Kensington Campus SW7 2AZ, United
Kingdom

*E-mail: anthony.dona@sydney.edu.au

1.1 History of NMR Hardware Development

The basic function of an NMR spectrometer is to measure the frequency of the resonance of a given nucleus. After the first decade of discovery (1946–1955), the basic NMR relationship was established (eqn (1.1)), which suggests the resonance frequency of a nucleus (ω) to the magnetogyric ratio (γ , specific to nucleus type) and the external magnetic field (B).

$$\omega = \gamma B \quad (1.1)$$

At this stage it was thought that the frequency of a nucleus depended entirely on the strength of the magnetic field in which it was placed. In 1949–1950, however, observations from ^{19}F and ^{31}P showed variations in frequency

[†]Current address - Northern Medical School, Kolling Institute of Medical Research, The University of Sydney, Royal North Shore Hospital, St. Leonards, NSW 2065, Australia.

between the two types of nucleus beyond the still rather large experimental error. Furthermore, after developments in the stability and homogeneity of the magnetic field, two separate resonances were observed from the hydrogen atoms in ethanol¹ and separately for nitrogen atoms in different chemical environments. This phenomenon soon became known as 'chemical shift' as the frequency of a resonating nucleus is largely dependent on the local chemical environment surrounding the nucleus itself. Further improvements in the resolution allowed separate resonance lines to be observed from a single chemical resonance, allowing the discovery of the concepts of indirect spin-spin coupling or scalar coupling.² In the late 1950s and the into the 1960s the strength of the field increased to 100 MHz and commercial instruments were developed by Varian that maintained a constant relationship between the magnetic field and the radiofrequency (RF) applied so spectra could be recorded at a known scan rate. In this time the first ¹³C spectrum was recorded,³ made difficult at the time by the low natural abundance of ¹³C atoms (1.1%). Carbon spectroscopy really became popular with the advent of double resonance, where two RF fields are applied to a sample simultaneously in order to measure one spin system while the other is perturbed. From double resonance, applications such as spin decoupling experiments and nuclear Overhauser effect (nOe) were introduced to aid studies of molecular conformation.⁴

In 1966, Ernst and Anderson published work⁵ showing that Fourier Induction Decay (FID) following a short RF pulse was all that was necessary to produce a spectrum measuring a range of frequencies. Additionally, mini-computers were being developed to interface directly with the spectrometer, allowing Fourier Transform NMR (FT-NMR). These hardware advances revolutionised NMR spectroscopy, supporting the enhancement of sensitivity (which was NMR's main disadvantage compared with other spectroscopy techniques) and enabling exploitation of time-dependent NMR phenomenon, namely relaxation.

Since these major advances in NMR hardware, many pulse sequences and applications have developed over the last 40 years, making NMR spectroscopy an incredibly versatile tool in chemical and biochemical research areas. Magnetic Resonance Imaging (MRI) was also made possible by imposing magnetic field gradients across a sample *in vivo*. Two-dimensional and eventually three-dimensional NMR imaging was made possible with wide bore magnets (wide enough to fit animals or humans through) and the measure of a frequency across a spatial gradient.⁶ Techniques are also currently being produced to integrate spectroscopy with imaging to obtain localised spectra in living creatures.

These days, the hardware available for modern-day NMR measurements allows for routine acquisitions with relative ease, in small (metabolite, organically synthesised) and large (protein) molecules, in either purified solutions (for molecular structure elucidation) or complex mixtures (for solution composition elucidation).

1.2 Components of NMR Hardware

1.2.1 Magnet

The NMR magnet itself is generally considered the most important part of the NMR spectrometer and is commonly the most expensive piece of equipment for a standard NMR laboratory. Before the advent of superconducting magnets, iron-core electromagnets had enabled field strengths of 2.35 T to be reached. At this stage iron saturates, so a great effort was required to develop a magnet that could be housed in a liquid helium Dewar to obtain the temperatures required to cool a superconducting solenoid. Today, commercial NMR magnets are generally superconducting and range in field strength from 6 to 23.5 T. Other physical chemistry laboratories are developing larger magnets, not intended for commercial distribution, with magnetic field strengths of up to 45 T.⁷ With increasing strength, not only does the size of the magnet need to increase but technologies need to be developed to break ground in new magnet generation. As magnets increase in size (and so magnetic field strength) their resolution and sensitivity of frequencies improves, but their cost also increases substantially, meaning larger magnets are much less accessible.

Superconducting wires are generally made from Nb_3Sn or $(\text{NbTaTi})_3\text{Sn}$, which are wrapped hundreds of times into a coil making up a length of wire up to 100 km long. Few higher temperature superconducting materials are suitable for manufacturing in the quality and quantity required for NMR magnets, so very low temperature superconductors (~ 10 K) are usually constructed. The wire has a rectangular cross section allowing maximum current density and therefore maximum magnetic field strength. This coil is kept inside a large Dewar containing liquid helium, keeping the coil at superconducting temperatures, which is in turn surrounded by a liquid nitrogen reservoir acting as a buffer between the room temperature air and the liquid helium. A significant property of the wire is the maximum critical current (I_c), which is a function of the operational temperature (T) and magnetic field (B). If the critical value is reached there is a transition in the wire from a superconductive to a resistive state, which in turn generates heat. The heat propagates rapidly through the coil prompting the energy store to be converted to heat, which induces the helium store to boil extremely rapidly. The loss of the superconductive state is known as a 'quench' and magnetic coils are developed to avoid quenches at all costs. Modern day magnets have an additional coil outside the main coil, which is used to contain the strong magnetic field by cancelling (shielding) the stray field, restricting it to a relatively small area.

The cryostat, which is the vessel surrounding the magnetic coil, must also be designed to be insensitive to variations in the environment and other disturbances such as helium evaporation rate, ambient temperature and pressure, and the cryogen levels inside the cryostat. New technologies in this area are focused around either using the complete enthalpy stored in the helium

gas to enable further cooling of the system or otherwise recycling helium such that it is not lost to the atmosphere. These technologies are aimed at having the lowest possible helium consumption as it becomes a rarer and more expensive commodity.

NMR laboratories are often forced to compromise on magnetic field strength owing to available funding and space. Modern day magnets between 400 and 600 MHz are commonplace in metabolomics laboratories as they are now constructed to fit in a room without roof height modification. With modern day shielding, the footprint required for the magnetic field is not much larger than the magnet itself and so magnets of this size can essentially be lined up next to one another. Magnets of 600 MHz are produced much more readily than those of larger size and so the production cost is far cheaper and more reliable, making them the magnet size of choice for routine metabolomics studies. On the other hand, magnet sizes smaller than 400 MHz do not enable researchers to resolve important metabolite signatures in complex biofluids and so these are generally overlooked when purchasing a magnet for metabolomics purposes.

1.2.2 Shim Coils

Shim coils are a set of conducting coils used to adjust the homogeneity of a magnetic field. In the past, shimming (the process of optimising the homogeneity of the magnetic field) a magnet consisted of attaching thin metal shims in various positions around the permanent magnet. Coincidentally, the term 'shimming' is used to describe the modern day process of homogenising the magnetic field across which the samples nuclei frequencies are measured. Modern high-resolution spectrometers alter the current in various conducting coils, which surround the external magnetic field, to alter its homogeneity.

When a spectrometer is installed, the local environment can disturb the magnetic field. Iron constructs in the walls and floor of the surrounding building disturb the homogeneity of the spectrometer's magnetic field. Therefore, upon installation spectrometers need to be roughly shimmed, after the initial activation of the magnetic field, with regards to the external environment. Once relative homogeneity is achieved, relatively minor changes in the magnetic field as a result of variations in the sample, the tube thickness and movement of ferromagnetic materials around the magnet are corrected before sample acquisition. These minor changes in the magnetic field are adjusted by changing the current in one or more (of up to 40) small shim coils with various gradients along all three spatial axes. In high-resolution spectrometers the magnetic field often demands homogeneity better than 1 part per billion in a tube, which is generally less than a millilitre in volume.

1.2.3 Sample Probe

The probe is the part of the spectrometer that does a lot of the physical experimental work and laboratories can often gain significant improvements in spectral quality by upgrading their spectrometer's probe. The probe contains

the RF coils tuned to specific frequencies to excite particular nuclei and coils to detect the NMR signal. Pulsed Field Gradient coils, which allow for the application of field gradients, are also commonplace, allowing for the application of field-gradient pulses. The probe must also consist of the necessary hardware to measure and control the temperature of samples (a thermocouple device).

An important aspect of probe design is the size of the bore, which can alter to accommodate various sizes of tube. Small volume probes (3 mm) or nanotubes (1.7 mm) are able to give the greatest sensitivity, benefiting from the dramatic decrease in the diameter of the NMR detection coil. Some modern day probes are able to record good quality spectra from microlitre or even picolitre sample volumes. Nevertheless, in many cases larger volume bores (5 mm, 10 mm or wide bore) are necessary as the solubility or concentration of the sample is an issue. Larger volume bores are also recommended when measuring samples of higher viscosity or samples with micro-scale inhomogeneities (blood or emulsions), and are necessary when imaging small animals.

1.2.3.1 Radiofrequency Coils

Modern day probes are constructed with two coils to record NMR signal (known as observe coils). They are wrapped around one another such that there is an inner coil and an outer coil, allowing the probe to respond to different frequencies during the one experiment. This design also allows multiple nuclei to be excited during one pulse sequence. There are two main approaches to the design of these coils:

- A 'Broadband Observe Probe' is constructed with the inner coil tuned to a broadband nucleus, and so is optimised for maximum sensitivity for nuclei at lower frequency (^{13}C , ^{31}P , *etc.*).
- 'Inverse Probes' or 'Indirect Detection Probes' have the inner coil tuned to measure the frequency of ^1H atoms and so get maximum sensitivity for proton experiments with much lower sensitivity when observing lower frequency nuclei.

1.2.3.2 Cryoprobes

Sensitivity of detection has dramatically improved with the advent of cryoprobes. These probes are significantly more expensive in design but ensure 2–4 times better sensitivity than standard probes. Cryoprobes have achieved the single largest jump in sensitivity enhancement by probe development in the last few decades by cryogenically cooling probe detector coils and pre-amplifier coils. As long as these coils, along with the tuning and matching circuits, are maintained at low temperatures the noise generated owing to random thermal motion of electrons is kept at a minimum. The resistivity in metals based in the preamplifier and filters is also cooled, improving the level of noise generated owing to the probe's electronic components.

There are currently two types of cryoprobe available, the first is cooled by a closed cycle helium cooler system, which yields a signal-to-noise enhancement up to a factor of five. The second type is a liquid nitrogen-cooled system, which has the advantage of being able to be cooled down and warmed up relatively quickly, although is only able to provide a sensitivity enhancement of up to a factor of three. Either one of these probes is generally applied to compensate for the lack of sensitivity owing to magnet strength, or simply to enhance a laboratory's run time capability.

1.2.3.3 *Microprobes*

The development of microprobes (or microcoil probes) was driven by the pharmaceutical industry where detection of small sample amounts was required with a relatively high throughput rate. The heart of the problem lied in developing a system with increased sensitivity for small sample volumes. The solution was simply to reduce the diameter of the RF coils; microprobes are generally considered to have an RF coil of less than 1.6 mm inner diameter. Therefore, there are few differences between standard NMR probes and microcoil probes apart from the fundamental sample volume required. However, as the volume required is much less, the microprobes allow for many more molecular biological applications.

While regular volume probes are based on Helmholtz saddle coils (coils that lie parallel to the external magnetic field),⁸ microprobes of less than 45 μL require a detection coil, which is typically a solenoid, that is placed orthogonal to the external magnetic field.⁹ The sample insertion into the magnetic field therefore cannot be performed traditionally from the top of the magnet using a conventional NMR tube. Sample insertion into a solenoidal coil is generally controlled by a flow injection probe (Section 1.2.3.4). Solenoidal probe interfaces are advantageous as they have a sensitivity level about 2–3 times larger than a traditional coil and only need an order of magnitude 1–2 times less than traditional NMR probes.

1.2.3.4 *Flow Injection Probes*

Flow injection probes were historically the first approach to conducting high-throughput NMR. Physically a flow probe contains a glass chamber within the RF coils with a total volume of between 1 μL and 120 μL . Flow injection probes with volumes less than 10 μL are considered to be 'micro-coil' flow probes (Section 1.2.3.3). The probe contains an inlet for solution to enter the chamber and an outlet for the sample to be evacuated and rinsed out. The probes can either be 'stop-flow', which means during the time of acquisition the sample is motionless in the probe's chamber, or alternately the probe can be 'on-flow' where the sample is flowing during the short analytical measurement of the sample by NMR. When using an on-flow probe, though, the shim quality is compromised owing to the flowing solution making it difficult to homogenise the magnetic field. Therefore, the results from

an on-flow probe are generally less resolved than a stop-flow probe, although the throughput is higher. Furthermore, the concentration of the metabolites of interest in the flowing sample needs to be considered before a measurement and optimisation of the methodology for each sample type is almost always necessary.

All flow techniques can be classified into one of two further separate categories. The first simply uses flow injection as a means to transport an untreated sample into the NMR analysis coil (almost always performed in a stop-flow environment to better distinguish samples and improve the shim quality), simplifying automation and speeding up throughput. The second category subjects the sample to a separation technique such as chromatography or electrophoresis before NMR analysis and can be performed in either stop-flow (separating fractions) or on-flow (continuous analysis of the probe's contents, which can be later processed separately). Flow injection probes have had many applications over the last couple of decades and more recently have been successfully applied to metabolic profiling of biofluids.¹⁰

Flow injection probes were invented at the turn of the century and have since been applied in multiple applications relating to the pharmaceutical industry and complex mixture analysis. Some examples of this are:

- (1) NMR screening applications to discover ligands with a high affinity to bind to proteins with various medical relevancies. Dubbed SAR-by-NMR, organic molecules are screened to develop an understanding of proximal subsites of proteins.¹¹ The molecules are either used themselves or as base molecules to develop high-affinity ligands.¹² The application is commonplace in the pharmaceutical industry and is a great example of a successful flow probe application as sensitivity and throughput are integral.
- (2) A high-throughput method has also been developed using microcoil probes for the production and analysis of large natural product libraries for drug discovery.¹³ High-performance liquid chromatography (HPLC) is coupled online with flow injection NMR to rapidly isolate and characterise bioactive products from plant roots, stems, leaves, flowers or fruits. Developing libraries of naturally occurring molecules through the use of HPLC-NMR is considered a technique as valuable to developing drug leads as combinatorial chemistry.
- (3) Identification of metabolite signals in complex mixtures is classically considered more robust when multiple forms of spectral information are acquired on a given sample or fraction. Alphanumeric metabolite identification metrics have been proposed recently to rank identification confidence based on the types of spectral data and comparisons to authentic standards.¹⁴ As a result of metabolic sciences seeking more robust metabolite identification, the direct coupling of HPLC with NMR has been extended into a triple hyphenated online technique enabling UPLC-NMR-MS.¹⁵ The technique splits the eluent after HPLC separation with conventional detection, with 95% of the eluent

flowing into an NMR flow injection probe and the other 5% directed to the inlet of an ion-trap multipole mass spectrometer. Owing to the technique's online approach, the development of automated routines and high-throughput applications is feasible. The approach initially allowed for the unequivocal identification of phenylacetylglutamine in human urine, an endogenous metabolite not previously identified by proton NMR owing to extensive overlapping with other resonances. This research underlined the technology's scope for success in metabolic profiling.

1.2.3.5 Magic-angle Spinning Probes

Magic-angle spinning (MAS) probes allow for broad resonances in solid materials to be narrowed by spinning the sample at the magic angle (54.74°) with respect to the magnetic field. This process averages internal interactions such as chemical shift anisotropy, dipolar coupling and first order quadrupole coupling, enabling high-resolution NMR by reducing line broadening. Spinning of the sample is generally achieved by using an air turbine system to drive a rotor at a rate of anywhere between 1–100 kHz. The complex design of these probes did mean that cryogenically cooling the coils wound around the sample is very difficult without decreasing the sample temperature at the same time. Recently, however, a cryogenically cooled MAS-NMR probe was designed allowing the measurement of the proton resonance of solid samples at room temperature. The desired increase in signal-to-noise was confirmed using this approach.¹⁶

1.2.4 Digital Filtering

The digital filter, when applied to NMR data, has many advantages over the pure application of analog filters, including improved baseline performance, better dynamic range, and improved overall sensitivity.¹⁷ The availability of fast digital signal processors (DSPs) has made the application of digital filtering a reality in NMR spectrometers, with very similar improvements in quality to other applications where source data from audio or visual sources are handled.

The specific path of data through the digital filter of an NMR spectrometer starts with unfiltered analog data entering an analog to digital converter (ADC). Because of a faster sampling rate, the number of points increases, which in turn needs to be reduced by extracting the spectral window of interest. The sampled digital signal is filtered and reduced in real time with a DSP. Finally, the filtered FID is sent to a workstation where Fourier transformations provide us with the spectral data for analysis. The quality of data has dramatically improved since the advent of digital filters as:

- The transition region of analog filters is rather large compared to digital filters and folding of signals and noise into the spectral window occurs.

- A digital filter has the ability to adapt and change the way in which a signal is processed.
- The characteristics of an analog filter are subject to drift and depend on the environmental surrounds, including the temperature and electromagnetic radiation. Digital filters are particularly stable with both time and temperature.

1.2.5 Computational Support to Hardware

Although computational support is not technically hardware, and so is not in the current chapter's scope, as hardware technologies improve, so does the computational support to achieve the full benefit of improvements in spectrometers. Areas of computational support need to constantly evolve with spectral instrumentation and data quality. Metabolomic data collection is still evolving and so the metabolomics community has a large volume of complex algorithms written in an array of languages to support the high-throughput analysis of metabolomics sample cohorts. There are already many user-orientated computational platforms specifically for processing complex NMR spectral datasets.¹⁸ Some of the fastest evolving elements of computation support with regards to NMR data are mentioned in this section.

1.2.5.1 Databases

NMR analysis has the advantage of being completely quantitative, accurately measuring the amount of protons, under given conditions. Spectral data can therefore be directly compared and spectral features of interest elucidated. Databases of standards are generally easy to come upon and there is a lot of information used to help assign NMR signals. Unfortunately, NMR suffers from comparably poor resolution and issues relating to molecules of similar structure producing similar magnetic resonance signatures, making them difficult to confidently assign without further work. Current research is aimed at producing reliable tools to deconvolve NMR signals from complex biofluid spectra producing an accurate measure of a number of metabolites automatically in large profiling studies. Issues of varying chemical shifts between samples, along with overlapping signals of varying shape, means creating automated packages to measure metabolites and their concentrations from spectral data is very difficult and is a form of on-going research. Metabolomics databases of NMR spectral data are readily available,¹⁹ including a urine-specific database.²⁰ Databases need to address a series of needs, including their availability, experimental validation, easily searched, readily interpreted and comprehensibility. These needs are constantly adapting and so databases are constantly in need of updating as spectral data improves as a result of hardware improvements.

1.2.5.2 Laboratory Information Management Systems

Sometimes referred to as a Laboratory Information System (LIS) or a Laboratory Management System (LMS), a Laboratory Information Management System (LIMS) supports the operation of a modern laboratory, particularly high-throughput laboratories where sample numbers are large and processes require mediation. Aside from the key functions of sample management, instrument application and data transfer, LIMS are also able to provide basic laboratory functions like document management, calibration, maintenance requirements, and data entry.

There are many open-source LIMS customised towards specific types of biomaterials or research data. More generic or general LIMS have come about as a necessity to catalogue large-scale biobanks of human specimens and manage and track their progress through a cascade of complex analytical procedures and experimental procedures.²¹

1.2.5.3 Metabolic Modelling

Metabolomics data sets, along with other phenomic data, generate increasingly complex data tables, which are becoming more difficult to summarise and visualise. For many years continuous spectral data, like that produced by an NMR spectrometer, has been statistically correlated with disease groups of other regions of the spectrum.²² Taking advantage of the multicollinearity of the intensities of signals in a set of spectra, spectral regions can be correlated with one another, enabling the association of peaks from the same molecule or similar biochemical pathways. Similarly, OPLS-DA analysis can be visualised in the same fashion by highlighting spectral regions associated with a discrete or continuous variable of physiological interest. Displaying highly correlated areas of a complex spectrum can even help identify the molecules responsible for the given signals or metabolic variation.

1.3 Automation of Metabolomic Profiling

The process of acquiring NMR spectral data begins with a submission of a request and finishes with data distribution and archiving. With the era of 'big data' upon us and the demand to automate processes, many types of hardware systems have recently become available to automate sample preparation, sample insertion and sample acquisition. These hardware accessories are linked with software packages often controlled by a LIMS to enable the organisation of spectral data and the corresponding metadata. The specific automation hardware systems are generally accessories to the components of NMR spectrometers (Section 1.2), although work in closely timed conjunction with the spectrometer to produce accurate and precise spectra without the need for constant supervision by a member of the laboratory.

1.3.1 Sample Preparation

In many NMR applications, as sample acquisition time is indeed the limiting factor for throughput, sample preparation does not need to be automated. In this case, the amount of a spectroscopist's time taken for sample preparation is considered negligible. In most situations robotic preparation is considered more precise and accurate; however, this is not always the case, and so there is no direct advantage to developing an automated preparation routine when sample preparation does not limit throughput. On the other hand, sample acquisition can be routinely very quick and therefore the manual preparation of samples can be tedious, tiring and unpractical. When hundreds, or even thousands, of samples are being acquired a day the development of automated routines for sample preparation is required.

Automation of metabolomic sample preparation for NMR is made difficult by the large variety of NMR applications and various types of biofluids. The variety in applications leads to a large number of preparative protocols that need to be catered for by robotic systems. Variations in total volume, solvent viscosities, tube diameters, sample temperatures and sample types all need to be accounted for when developing an automated routine to prepare a cohort of samples. The sample preparation (much like the collection) is ideally performed in such a way that as little variation as possible in the metabolic content of samples is encouraged during the process. In many cases, however, preparative and analytical variation cannot be avoided. Inconsistencies in quantification of metabolites can arise from many sources during sampling, sample storage, sample extraction, derivatisation, analysis and/or detection. During the sampling and preparation of samples, undesired changes in the metabolic content of samples may occur owing to enzyme activity, high reactivity of metabolites, or breakdown or degradation of metabolites. Owing to the nature of metabolomics samples, a very inert analytical preparation is required to minimize absorption and degradation of metabolites, particularly relatively polar compounds. Furthermore, the degree of adsorption and degradation can vary between different samples with different biomass concentrations and different sample matrices. Consequently, such matrix effects should be evaluated.

Preparation of quality control samples (both internal and external) in the project design helps understand the amount of preparative and analytical variance introduced during a study. Three sample types are useful for these measures, a composite quality control (QC), a stable quality control, and an external quality control. A composite study reference (SR) is simply a reference solution made from equal amounts of each sample within the study. A composite QC is made from the SR in the same way that each individual sample is prepared. A stable QC is prepared on a much larger scale with the reagents or buffers necessary and then transferred into multiple tubes or vials for analysis. An external QC is simply produced with samples of the biofluid of interest, obtained completely independently of the study, and prepared similarly to each individual sample. Buffer blanks can also be prepared

by simply replacing a sample with pure water. These samples are important to ensure the robot is performing as expected and that the prepared buffer or samples are not being contaminated during the preparation process.

1.3.1.1 Robotics for Sample Preparation

There are many brands (more than 15), each with multiple makes, of commercial liquid handling devices currently available to aid laboratory preparation of samples. A complete review of these devices is outside the scope of this chapter; however, it is important to note that many devices do not cater for NMR tube sample preparation. Robotic preparation into standard 5 mm tubes requires that samples are prepared with a syringe and the robot is able to clear the classical 7 inch length of the tube. The nature of how the tubes are held in position within the robot while not confusing the individually held tubes is also an issue. A simple solution to this came about with the pioneering of SampleJet Racks (compatible with SampleJet automation systems, Section 1.3.2), which position 3 mm or 5 mm tubes into a 96 well plate format. The tubes are also only 4 inches in length and so are generally compatible with robots that are capable of preparation using well plates (Figure 1.1).

Preparation of samples within a 96 tube rack generally occurs one at a time, or alternately a robotic solution with a multi-syringe is required. Furthermore, robotic preparation requires that each tube either has its cap off during preparation (which entails the tedious process of taking each tube cap off before preparation and replacing after preparation) or alternately a small entry hole in the cap to allow the preparative syringe to enter through. The most straightforward solution to this problem currently is to have a small



Figure 1.1 96 tube sample rack containing 5 mm sample tubes developed to enable sophisticated automation of the NMR preparation and acquisition process of biofluids.

hole in each cap (Figure 1.1) and then following sample preparation sealing each cap with a small ball (POM) to contain the biofluid sample inside.

Further complications are met as buffer solutions, urine, plasma/serum and other biofluids manipulated regularly for routine metabolomics analysis all have varying viscosities. Varying viscosities lead to variations in aspirated and dispensed volumes when using robots that use either push solvents or air pressure. Furthermore, during study design the amount of biofluid available is generally small and metabolomics assays can require relatively large volumes compared with other available screening techniques. Modern bio-banks spend a lot of time debating sample amounts that should be spared for particular types of analysis. During the preparative procedure it is therefore also important that the volume of sample 'wasted' by preparative robots is considered and preferably kept to a minimum, making metabolomics approaches more reasonable for clinicians of population research groups. All these complications lead to many forms of optimisation and understanding of robotic preparation before routine high-throughput preparation can occur.

Currently, the Gilson Automatic Liquid Handler is the most versatile robot available for preparation into a SampleJet 96 tube rack, although not necessarily the best for compatibility with tubes >7' in length. The Gilson 215 robot has a single syringe format, so samples must be prepared one at a time, while the Gilson 274 allows preparation of up to four samples at a time. Both robots are well-suited to house refrigeration mats while preparation is occurring, minimising the time biofluids are exposed to room temperature. A Liquid Handler 215 takes 3 h to prepare a 96 tube rack, including the addition of reagent to the sample, injection into the tube, and a mixing and cleaning procedure for each sample.

1.3.2 Automated NMR

Given the large numbers of samples to be analysed and the need for exceptional levels of reproducibility, new analytical protocols have had to be developed for high-throughput metabolomics by NMR.²³ This protocol describes a significantly enhanced and augmented version of the NMR spectroscopy protocol²⁴ previously published for any large-scale phenotyping projects utilizing the latest generation digital spectrometers. The previous version of the protocol is still suitable for small scale studies that do not rely on multiple batch analysis and where interstudy and interlaboratory comparisons are less important. NMR spectroscopy is intrinsically both quantitative and, with the current state of commercially available equipment, highly reproducible. With modern digital electronics-based machines, exact parameter settings can be set, but there are still other considerations and procedures required to obtain reproducible sample chemistries. Before performing the spectroscopic analysis, it is necessary to consider aspects such as sample numbers per group and randomization. In order to be able to validate results, sufficient sample numbers should be analysed to achieve statistically significant

results (the numbers required can be estimated using multivariate power calculations²⁵).

Generally, NMR-based metabolic studies of biofluids have shown very high reproducibility,²⁶ but new protocols linked to new technology as described herein have enabled even greater levels of reproducibility to be obtained without compromising the level of throughput. Using current automation provided by the SampleJet system, samples can be kept chilled while awaiting analysis (Section 1.3.2.1) and problems with insertion of samples into the magnet are rare. It is important to keep aliquots of samples at the sample collection stage in order to be able to repeat acquisitions or for subsequent 2D NMR spectroscopy necessary for biomarker identification. Consideration has to be given to quality control at all stages of the phenotyping process (*i.e.*, quality of the sample subject, quality of the study design, quality of the sample collection, quality of sample aliquoting, quality of the sample storage, quality of the preparation of samples for NMR, quality of the acquiring of data, and quality of the upkeep of the parameters desired over time). Ultimately, to achieve a spectral snapshot of the metabolic content of a biofluid at the time of collection, care must be taken at each stage of sample handling to ensure that a true indication of the biofluid content is observed.

1.3.2.1 Hardware for Automation

Many sample automation units are available for high-throughput analysis. Historically, sample automation was performed by a carousel (generally holding either 24 or 60 samples), which simply rotated as each sample was analysed. The user would need to place each sample in a regular tube and then the tube into a spinning shuttle. The tube already measured to a depth in the shuttle was placed in a specific position in the carousel along with the small cohort of samples and through the relevant software programmed to guide the spectrometer to perform particular experiments on each sample. The carousel could be emptied of acquired samples and restocked with further samples if necessary. There are four main types of technology currently available for installation onto an NMR magnet, SampleMail™, SampleCase™, SampleXpress™ and SampleJet™, each designed to completely automate the acquisition process. The first three automation devices are based on the classical carousel whereas the SampleJet provides a complete makeover of high-throughput NMR and is the ideal device for automating metabolomic profiling research.

- The SampleMail is simply a device that allows a single sample to be placed into and out of the magnet without the need for reaching over the top of the magnet (avoiding the use of a ladder).
- The SampleCase is a similar system to the SampleMail however allows up to 24 samples to be held at user height in a carousel. The SampleCase, like older carousels, still requires samples to be placed into individual shuttles by the user, ordered in the carousel manually and programmed *via* the software. The major advantage to metabolic research over

classical automation methods is that samples can be cooled to a given temperature (refrigerated) while they wait on the carousel to be analysed.

- The SampleXpress is not a refrigerated storage unit reducing its potential for running large biofluid sample cohorts. It does however incorporate an automatic barcode reader, which automatically locates and identifies samples that are being analysed, avoiding user error.
- The SampleJet is the ideal solution for automated high-throughput metabolomics research as it was designed specifically with biofluid applications in mind (Figure 1.2). Samples are able to be prepared as usual or in a 96 tube well plate format. When samples are prepared into a 96 tube rack, the sample tubes do not need to be placed into individual shuttles but rather a single shuttle, which is controlled by the robot, is placed in the magnet's shaft and samples are simply inserted into and removed from the shuttle automatically. Samples can be refrigerated to a given temperature in one of five rack positions while they are queued for analysis. Tracking sample position can be performed by imaging barcodes on individual tube caps or alternately samples can be tracked by simply using their position in a given rack. Finally, SampleJet automation units are also compatible with classical individual 7 inch tubes; however, unlike classical automation devices, NMR tubes do not need to first be placed in a shuttle (Figure 1.2c).

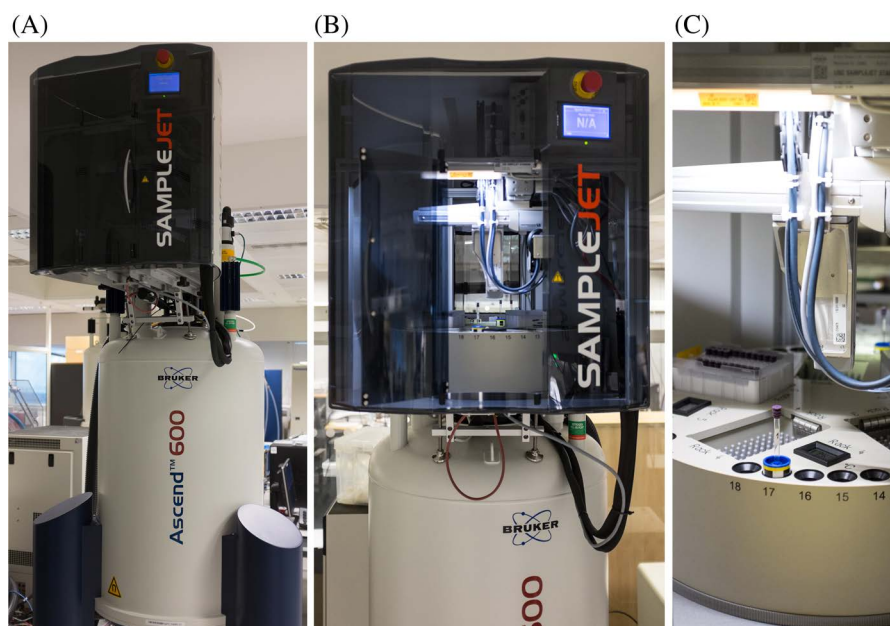


Figure 1.2 Bruker SampleJet automation unit installed on a 600 MHz magnet (A); access to temperature-controlled SampleJet unit (B); and interior of SampleJet system displaying samples queued for analysis in both 96 well plate racks (at rear) and individual 7 inch tubes loaded into NMR spinner collars (C).

Whatever automation system is chosen for a modern metabolic profiling lab, the system should be heavily integrated into the laboratory's workflow and LIMS. Automated NMR is always a central point of exchange between sample preparation and sample acquisition, making it an important key to any metabolic profiling laboratory's sample tracking and throughput.

1.3.2.2 Temperature Control

The temperature of urine or plasma/serum samples during storage, preparation, preacquisition and analysis is of particular importance in the metabolomics field as temperature changes can influence the metabolic content. Maintaining constant temperature conditions across sample cohorts is essential to identify subtle biological differences of interest. Storage of samples at $-80\text{ }^{\circ}\text{C}$ is widely accepted in the metabolomics community, ideally with no freeze-thaw cycles during the period of storage.²⁷ From the point of thawing samples, samples should be kept refrigerated during preparation to avoid large changes in temperature between thawing and analysis. Samples can be prepared on ice packs when manually preparing samples or for robotic preparation refrigeration mats compatible with the specific robot can be placed underneath well plates and tube racks. In an ideal world this maintains a sample temperature below $10\text{ }^{\circ}\text{C}$ during the preparative process.

Prepared samples should be immediately placed in an NMR automation system. Laboratories with automation systems that contain temperature control have the advantage of being able to set the temperature ($\sim 5\text{ }^{\circ}\text{C}$) of samples queued for analysis. SampleJet systems can be programmed so that one or two samples (depending on sample analysis time), following the sample currently being analysed, can be placed in a heating block set at the same temperature of NMR acquisition (generally $27\text{ }^{\circ}\text{C}$, or $37\text{ }^{\circ}\text{C}$ to mimic biological conditions). The heating block coincidentally reduces the time taken for temperature equilibrium to be reached inside the magnet during the automatic acquisition routine (Section 1.3.3). Finally, once the sample is placed inside the magnet, the temperature is given a set amount of time to reach equilibrium ($\sim 90\text{ s}$) followed by a set amount of time ($\sim 10\text{ s}$) to record any temperature change. If a change in temperature is registered, the equilibrium process is repeated; if not, equilibrium is deemed to have been reached and acquisition can proceed. Post-acquisition, samples should be placed back into a freezer at $-80\text{ }^{\circ}\text{C}$ for storage. Although the sample itself is no longer considered fit for cohort metabolic profiling, it may be useful in identifying structures of unknown metabolites of interest.

1.3.3 Automated Acquisition

The current automation programs for NMR include not only the acquisition routine for locking, tuning, matching, shimming, pulse calibration, and optimized presaturation power for individual samples, but there are also now

automated routines for processing that include phasing, baseline correction, and calibration. These routines can be formulated in many different fashions, but they are essential for the analysis of large sample cohorts.

1.3.3.1 Instrumental Calibration

It is important before running a large metabolic profiling study that the hardware is performing within specific requirements of accuracy and precision. These requirements include:

- *Temperature Calibration.* The temperature for a urine sample set should be adjusted to 27 °C while the temperature for a plasma/serum sample set should be adjusted for a real temperature of 37 °C.²⁸ This is accurately achieved by running a standard 90° proton parameter set using a pulse length of 1 μs on pure MeOD.²⁹ The short pulse length and highly deuterated methanol ensure the sample is not affected by radiation dampening. When the experiment is processed without any line broadening, the methanol peak at 3.33 ppm (from CD₂H-) should have a distinct 1:2:3:2:1 multiplicity due to H–D spin coupling and be symmetrical. Once achieved, the experiment is processed with a line broadening of 3.0 Hz and the real probe temperature calculated by measuring the distance in Hz between the two methanol peaks, the CH₃ and OH resonances, and by referring to a calibration curve supplied by the instrument manufacturer (1.526 ppm for 27 °C and 1.428 ppm for 37 °C). If required, the target temperature of the probe is adjusted and the procedure repeated until the actual calculated temperature of the sample is within ±0.05 °C. The temperature is recorded and later stored in the relevant parameter set for running under automation.
- *Water Suppression.* A standard 2 mM sucrose sample (containing 0.5 mM TSP and 2 mM NaN₃ in 90% H₂O:10% D₂O) is used to check the performance of the water suppression functionality. Assuming the pulse frequency has been optimised, the water suppression performance is evaluated by acquiring a full cycle (eight scans) experiment and relatively long delays (~10 s). The signal-to-noise value should be higher than 300 (as measured on the anomeric proton of sucrose), the resolution better than 15% (a measure of the height of the minimum of the anomeric proton as a percentage of the entire peaks signal) and the water hump not bigger than 40/80 Hz (as measured at 100% and 50% of the TSP signal intensity, respectively).

1.3.3.2 ERETIC

Quantification based on reference to a synthetic signal in NMR data has been available for many years. Namely, the ERETIC method³⁰ (Electronic Reference To access *In vivo* Concentrations) provides a reference signal, synthesized

by an electronic device, which can be used for the determination of absolute metabolite concentrations.³¹ Lately, methodologies have been developing to allow even metabolites in complex biofluid matrices to be quantified by these methods.³² It is therefore very useful to run a calibration monthly (or before each sample set) on a sample of quantified, resolved signals for quantification reference. A good quantification reference sample consists of approximately 20 standard metabolites commonly found in biofluids that are known to be chemically stable over the long term as a mixture. Ideally a universal reference is run across all systems enabling the comparison of all instrumentation used. An experiment using the same parameter set as that used for sample acquisition should be acquired on the reference, and care taken during deconvolution for quantification of resonances as biofluids are complex and signals often unresolved.

1.3.3.3 Automation Sequences

Automation procedures allow for the exact same set of tasks to be performed on each biofluid sample. There are often different ways to get to the same result using NMR procedures, however where interstudy and interlaboratory comparisons are planned, it is essential that these routines are kept consistent. A typical routine will cover the following steps:

- (1) Tuning and matching will be adjusted.
- (2) The sample will be locked to the solvent for urine or plasma/serum samples (note: solvent locking parameters can be optimized for particular biofluids).
- (3) The temperature is equilibrated as explained previously (Section 1.3.2.2).
- (4) The sample is shimmed using an automated routine.
- (5) The 90° pulse length is calculated along with the presaturation frequency and power.
- (6) The data set is updated with the parameters optimized.
- (7) The data are acquired using the optimal parameter set(s).
- (8) The routine should read in the optimal processing parameters and automatically Fourier transform, phase, calibrate and baseline the data.

For the NMR experiments, a throughput of 96 urine samples (excluding reruns) can be achieved within 24 h (~15 min per sample) when running a NOESY-presat and JRES. Plasma and serum samples NMR data can be acquired at a rate of 72 samples (excluding reruns) per 24 h when running a NOESY-presat, CPMG, and JRES (~19 min a sample) spectrum set. An acceptable rate of reruns in modern day systems is in the order of 5%.

Quality assurance of spectral results can be conducted by running programs in real time to advise spectrometers to either rerun samples and/or

exclude spectral data from multivariate analysis and biomarker discovery stages. The main criteria assessed are:

- *Line Width:* In urine the signal resulting from the TSP should be symmetrical and at half height the line width should be <1 Hz after a line broadening of 0.3 Hz has been applied to the data.
- *Baseline:* The spectral noise (from a line through zero relative intensity) should have no residual trend outside the -0.5 – 10 ppm range for urine and outside the -5 – 15 ppm range for plasma/serum samples.
- *Water Suppression:* A urine NMR spectrum should not be affected by the residual water resonance outside of the range 4.7 – 4.9 ppm. As plasma/serum is run at 310 K the residual water resonance shifts to a higher ppm value but this should not affect the resulting spectrum outside of 4.6 – 4.8 ppm.
- *Phase Error:* Checking the zero-order phase manually (after automatic processing) around the TSP signal, there should be less than a 0.2° distortion in the value if correcting by eye. A first-order phase correction should not be necessary (set to 0.0°).
- *Receiver Gain:* Baseline artefact problems (sinc-wiggles) can occur when the sample concentrations are too large because the analog–digital converter in the detection process is overloaded. For consistency, it is preferred that the receiver gain is not adjusted during a run of samples, but if overload does occur, then the affected spectra should be discarded and the samples should be re-run with a readjusted receiver gain value.

1.3.4 Integrated Metabolic Profiling NMR Systems

Innovative solutions to the analysis of complex mixtures of small molecules have already been developed and commercialised. Effectively, the profiler systems are push button (do not require an expert to be operated) and ultimately feedback a wealth of information based on the metabolic content of solutions like fruit juices or wines. The user aliquots a small volume of sample (~ 0.5 mL) into a well plate format, which is then placed in a liquid handling robot for preparation. Once prepared, the NMR tube is taken from the preparation robot, a barcode is placed on the tube and the sample placed in an autosampler. At this stage the user's interaction with the sample is finished and the unique barcode tracks the sample through the LIMS for the duration of the analytical process, quality assurance of data and presentation of results. Finally, a sample-specific report is produced that presents accurate concentration measures of a considerable number of metabolites, distribution analysis in comparison to similar samples and basic multivariate analysis within a similar population.

Juice and wine manufacturers have a particular interest in understanding the chemical compositions of their product for both legal and quality reasons. Juice manufactures, for example, are given a specific regulation for the ethanol concentration in a product. The concentration of ethanol, along with

other metabolite amounts, is regulated in many nations to varying thresholds. Metabolic profiling allows juices to be screened to ensure they meet the legal concentration thresholds of countries their product is for sale in *via* a single NMR assay. Wine manufactures are also considerably interested in the chemical composition of their final product. Moreover, metabolic profiling is able to provide a basic insight in how similar a current batch of wine is to other batches, year's yields, grape varieties, and similar award winning fermentations.

With metabolic profiling of 'less complex' fluids becoming routine, the strategic aim of current research is to develop rigorous systems to access human phenotypes through biofluid analysis. The metabolic content of the population's biofluid can then be used to assess a person's health status and guide more informed health interventions and more advanced medical treatments.

References

1. J. Arnold, S. Dharmatti and M. Packard, *J. Chem. Phys.*, 1951, **19**, 507.
2. N. Ramsey and E. Purcell, *Phys. Rev.*, 1952, **85**, 143.
3. P. Lauterbur, *J. Chem. Phys.*, 1957, **26**, 217.
4. F. Anet and A. Bourn, *J. Am. Chem. Soc.*, 1965, **87**, 5250–5251.
5. R. Ernst and W. Anderson, *Rev. Sci. Instrum.*, 1966, **37**, 93.
6. P. Lauterbur, *Nature*, 1973, **242**, 190–191.
7. M. Bird, *IEEE Trans. Appl. Supercond.*, 2012, **22**, 4300504.
8. D. Hoult and R. Richards, *J. Magn. Reson.*, 1976, **24**, 71–85.
9. M. Lacey, R. Subramanian, D. Olson, A. Webb and J. Sweedler, *Chem. Rev.*, 1999, **99**, 3133–3152.
10. L. Da Silva, M. Godejohann, F. Martin, S. Collino, A. Bürkle, M. Moreno-Villanueva, J. Bernhardt, O. Toussaint, B. Grubeck-Loebenstein, E. Gonos, E. Sikora, T. Grune, N. Breusing, C. Franceschi, A. Hervonen, M. Spraul and S. Moco, *Anal. Chem.*, 2013, **85**, 5801–5809.
11. S. Shuker, P. Hajduk, R. Meadows and S. Fesik, *Science*, 1996, **274**, 1531–1534.
12. M. Pellecchia, D. Sem and K. Wüthrich, *Nat. Rev. Drug Discovery*, 2002, **1**, 211–219.
13. G. Eldridge, H. Vervoort, L. Lee, P. Cremin, C. Williams, S. Hart, M. Goering, M. O'Neil-Johnson and L. Zeng, *Anal. Chem.*, 2002, **74**, 3963–3971.
14. L. Sumner, Z. Lei, B. Nikolau, K. Saito, U. Roessner and R. Trengove, *Metabolomics*, 2014, **10**, 1047–1049.
15. J. Shockcor, S. Unger, I. Wilson, P. Foxall, J. K. Nicholson and J. C. Lindon, *Anal. Chem.*, 1996, **68**, 4431–4435.
16. T. Mizuno, K. Hioka, K. Fujioka and K. Takegoshi, *Rev. Sci. Instrum.*, 2008, **79**, 044706.
17. D. Moskau, *Concepts Magn. Reson.*, 2002, **15**, 164–176.
18. J. Xia, N. Psychogios, N. Young and D. Wishart, *Nucleic Acids Res.*, 2009, **37**, W652–W660.

19. D. Wishart, D. Tzur, C. Knox, R. Eisner, A. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. Duggan, G. Macinnis, A. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. Sykes, H. Vogel and L. Querengesser, *Nucleic Acids Res.*, 2007, **35**, D521–D526.
20. S. Bouatra, F. Aziat, R. Mandal, A. Guo, M. Wilson, C. Knox, T. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z. Dame, J. Poelzer, J. Huynh, F. Yalou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring and D. Wishart, *PLoS One*, 2013, **8**, e73076.
21. S. Bloch, J. Lee, T. Syburra, U. Rosendahl, M. Griffiths, P. Kemp and M. Polkey, *Thorax*, 2015, **70**, 219–228.
22. O. Cloarec, M. Dumas, A. Craig, R. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes and J. K. Nicholson, *Anal. Chem.*, 2005, **77**, 1282–1289.
23. A. C. Dona, B. Jimenez, H. Schafer, E. Humpfer, M. Spraul, M. R. Lewis, J. T. M. Pearce, E. Holmes, J. C. Lindon and J. K. Nicholson, *Anal. Chem.*, 2014, **86**, 9887–9894.
24. R. Barton, J. K. Nicholson, P. Elliot and E. Holmes, *Int. J. Epidemiol.*, 2008, **37**, 37–40.
25. M. Chadeau-Hyam, T. Ebbels, I. Brown, Q. Chan, J. Stamler, C. Huang, M. Daviglus, H. Ueshima, L. Zhao, E. Holmes, J. Nicholson, P. Elliott and M. De Iorio, *J. Proteome Res.*, 2010, **9**, 4620–4627.
26. H. Keun, T. Ebbels, H. Antti, M. Bollard, O. Beckonert, G. Schlotterbeck, H. Senn, U. Niederhauser, E. Holmes, J. Lindon and J. Nicholson, *Chem. Reson. Toxicol.*, 2002, **15**, 1380–1386.
27. M. Rist, C. Muhle-Goll, B. Görling, A. Bub, S. Heissler, B. Watzl and B. Luy, *Metabolites*, 2013, **3**, 243–258.
28. R. Farrant, J. Lindon and J. Nicholson, *NMR in Biomedicine*, 1994, **7**, 243–247.
29. M. Findeisen, T. Brand and S. Berger, *Magn. Reson. Chem.*, 2007, **45**, 175–178.
30. S. Akoka, L. Barantin and M. Trierweiler, *Anal. Chem.*, 1999, **71**, 2554–2557.
31. G. Wider and L. Dreier, *J. Am. Chem. Soc.*, 2006, **128**, 2571–2576.
32. M. Albers, T. Butler, I. Rahwa, N. Bao, K. Keshari, M. Swanson and J. Kurhanewicz, *Magn. Reson. Med.*, 2010, **61**, 525–532.