The Handbook of Metabonomics and Metabolomics
John C. Lindon, Jeremy K. Nicholson and Elaine Holmes (Editors)

Chapter 3

NMR Spectroscopy Techniques for Application to Metabonomics

Alfred Ross, Goetz Schlotterbeck, Frank Dieterle, and Hans Senn

Pharma Research, F. Hoffman La-Roche AG, Basel Switzerland

3.1. Introduction

Since its discovery in the 1940s, **N**uclear **M**agnetic **R**esonance (NMR) Spectroscopy has become a powerful, interdisciplinary method. A brief historical review would reveal as many as nine Nobel Prize laureates since the time when Isador I. Rabi developed resonance methods for recording the magnetic properties of atomic nuclei and was awarded the Nobel prize in physics (1944). The NMR phenomenon was soon later demonstrated for protons. After years of continuous development, Fourier Transform (FT) NMR entered the scene in the 1960s, followed by the evaluation of non-invasive preclinical and medical imaging in the early 1980s. At the same time, three-dimensional (3D) structure elucidation of proteins at atomic resolution in aqueous environment was developed [1]. This breakthrough was made possible also by the availability of superconducting materials and stable and robust electronic equipments.

NMR has since been used in an almost unlimited variety of ways in physics, chemistry and biology. For the investigation of biological systems it is convenient to distinguish between three types of applications [2]: (1) to study structure and function of macromolecules, (2) to study metabolism, and (3) to obtain *in vivo* images of anatomical structure and functional (physiological) states.

The use of ^{1}H NMR for metabolic studies was described as early as 1977 when it was shown that ^{1}H signals could be observed from a range of compounds in a suspension of red blood cells, including lactate, pyruvate, alanine and creatine [3]. A great deal of metabolic information can be derived from such metabolic studies and it was soon recognized that ^{1}H NMR of body fluids has a considerable role to

play in areas of pharmacology, toxicology and the investigations of inborn errors of metabolism [4–6].

Since these early applications, ^{1}H NMR of biofluids and cell extracts including high-resolution magic angle spinning (HR-MAS) NMR of soft tissues have been successfully applied to investigate numerous diseases and toxic processes [7–9].

This chapter is meant to give an introduction and overview on NMR under the view point of its application in metabolic profiling (metabonomics). It is not meant to review all the different applications of NMR in this field but rather to concentrate on the essentials and prerequisites of its successful implementation as a metabolite profiling tool. To this end, specific NMR hardware requirements for metabolite profiling are reviewed and compared as well as automation and robotics control of the work flow, which are crucial for achieving high throughput and consistent quality of results. It is then shown that sample preparation and handling is of utmost importance for meaningful comparison of hundreds of samples and thousands of spectral variables in a metabolite profiling study. Therefore all aspects known to us which may change the sample property of biofluids are included and extensively discussed. The handling and preparation of urine samples is especially demanding in order to control and handle the wide variability of conditions that influence the ^{1}H NMR spectrum. These problems are less severe with blood serum, plasma or **c**erebro**s**pinal **f**luid (CSF) where homeostasis ensures a much narrower range of sample variability. In the section on NMR Experiments and Processing, the information which is necessary to obtain high-quality one-dimensional (1D) and two-dimensional (2D) NMR spectra of biofluids is summarized. Finally, the state-of-the-art of data pre-processing, which is the intermediate step between recording NMR raw spectra and applying uni- or multivariate data analysis and modeling methods, is discussed in great detail. In metabonomics, this is an important step to make subsequent analysis and modeling easier, more robust and more accurate.

3.2. Principles of NMR

The theory of NMR is highly developed and the dynamics of nuclear spin-systems fully understood [10, 11]. A first-principle quantum mechanical description is available, but beyond the scope of this introduction. What is missing to complete the theoretical picture is a reliable and accurate prediction of the chemical shift of spins. Therefore, recourse to spectral databases is needed if a chemical interpretation of metabonomics data is envisaged.

We present a phenomenological description of magnetic resonance needed for the understanding of the metabonomic literature. It has to be stressed, however, that nearly any aspect of modern NMR spectroscopy is of importance for the acquisition of high-quality data needed for a reliable biological interpretation of results. After an

introduction into the principles of NMR (magnetization, chemical shift, relaxation, J-coupling) we touch upon more advanced concepts involving chemical exchange, 2D and heteronuclear experiments. We will see that all effects are of relevance for metabonomic research. We do not seek for a complete review of literature of the topic. If possible we provide examples from our own work.

3.2.1. Magnetism

Magnetism and spin: Matter is composed of molecules built of atomic nuclei with a characteristic proton/neutron composition. Nuclei are surrounded by electronic "clouds". Besides charge and mass, a further property of protons, neutrons and electrons is an angular momentum $\vec{I}$ known as spin. Because of the magnitude of this angular momentum $\vec{I} \cdot \vec{I} = \frac{h}{2\pi} \cdot \frac{3}{4}$ the aforementioned particles are known as spin-1/2 particles.[1] The total spin of a nucleus depends on its nucleon content. Some nuclei carry a total spin (^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P, . . .)[2] resulting in a magnetic moment $\vec{M} = \gamma_X \cdot \vec{I}$ of different magnitude and sign, others (e.g. ^{12}C,^{14}C, ^{16}O, . . .) do not; γ_X is the gyromagnetic ratio of the atomic nucleus X.

In NMR-spectroscopy samples of liquid or solid material[3] are exposed to an external static and homogeneous magnetic field referred to as B_0. The direction of this field is usually defined along z. The magnetic moments in the sample align along B_0 according to a Boltzmann distribution. In contrast to classical physics, quantum mechanics shows that magnetic moments due to spin-1/2 particles can only align parallel (called up) or anti-parallel (called down) with respect to this external field, these two states have a difference of energy given by $\Delta E = \gamma_X \cdot B_0$. Due to statistic averaging, the magnetic moment of any macroscopic sample can be treated in many respects like a classical macroscopic magnetic moment $\vec{M}$. In accordance with the definitions above, the thermal equilibrium magnetization of a sample aligned along z is called M_0. For the detection of the NMR signal, M_0 is flipped orthogonal to B_0 by use of a high-frequency magnetic field B_1 applied for a defined time period. B_1 is also applied orthogonal to z (see 3.2). This is called a B_1-pulse. Classical physics shows that $\vec{M}$, now aligned without loss of generality along the x-direction, will precess with a (resonance) frequency given by

$$f_0 = \frac{\gamma_X}{2\pi} \cdot B_0 \tag{3.1}$$

[1] h is known as Planck's constant.

[2] H is a spin-1 particle. To date there is no application for ^{2}H and other "higher spin" nuclei reported for metabonomic research.

[3] There are metabonomic applications for tissues. Tissues are not real solids, but more gel-like and soft structures where the spin-physics can be transformed to liquid-like behavior by use of the so-called HR-MAS technique.

This is the reason why NMR spectrometers are normally classified in frequencies. The magnetization of a sample of hydrogen atoms experiencing a magnetic field of 14.1 T will, for example, rotate with 600 MHz. Such a spectrometer is called a 600 MHz apparatus. Other nuclei will rotate in the same field with another frequency. A ^{13}C spin will precess in the same 600 MHz magnet at about 150 MHz.

Detection of the signal: The sample is positioned inside a detection coil (see Figure 3.1). According to the law of inductivity the precessing magnetization will induce a voltage U_{ind} modulated with f_0. The amplitude of this voltage is directly proportional to $\vec{M}$ and thus with the number of spins rotating with f_0 located inside the observe volume of the apparatus. In the NMR literature the signal detected is called a **F**ree **I**nduction **D**ecay or FID.

Due to diamagnetism of the electron clouds of the molecule the local magnetic field experienced by a spin is changed by a small amount compared to B_0. Thus the frequency f_A measured for spin A directly reports the electronic/chemical neighborhood of the nucleus observed. The value $\frac{(f_A - f_0)}{f_0} \cdot 10^6$, measured in parts per million (ppm) in respect to B_0, is called the chemical shift of this spin. This definition is independent of B_0 – thus data taken at different field strength can be compared easily. The chemical shift of different spin species covers different ranges: ^1H nuclei resonate in most cases within a width of 15 ppm; ^{13}C nuclei cover more than

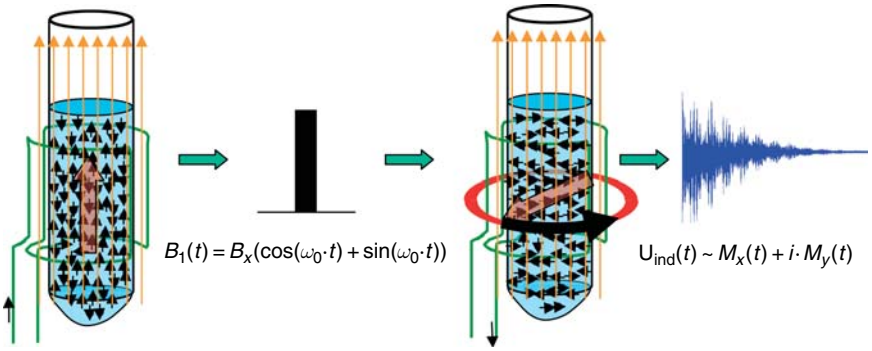

Figure 3.1. Detection of an NMR Signal: (left) a sample with many spins ($>10^{12}$) is placed inside a strong external magnetic field B_0 (orange) oriented along the z-direction. The sample is enclosed in a detection coil (shown in green). At thermal equilibrium about one spin in every 10,000 contributes to a macroscopic magnetization M_0 shown in transparent red. After short application of a high-frequency B_1 field (90° pulse: black bar and equation at the left panel) the magnetization is aligned along the x-axis of the rotating B_1 field, and starts to precess around B_0. The voltage, U_i, induced in the detection coil (dark-blue and equation at the right panel) can be described by the equation shown on the bottom of the right panel.

200 ppm. For the definition of f_0 normally a reference compound is added to the sample.[4]

Relaxation [12]: Each spin creates a small dipolar magnetic field spread over space. The magnitude of this additional magnetic field experienced by another spin species in the neighborhood (e.g. the same molecule) is dependent on the angle and the length of the vector connecting both spins with respect to B_0. As Brownian motion is active in liquids this angle is time-dependent. Thus besides a huge constant magnetic field each spin experiences an individual small time-dependent "jittery" magnetic field. As a consequence the frequency of rotation is time-dependent – each molecule experiences its own time-dependence. The signal measured is an ensemble average. This time-dependence will lead to an individual precession trajectory of the magnetization for any molecule of the sample reflecting itself in a "dephasing" of the measured macroscopic magnetization $\vec{M}$. The same happens with the induced voltage U_{ind}. The magnitude of the signal will decay over time. This phenomenon is known as dipolar relaxation. There are other sources of relaxation like quadrupolar relaxation. The effect of all relaxation mechanisms active for a certain spin species manifest themselves as an exponential decay of the signal characterized by a constant T_2 known as transverse relaxation.

If an unmagnetized sample is placed into a magnet the built up of M_0 will take a certain time. The time-constant necessary to approach this thermal equilibrium is known as longitudinal relaxation T_1. The physical mechanisms (e.g. dipolar interaction of spins) behind this phenomenon are essentially the same as those for T_2 but they are active in general with different rate constants here. For small molecules as a rule of thumb $T_1 \approx T_2$ is found. For macromolecules $T_1 \gg T_2$ is valid.[5] Stokes law $\tau_C \sim R_H^3$ $T_2^{-1} \sim \tau_C$ (τ_C, R_H being the rotational correlation time and the hydrodynamic radius of the molecule respectively) ensures that NMR signals of large molecules decay much faster than those of small molecules. This has been exploited in metabonomics for the suppression of protein-related signals [13] (see Section 3.5.1.2). For small molecular metabolites mostly investigated in metabonomics, the relaxation times are in the range 1–3 s. The magnitude of the "jittery field" and thus dipolar relaxation depends on the gyromagnetic ratios of the spins involved. Nuclei with a small gyromagnetic ratio (e.g. ^{13}C) relax more slowly and longer relaxation times between repetition of pulses is needed for their NMR detection (see Section 3.2.1.1).

[4] In metabonomics, the samples are usually aqueous solutions (e.g. urine) to which a small amount of partially deuterated TSP (**t**rimethyl**s**ilyl**p**ropionic acid-d_4) is added for reference purposes. It must be noted, however, that TSP does bind to proteins (e.g. human serum albumin, HSA). This has to be taken into account for biofluids like blood plasma or serum, as protein binding influences the chemical shift of TSP (see Section 3.2.2.1). The addition of TSP also allows the monitoring of the quality of the B_0 homogeneity across the sample, as the line width of this signal is expected to be identical for all samples as long as HSA binding can be excluded.

[5] T_1 and T_2 are dependent slightly on B_0, but this can be neglected for low-molecular-weight molecules investigated in metabonomics analysis.

Bloch equation: All what has been described above can be summarized phenomenological in the so-called Bloch equation: here the time evolution of any type of magnetization $\vec{M}$ can be described if direct spin–spin interactions (J-coupling, NOE, see Section 2.1.3) are neglected:

$$\frac{\partial}{\partial t}\begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} = \gamma_X \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} \times \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} - \frac{1}{T_2}\begin{pmatrix} M_x \\ M_y \\ 0 \end{pmatrix} - \frac{1}{T_1}\begin{pmatrix} 0 \\ 0 \\ M_z - M_0 \end{pmatrix} \qquad (3.2)^6$$

As a consequence of Equation 3.2, z-magnetization rotates in the z-y plane by application of a x-pulse. Magnetization in the x-y plane will evolve as a superposition of cosine-modulated and T_2-damped contributions under the presence of B_0. Irradiation, long on the timescale of T_1 and T_2, will lead to a substantial reduction of the NMR signals. This phenomenon is known as saturation and does find its application for the purpose of solvent suppression in metabonomics (see Section 3.5.1.1).

The NMR signal: The amplitudes $M^i_{z,0}$ of the individual contributions originating from one molecular species in a mixture are given by the number of spins in a molecule which resonate at a certain frequency f_i multiplied by the concentration c_j of the respective molecular species.[7] Thus the signal detected can be described as:

$$\begin{aligned} \vec{M}_{\text{molecule}}(t) &\sim \sum_i M^i_{z,0} \cdot \left[\cos(2\pi \cdot f_i \cdot t)\cdot \vec{e}_x + \sin(2\pi \cdot f_i \cdot t)\cdot \vec{e}_y\right] \cdot e^{\frac{t}{T_2}} \\ \vec{M}_{\text{total}}(t) &= \sum_j c_j \cdot \vec{M}^j_{molecule}(t) \end{aligned} \qquad (3.3)$$

Here the summation over "i" is done over the spin-types of a molecular species. The summation over j is performed over the different molecular species composing the sample.

Fourier transformation (see Section 3.2.1.2) of Equation 3.3 leads to the NMR spectrum of the sample (see Figure 3.1). The spectrum of a molecule is given

[6] All quantities in Equation 3.2 except relaxation times and the gyromagnetic ratio are in general time-dependent. Only time-independent magnetic fields are analytically solvable.

[7] $M^i_{z,0}$ is equal to the thermodynamic equilibrium value only, if the sample is in full equilibrium for each repetition of the experiment. (This is needed to achieve high signal-to-noise (see below).) Otherwise the simple assumption that a proton signal emerging from a methyl-group is three times higher compared to that of a methine group will break down. Equilibration is achieved to a good approximation by a waiting time between repetitions of the experiment in the order of 3 s. Experience shows that in this regime substantial falsification of amplitudes is prevented.

by resonance lines with defined frequencies and amplitudes.[8] The spectrum of a mixture is given by the concentration-weighted summation of the spectra of the molecular species constituting the mixture. This feature of perfect linearity makes the detection of NMR signals highly attractive for the quantitative determination of the concentrations of components of a mixture.

3.2.1.1. Noise, sensitivity, limit of detection

The detected signal is proportional to the total magnetic moment of a sample (law of induction). The latter depends linearly on the number of spins N (atomic magnets) in the sample and the difference between spins aligned in parallel with the B_0-field and those aligned anti-parallel. According to the high-temperature approximation of Boltzmann's law this difference increases linearly with B_0. There is an additional B_0 dependence in the detection process itself (the energy provided by any spin flip). Other important factors influencing the sensitivity of the measurement are the gyromagnetic ratio γ_x, the sensitivity of the detector and the noise created by the sample, the latter two summarized in S_D. The sensitivity defined as the signal-to-noise ratio (S/N) of a single repetition of the experiment is given by [14]:

$$\text{Sensitivity}_{\text{scan}} \sim \gamma_X^3 \cdot N \cdot B_{\text{o}}^2 \cdot S_{\text{D}} \qquad (3.4)^9$$

There exist two approaches to increase S_D (see Sections 3.1 and 3.3): (a) Noise in electric circuits is caused by thermal motion of electrons in wires. Consequently this noise can be reduced by cooling the wires of the detection coil and the preamplifier used to cryogenic temperatures of a few K. Such probeheads are called cryogenic or cold-probes [15, 16] (b) S_D can be increased by reducing the size of the detection coil [17, 18] – naively it is clear that shorter wires create less noise. It must be noted that for biological samples used in metabonomics the concentration of the sample is given by biology. Here the application of miniaturized detection devices is only recommended for volume-limited samples (e.g. CSF taken from a mouse).[10]

A sample of high conductivity (salt) reduces the sensitivity of the detection by induced eddy-currents. This situation is often encountered for biological samples having high ionic strength. The effect is more pronounced for cryogenic probeheads [19].

---

[8] Due to rotational averaging the chemical shifts of fast rotating structural moieties (e.g. methyl groups) will appear at the same frequency – they are called "degenerate". The same holds true for spins in fast chemical exchange between two sites (see Section 3.2.2.1).

[9] A more thorough calculation shows proportionality in $B_0^{7/4}$ and $\gamma_X^{11/4}$.

[10] This is true as the gain achieved for S_D by miniaturization is normally more than counterbalanced by the reduction of N if a smaller sample volume is used.

From Equation 3.4 it is also clear that a higher static magnetic field does increase the sensitivity of the NMR detection. This is one reason why magnets with higher B_0 field are attractive.

If the same experiment is repeated by adding the detected signal, it is important to realize that the signal sums coherently (proportional to the number of repetitions called scans (NS) or transients). Statistic averaging leads to an increase of noise linear in $\sqrt{\text{NS}}$. Thus a needed increase in signal-to-noise by two enforces a fourfold increase in experimental time.

Another important factor influencing the outcome is the rate of repetition of scans. A slow rate, that is long inter-scan delays, allows for full relaxation of the magnetization. This ensures that the NMR signal is perfectly linear to the concentration of the analytes. The trade-off in this regime is a long experimental time for a given number of repetitions. Analysis by use of Equation 3.2 shows that highest sensitivity under a loss of perfect linearity is obtained if the repetition rate is $1.26^{*}T_1$. The compromise used in metabonomics is given by a repetition of scans every 3 to 4 sec.

3.2.1.2. Principles of spectral processing [20]

Modern NMR spectrometers detect magnetization along the x- and the y-direction over time. This process can be visualized as if the measurement of induced voltages were done with two detection coils arranged along the x- and y-direction respectively.[11] For any time-point, two values M_x and M_y are recorded – this technique is called phase-sensitive detection.[12] For the calculation of the NMR spectrum, a complex valued time domain signal, constructed as $M_x(t) + iM_y(t)$ (with $i := \sqrt{-1}$), is used as input for a complex valued FT. The mathematics of FT tells that the resulting real part of this calculation performed on a time domain signal expressed in Equation 3.3 will show individual lines with positions given by the chemical shifts. The width of the lines is characterized by $(\pi^{*}T_2)^{-1}$ and shows a so-called absorptive lorenzian line-shape. This holds true only if all magnetization flipped in the x-y plane is perfectly aligned along the x-axis at the beginning of the detection. Due to limited strength of the B_1 field in use[13] and due to limitations of spectrometer electronics this does not hold true in practice. Magnetization is oriented under a small

[11] In reality the detection is done with a single coil using two-channel phase-shifted high-frequency mixing.

[12] Not only the magnitude of the transversal magnetization is measured, but also the orientation in the x-y plane, normally expressed as an angle (phase) with respect to the x-axis.

[13] Typically a B_1 pulse needed for flipping the magnetization by 90° lasts less than 7.5 µs for modern 100 W amplifiers. Thus a 360° rotation with the same pulse will last 30 µs. The reciprocal of this value defines the B_1 field strength in Hz. In this example, we have a B_1 field of 33 kHz. The excitation bandwidth of a pulse given in Hz is approximately $2^{*}B_1$. A pulse of 33 kHz excites at a 600 MHz spectrometer, a bandwidth of 60 kHz or 100 ppm for protons, easily covering the typical chemical shift range found in biofluids. The situation changes if multiple pulses are applied.

initial and frequency-dependent phase with respect to the *x*-axis. As a consequence, the real part of the FT will also contain a frequency-dependent contribution of the imaginary part of a perfect signal. This imperfection can be removed after FT by a process called "phasing". Another artifact found in NMR spectra is a distortion of the baseline due to imperfections and non-linearities of the electronic detection process. This baseline distortion can be corrected by the subtraction of a tailored polynomial from the raw spectrum obtained.

Unfortunately, both processes, phasing and baseline correction, can be performed automatically only with limited reliability. Visual inspection of processed spectra for artifacts is of importance.

The signal-to-noise and line-shape of spectra can be tailored according to needs if the time-domain signal is multiplied with a so-called window function [21] prior to FT. In metabonomics, the window function used for this purpose is typically the so-called exponential-window given by $e^{-t \cdot \pi \cdot \mathrm{LB}}$ employed with a so-called line-broadening LB of 1 Hz (see Figure 3.2). This offers an acceptable compromise between signal-to-noise and spectral resolution. If only a short acquisition time (e.g. in 2D applications, see Section 3.5.2.) is allowed, a shifted sine or squared sine-shaped window ensures the absence of processing artifacts.

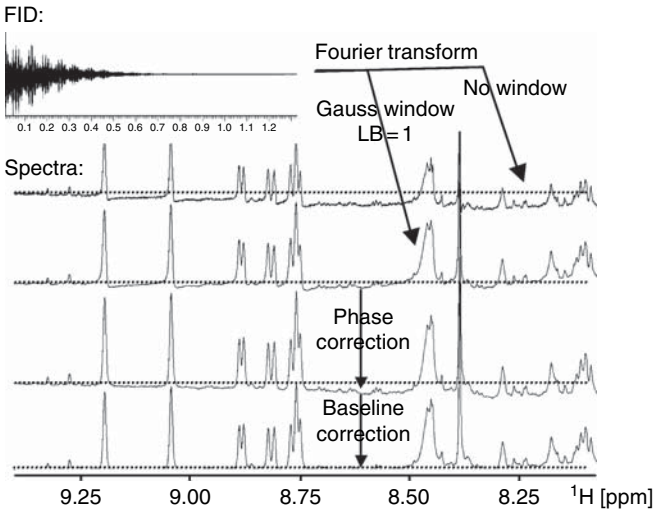

Figure 3.2. Processing of a recorded time-domain signal (top). The FT of the signal is shown second trace: a low signal-to-noise ratio, baseline-offset and dispersive contribution of the lines are seen here. Application of an exponential window function of 1 Hz LB prior to FT results in the third trace. Here signal-to-noise is substantially improved but the other artefacts remain. A spectrum suited for metabonomics interpretation is only obtained after phase correction (fourth trace) followed by baseline correction (fifth trace, bottom). The last two steps have to be inspected visually.

It is of importance that the integral of an NMR signal in the spectrum (not its amplitude) is linear in the number of NMR active nuclei present in a molecular moiety. The concentration of this molecule can be determined from this integral if the signal was assigned to one or a group of nuclei in the molecule. Comparison of integral values with a reference compound or an electronically created standard [22] allows an absolute value (e.g. mg/ml) determination of concentration. More details for spectral processing can be found in Section 3.5 along with the detailed description of NMR experiments.

3.2.1.3. Spin–Spin interaction

The description of NMR based on the Bloch Equation breaks down if spin–spin interaction beyond T_1 and T_2 relaxation is occuring. The detailed description of the physics behind these phenomena is beyond the scope of this introduction. Only a qualitative description of the phenomena will be presented.

J-coupling: Mediated through the spin-orbit coupling of the nuclear spins with the electronic system there is an interaction between two active A and B spins, if both spins share a common electronic orbital. This is often (but not necessarily) fulfilled if both spins are separated by less than 5 chemical bonds in a molecule. As a consequence, the magnetic field seen by spin A is dependent on the orientation of spin B and vice versa. Thus the signals of spins A and B appear as a doublet of 1:1 intensity ratio. The difference between both components (expressed in Hz) is a direct measure of the electronic overlap and thus of the nature and topology of the chemical bond(s) linking both spins.[14] This difference is B_0 independent and is called scalar coupling (constant) J_{AB}. Besides homonuclear scalar couplings (e.g. coupling from ^{1}H to another ^{1}H) there also exist heteronuclear couplings (e.g. ^{1}H to ^{13}C) [23].[15] Normally the number of bonds separating the two spins is provided as additional information: $^3J_{HC}$ refers to a three-bond coupling measured for a ^{1}H nucleus separated from a ^{13}C nucleus by three bonds. If a spin has a scalar coupling to more than one spin, complicated but highly informative coupling patterns are seen in the spectrum. These patterns help to classify spin systems: an AX spin system consists of two spins sharing one coupling and in the NMR spectrum two doublet signals (1:1 ratio) of equal intensity are observed. An AX_2 spin system has three spins: spin A corresponds to one nucleus (e.g. CH), spin X is given by

[14] For a given type of bond (e.g. sp^3 hybridization) the size of the coupling constant is related to the dihedral angle of the bond topology via the so-called Karplus Equation.

[15] Homo and heteronuclear J-couplings (multiplets) can be removed from the spectrum by decoupling. With this technique one or more spins of the spin system are selectively irradiated with a B_1 field (strong compared to J_{AB}); for example at the resonance frequency of spin B. This field will lead to a fast exchange of the up and down state of this spin. Thus the coupled spin A in A-B only 'sees' an averaged state of spin B leading to a collapse of the spin A signal to a singlet.

two nuclei (e.g. CH_2) having identical chemical shifts and coupling topology to spin A. The X spins are called magnetically equivalent and no coupling between the two X spins is seen in the NMR spectra. The spectrum of the AX_2 system is given by one triplet signal (1:2:1 ratio of the three components) with intensity 1 for spin A, and one doublet signal with intensity 2 for spins X_2. More complicated spin-topologies can occur in molecules. It must be noted that a simple interpretation of spin-systems based on their multiplet structure is possible only if the coupling constants involved are much smaller than the differences of their chemical shifts (both in Hz). This regime is called "weak coupling". If this assumption is not valid, the spin system is strongly coupled and only a numerical fit can resolve the individual coupling constants. It is clear that a strong coupling effect at a low B_0 field strength can become weak coupling at higher field strength.[16] An important example of a strongly coupled spin-system in metabonomics is citrate (see Figure 3.10).

Complex spin systems can be resolved by use of 2D NMR methods if individual components are overlapping in the 1D spectrum (see Sections 3.2.1.4 and 3.5.2.1).

Nuclear Overhauser Effect (NOE): A second spin–spin interaction of importance for the richness in information of NMR spectra is the so-called NOE. It rests on the motional average of through-space dipolar interaction between spins.[17] In contrast to J-coupling, NOE interacting spins do not have to share a common electronic orbital – the strength of this effect scales with the spatial distance between the interacting spins. If the NMR signal of a spin A is saturated (see above) by irradiation with a B_1-field, this distortion of signal amplitude will progress to other spins B_i having dipolar interactions with the saturated spin. As a consequence, the amplitudes of the signal detected for B_i spins will also change. This is due to double and zero quantum transitions in the interacting spin system. By bookkeeping of all populations of the different energy levels,[18] a phenomenological equation known as the Solomon equation [24] is derived. The Solomon equation describes the evolution of longitudinal non-equilibrium magnetization $\Delta M_z^{A/B}(t) := M_z^{A/B}(t) - M_0^{A/B}$ of two spins with dipolar interaction:

$$\frac{\partial}{\partial t}\begin{pmatrix}\Delta M_z^{A}\\ \Delta M_z^{B}\end{pmatrix} = -\begin{pmatrix}1/T_1^{A} & \sigma\\ \sigma & 1/T_1^{B}\end{pmatrix}\cdot\begin{pmatrix}\Delta M_z^{A}\\ \Delta M_z^{B}\end{pmatrix} \tag{3.5}$$

[16] On a 400 MHz spectrometer a chemical shift difference of 0.1 ppm corresponds to 40 Hz. This difference doubles at 800 MHz.

[17] If motional averaging is restricted, for example in tissues due to high viscosity, the dipolar interaction manifests itself in line-broadening of signals making their interpretation impossible. This effect can be reduced by application of the HR-MAS technique here the sample is rotated at high frequencies of several kHz. This introduces the motional averaging needed to narrow lines by reducing the strength of the dipole-dipole interaction. Applications of this technique are discussed in Section 2.1.4. A generalization is also possible for powder or crystalline samples.

[18] A two spin AB system has four energy levels, given by the four states up_A-up_B, up_A-$down_B$, $down_A$-up_B, $down_A$-$down_B$.

Details of the parameters T_1 and σ depend on the gyromagnetic moments and the rotational tumbling of the molecule involved. In general the effect is more pronounced for high-molecular-weight compounds. If the spins of interest are both protons, the saturation of spin A will result in a reduction of the signal observed for spin B. The NOE is used for the determination of the 3D fold of molecules, as the effect is proportional to r_{AB}^{-6} with r_{AB} being the distance between both spins. Formally an equation identical to Equation 3.5 is obtained if spin A is in chemical exchange with spin B. This is, for example, the case for protons of urea in exchange with those of water. Thus in full analogy with the NOE, the saturation of water (used for solvent suppression) will influence the amplitude of any proton ($-$OH, $-$NH, $-NH_2$) in chemical exchange with water.

3.2.1.4. Two-Dimensional methods

In 1D NMR the molecule is characterized by a spectral amplitude plotted versus a single frequency axis. This method allows the determination of the composition of a sample only if there is no severe overlap of signals occurring in the spectrum.

In the mid-1970s, a method called 2D NMR was developed [25]. Here a series of 1D NMR spectra is acquired under systematic variation of an experimental parameter of a sequence of B_1 pulses (see Figure 3.3). In most cases the parameter varied is an inter-pulse delay between two pulses (or sequences thereof). During this variable interpulse delay, named t_1, a magnetic coherence state[19] which was prepared at the beginning of the pulse sequence is evolving. Step-wise incrementation of the delay allows the monitoring of the evolution of the prepared spin state. If this evolution is, for example, a chemical shift precession, the chemical shift of the spins involved in this prepared state can be measured. Thereby, the evolving spin state is transferred at the end of t_1 to magnetization resting on a 2nd spin followed by the detection of the modulated NMR signal evolving under t_2. The 2D spectrum obtained after FT with respect to the step-wise increased t_1 value and the t_2 acquisition time will result in a 2D map.[20] Here the spectral amplitude is plotted as a function of two frequency dimensions. A cross-signal in this map will indicate a pair of spins between which a transfer of magnetization has occurred during the pulse sequence. Details of the architecture of the sequence determine which interaction and which types of spins are selected. If both spins are of the same type (e.g. hydrogen atoms) the experiment

[19] Coherences are states of the spin-system which cannot be described in classical terms. The quantum mechanical picture is needed here [10]. Important examples of coherences are the so-called anti-phase state used for *J*-coupling-based transfer of magnetization between spins (see Section 3.5.2.1) and multiple-quantum coherences (see Section 3.5.2.2) evolving with the sum or the difference of the chemical shift of the spins entangled in the state.

[20] Phase sensitive detection is realized along t_1 by always acquiring pairs of experiments for identical settings of t_1. In these pairs, *x* versus *y* direction coherence is the read out at the end of the evolution period.

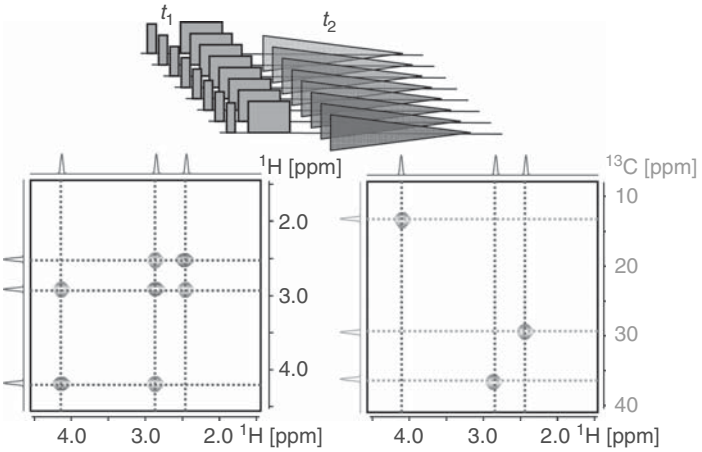

Figure 3.3. Principle of 2D NMR spectroscopy: A set of 1D spectra is acquired under incrementation of a t_1 evolution delay (top panel). FT is performed with respect to t_1 and t_2. A homonuclear shift-correlation spectrum is shown (left bottom). Diagonal signals (in blue) code for magnetization which was initially on a Spin A and was not transferred to a 2nd spin along the sequence. Thus both chemical shifts are identical. The red off-diagonal cross-signals code for magnetization which was transferred by a selected interaction mechanism to a 2nd spin B. It is clear that a heteronuclear shift correlation spectrum (bottom right) only contains non-diagonal cross signals. Here proton magnetization was transferred to carbons (evolution along t_1) and back to protons for detection along t_2. Here axes show the chemical shift of ^{13}C and ^{1}H respectively.

is called homonuclear otherwise it is termed heteronuclear.[21] The 2nd dimension can also code a physical relaxation or a diffusion parameter measured for a certain spin.

Popular examples of 2D NMR which are of some significance for metabolic profiling are discussed in Section 3.5.2. For all 2D methods the detection of the NMR signal has to be performed several hundred times under incrementation of the t_1 duration. Therefore the time needed for such experiments normally comprises several hours up to days. It is clear that this type of experiments normally cannot be done on large sample arrays. It is noteworthy that all that has been said for 1D NMR methods with respect to sensitivity, linearity of detection, methodology of spectral processing can be extended to the 2nd dimension here. Thus relative or absolute quantification of metabolites for comparison of different samples can also be achieved on signals of 2D NMR spectra. Absolute quantification (in mg/ml) is

[21] It is of importance to realize that the sensitivity of the experiment is given by the sensitivity of the nucleus detected during t_2 multiplied by the lowest natural abundance of all nuclei involved in the transfer. This offers the possibility that the chemical shift of carbon nuclei can be detected with a much higher sensitivity compared to the direct measurement of a ^{13}C spectrum, namely via indirect ^{1}H detection of its ^{13}C-coupled ^{1}H nuclei. It has to be kept in mind that the sensitivity of this experiment is still hampered by the low natural abundance of ^{13}C – only 1% of the sample will contribute to the measurement.

more difficult, as the transfer steps of the pulse-sequences involved may reduce the magnitude of the signals because of dependence on many molecular parameters (T_2, T_1, J_{AB} . . .).

Multivariate analysis of 2D NMR data has been described in the literature [26]. Interestingly pseudo 2D spectral information was achieved in metabonomics recently as a spectral map of the correlation coefficient between data points calculated for 1D spectra of a large ensemble of samples of a metabonomic study. Analysis of this map allows for the identification of up- or down-regulated metabolites [27].

3.2.2. Special aspects for biological samples

3.2.2.1. Exchange between different states

The NMR parameters, such as chemical shift, T_1 or T_2, are influenced if molecules exchange between different chemical states. The molecule will have different resonance frequencies for its NMR signals in the different states which are separated by Δf [Hz]. The parameter $\tau = (2 \cdot \pi \cdot \Delta f)^{-1}$ defines the NMR timescale of exchange in seconds.[22] If the exchange time τ_{ex} is slow compared to τ, this will reflect itself in the NMR spectrum as individual lines with amplitudes reflecting the populations of the different chemical states. This situation is referred to as the slow-exchange regime. In the fast-exchange regime given by $\tau_{ex} \ll \tau$ the NMR spectrum will be a single line with a position given by the population weighted average of the chemical shifts of all chemical states contributing. In between, the situation is more complicated and can be described analytically only for the two-site exchange by the so-called McConnell equation [27, 28]. This equation allows the fit to kinetic and stochiometric parameters of the exchange process by detailed analysis of the line-shape. Another interesting case is the so-called coalescence point given by $\tau_{ex} \approx 1.1 \cdot \tau$ where the signals of the exchanging spins will just merge and can become very broad and become nearly invisible in the spectrum.

In this description different states can be represented by different molecules which are interconverting by spin exchange (e.g. protons), or different conformations (e.g. bound and free, or chair and boat forms) of the same molecule which are in dynamic exchange. Chemical exchange is, for example, seen for any signal in a molecule where protons (OH, NH, NH_2) exchange with water. The exchange can also occur between the free-state and a bound-state of a molecule (which is, e.g., seen for citrate forming a non-covalent complex with Ca^{2+} or Mg^{2+}, see Section 3.4.2.3) [30]. The referencing compound normally used in metabonomics TSP (trimethylsilylpropionic acid) has a certain binding affinity to the protein HSA (human serum albumin). As a consequence, the referencing of chemical shift (and

[22] τ is dependent on B_0.

therefore of all lines in the spectrum) will change if samples with different HSA/TSP ratios are compared. Another origin of exchange phenomena is interconversion between different conformations of the same molecule. In this context it is important to realize that the population of a protonated species of a molecule is shifted towards the deprotonated state if the pH of the sample is increased. Thus it is clear that the position of NMR signals of molecules having a titrateable group will depend on the pH of the sample (see Figure 3.10). This effect is compensated for in many applications by introduction of a buffering agent into the biological sample prior to the NMR measurement. It is also clear that any conceivable origin of exchange phenomena can be encountered in complex mixtures such as biological samples. The ideal situation to control the positions of all NMR lines of identical metabolites in biofluids across all samples of a biological study perfectly is therefore approached only asymptotically. Our approach to improve on this topic is discussed in Section 3.4.2.

3.2.2.2. Diffusion

It has been described in Section 3.2.1 that the tumbling of the molecules in mixtures can be used to edit spectra with respect to molecular size. If a so-called T_2-filtration is applied, only NMR signals of low-molecular-weight molecules will be seen in the spectrum (3.5.1.2). Another parameter sensitive to molecular size is translational diffusion characterized by the diffusion constant D_t. In a properly tailored NMR sequence, small molecules with fast Brownian motion can be filtered out of the spectrum [31]. Only macromolecules will be seen in the final data (see Figure 3.18). By systematic variation of a gradient, pulse mixtures can be characterized based on their propensity of different diffusion timescales. This is, for example, highly attractive for the classification of blood samples with respect to different sizes of lipoprotein fractions for which LDL and HDL values reported by classical clinical chemistry are only the most prominent species in a distribution of 14 differently sized fractions with different D_t [32, 33].

3.3. Hardware requirements and automation

The specific demands on the NMR technology and concomitantly on the NMR hardware have been changing over time, although the basic physical principles remain the same. Structural elucidation of large biomolecular structures require multi-dimensional experiments with long data acquisition times depending on complex pulse-sequences and spin physics. But rarely more than a handful specifically prepared samples, for example of a protein are involved [34]. This has changed significantly as the methodological developments has begun to include screening for ligand binding to target proteins [35], and, increasingly also for metabolic profiles of biofluids [8]. Metabonomic studies involve large sample arrays with several

hundred biofluids or cell extracts. As a consequence, the NMR data acquisition needs to be fast. This requires relatively simple but highly robust 1D or simple 2D ^{1}H NMR experiments which are amenable for high sample throughput. Therefore a high level of automation is required starting from sample preparation, followed by automatic transfer of the sample in and out of the magnet, and ultimately also for data processing and evaluation including the build-up of biologically annotated spectroscopic databases. Another recent challenge in hardware development was dictated by the ever increasing need to optimally handle and measure mass- or volume-limited samples from biological sources or from combinatorial chemistry and HT-screening. This has led to the miniaturization of the probehead and the adaptation of the cryoprobe technology to reduced volume flow cells. In the following, the state of the NMR hardware for high-throughput applications in metabolite profiling is discussed.

3.3.1. Magnetic field strength

Most metabolite profiling studies reported in the literature are carried out at high magnetic field strengths ($\geq$600 MHz) to achieve a good spectral dispersion [8]. Nevertheless, the spectral overlap problem in ^{1}H NMR spectra of complex biofluids, especially of urine containing thousands of signals arising from hundreds of endogenous molecules, remains severe, and the spectra are comparable in complexity with spectra of large proteins, [1, 36]. In principle the application of higher magnetic fields (700–900 MHz) would be advantageous for resolving more individual metabolite signals. However, to control the data resolution across a set of many samples in a practical metabonomics study is a challenging problem. This is mainly due to the observed variation of chemical shifts between corresponding metabolite signals as a consequence of varying biofluid compositions, especially of urine samples (see Section 3.4.2). Thus it is often mandatory to reduce artificially the originally obtained spectral resolution of 16 or 32 k data points to a much smaller number of bins in order to gain control over the spectral variable for subsequent multivariate analysis (see Section 3.6). To make full use of the enormous dispersion power of very high field spectrometer, the data points of identical structures across all the biofluid samples of a metabolite profiling study must be related and aligned in one way or another. This remains a challenge for sample preparation (see Section 3.4.2) and post-measurement alignment methods [37–39].

3.3.2. Spectrometer console

The electronic hardware for metabonomics studies should ideally include two independent high frequency channels with full phase and pulse shape control for ^{1}H

(proton) and for ^{13}C (carbon) excitation and acquisition. In addition, a gradient amplifier unit for generation of pulsed field gradients (z-gradient pulses) is mandatory in almost any pulse sequence used today. Pulsed field gradients are applied to select the magnetic spin state(s) (coherences) of interest or to destroy unwanted magnetization without the need of lengthy phase cycling. Pulse field gradient pulses are also important for efficient water suppression by purge pulses. Thus the experiment time can significantly be reduced and spectral quality dramatically increased. Further, signal detection with optimal dynamic range of the electronic receiver system is possible as only the signal of interest is detected.

In addition to higher magnetic fields, also technical means to control and improve the homogeneity of the magnetic field have been constantly improved over the years, for example by new field lock and shim technologies. Most importantly, also the sensitivity has been steadily increased over the years by improved low-noise receiver technology, digital filters, oversampling and notably boosted by new probehead technologies [16].

3.3.3. Probeheads and sample volumes

The probehead is the sensor positioned in the center of the magnet containing a coil that is used both to send radiofrequency (RF) pulses to the sample and to detect the NMR signals returning from the excited atomic nuclei of the sample. The sensitivity of the NMR spectrometer depends, besides the strength of the applied magnetic field, primarily on the inherent sensitivity of the probehead (see Section 3.2.1.1). There are different probeheads available to detect different nuclei with optimal signal-to-noise. For metabolite profiling studies where the endogenous molecules in the biofluids are mostly rather diluted, the probe should be optimized for ^{1}H detection and the NMR observe volume of the probe coil has to be filled completely with the sample to allow for highest sensitivity. In this case the filling factor of the probe is 1. The volume of the sample residing within the boundaries of the NMR active detection region of the coil is called the observe volume. It is always smaller than the overall volume of the sample which is needed to fill the probe such that the sample extends homogeneously over the two coil ends. There are probeheads available with different coil sizes, ranging from 10 to 1 mm in coil diameter and with corresponding sample volumes from a few ml to as low as 2 µl. The standard NMR probehead is usually equipped with a 5 mm coil and requires 550–600 µl sample volume (see Figure 3.4).

The sample volume depends critically on how the sample is transferred to the probehead. It can be provided in a discrete sample glass tube as shown above or by flow injection through a transfer capillary. In the latter case the probe must contain a flow detection cell which is linked to the transfer capillary. Such a probehead is called a flow probe (see Figure 3.5). The sample volume needed for filling

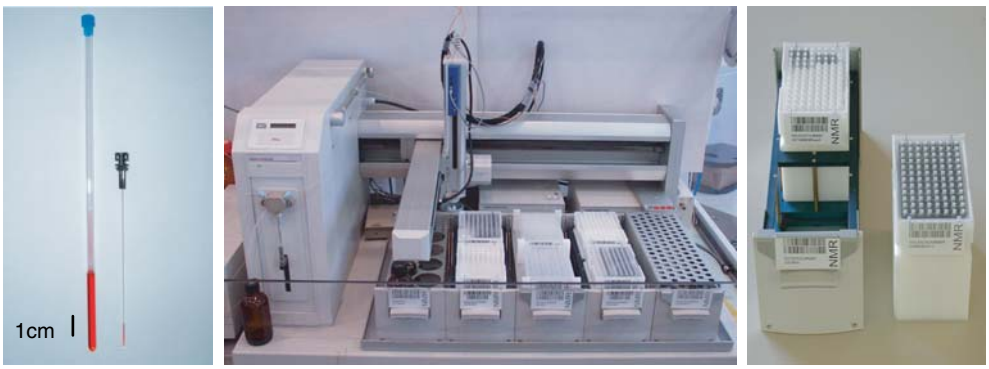

Figure 3.4. *Left:* Standard NMR sample tube with 5 mm diameter (left) and NMR capillary tube with 1 mm diameter (right). The sample volume needed is 550 µl and 5 µl, respectively. *Middle:* Liquid handling Robot (Gilson with NMR micro addition) which can automatically fill NMR sample tubes from 5 mm down to 1 mm in diameter. *Right:* NMR samples are organized in standard 96-well-plate-sized NMR tube racks. These racks can hold, for example, sample tubes and capillaries with 5 mm and 1 mm in diameter, respectively. Each rack is identified by a bar code and each sample by a dot code on the sample cap.

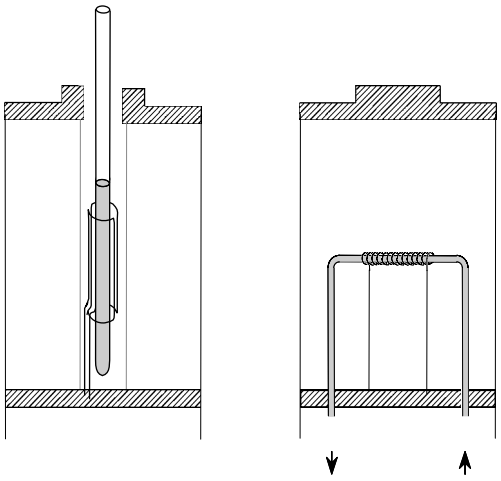

Figure 3.5. Schematic view of the inner part of an NMR probehead. The orientation of the static magnetic field B_0 is vertical. *Left:* NMR detection coil with Helmholtz design filled with a discrete NMR tube. Probeheads are commercially available with coils ranging from 10 to 1 mm in diameter with corresponding observe volumes as small as 2 µl. The Helmholz coil design lends itself also for flow application (direct injection or on-line LC-NMR). Flow probes with various cell volumes down to 30 µl are available. *Right:* NMR detection coil with solenoidal design [18]. Probeheads with this design can only be used for flow application (direct injection or online LC-NMR) as the cell axis is perpendicular to the opening of the magnet and the B_0 field. A probehead equipped with a 5 µl flow cell having 1.5 µl NMR active observe volume (observe volume) is commercially available.

an identical active volume is always larger for a flow cell than by using discrete sampling tubes. In case of the 1 mm coil, approximately two times more sample volume is needed to fill the flow cell (9 µl) than to fill a discrete 1 mm capillary (4.5 µl). The choice of the appropriate transfer method depends on sample properties such as volume, viscosity and solute concentration (see below). Also the sample exchange time varies between the two basic supply modes.

With decreasing diameter of the NMR detector coil, the mass sensitivity (S/N per mole) increases with 1/d to a first approximation [18, 40–42]. To acquire high-quality spectra of volume-limited biofluids, for example CSF from mice, where only a few µl are available, the measurement is ideally done in a 1 mm probehead having an NMR active volume of ~2.5 µl [17] with a fourfold to fivefold increased mass sensitivity when compared to a conventional 5 mm probe. Micro-probeheads are available for discrete tube and flow sampling [17, 43]. Alternatively, the CSF sample could be measured in a cryogenically cooled probe [44] with approximately the same mass sensitivity available for operation with 5 and 3 mm sample tubes (500–200 µl sample volume), either by using the same 1 mm capillary or, after dilution with buffer, in a 3 or 5 mm tube. The cryogenically cooled probehead technology enhances the signal-to-noise of larger sample volumes by a factor of approximately 4 compared to a conventional probe, or alternatively it allows for a 16-fold reduction in measuring time (see Figure 3.6). The high sensitivity of this technology has already notably stepped up the efficiency of drug metabolite identification and structure elucidation [44] . The use of a cryogenically cooled probe will be particularly beneficial in metabolite profiling and biomarker identification where ultimate sensitivity is important. It emerges to be the first choice for sample volumes >10 µl, especially since a convertible cryogenic flow probe with removable

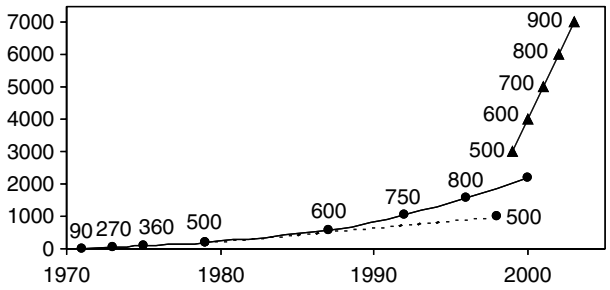

Figure 3.6. The specified signal-to-noise ratio of 0.1% ethylbenzene (EB) in $CDCl_3$ for ^1H-observe probes plotted as a function of time. The black dots denote the sensitivity of a conventional probe at the launch of a magnet operating at a particular field, and the triangles mark the launches of cryogenic probes at different fields (all data from Bruker BioSpin). The magnetic field (indicated by the ^1H operating frequency MHz) is given above the marker. The dashed line indicates the increase in specified sensitivity during two decades for a conventional probe operating at 500 MHz.

insert cells is becoming available [44]. Thus this probe allows switching between flow injection and discrete sample mode.

Important practical accessories to the probehead are a temperature control unit and a motor device which allows for automatic matching and tuning of the resonant circuit of the coil to the impedance of each individual sample.

3.3.4. Robotic sample changer

To monitor beneficial or toxic effects of drug candidates, or to recognize metabolic disease patterns, often demands the screening of hundreds of biofluid samples [8]. In case of small animals such as mice and rats, the biofluids are available only in limited volumes, CSF from 2 to 50 µl, blood from 50 to 200 µl, and urine from $<$1 ml to few ml. The fast supply of large sample arrays of different volumes to the spectrometer magnet is a critical issue in metabolic profiling of biofluids, in biomolecular NMR screening and in structural analysis of combinatorial chemistry products.

As described above there are two fundamentally different ways to supply and exchange the samples to the magnet. Conventionally, glass NMR tubes of 5 mm in diameter, also available as disposables, are filled and supplied to the magnet. This process has been in operation with a low-speed robotic sample changer over 20 years [45, 46]. It is very limited in efficiency for high sample throughput as the time for sample exchange can be much longer than the actual measuring time needed for data acquisition in the magnet. An alternative approach was introduced in 1997 [47, 48] in a study employing urine and samples from combinatorial chemistry. It uses a flow probe in which a direct transfer of samples is possible from a 96 well plate by a flow-injection device resulting in a significantly increased rate of sample throughput. If the washing step of the transfer capillary is not optimal, flow injection may suffer from sample spillover and contaminations or even from bacterial infections within the transfer tube. This could particularly be the case with viscous or concentrated samples, for example blood plasma. An injected sample may extend over 1 m in the transfer capillary and is thus exposed to a huge surface area. When it enters in the wider NMR flow cell, the sample lengths is dramatically ‘contracted’ to a few mm in length with corresponding changes in flow dynamics. Besides this fluidic problem, sample dilution will also occur if a system solvent is used between sample plugs. The recovery of the sample is thus further hampered. As previously discussed, flow injection needs a sample volume to fill the cell which is approximately two times larger than the flow cell volume. This is due to the large dead volume of the flow system. In the case of discrete sampling with tubes or capillaries, there are only minimal dead volumes involved and the samples are completely shielded from each other. This difference may become significant if only limited and small sample volumes of biofluids are available, for example from small animals.

In the past few years, both discrete and direct flow-injection sampling methods have been further developed and miniaturized. This development occurred in parallel with the miniaturization of the NMR detection coil of the probehead [17, 43]. Today, discrete 1 mm NMR sample capillaries can easily, quickly and reproducibly be filled with 5–8 µl sample by a liquid handling robot. Also automatic flow-injection systems are in place for handling sample volumes as small as a few microliters.

A very promising development has recently led to a new innovative sample changing system (Figure 3.7) called SampleJet (Bruker Biospin). It combines high speed and throughput with the advantages of handling discrete NMR sample tubes having sample volumes from 600 to 5 µl and sample diameters from 5 to 1 mm, respectively. The NMR samples are spatially organized in the standard well-plate format that lends itself ideally for high-speed, automated sample handling systems. Sample exchange times as short as 30 s can be achieved. There is room for even higher throughput using this platform. The exchange time is thus comparable to those achieved with fast flow-injection systems. Continuous NMR sample tube automation is available in the microtiter-plate format including automated liquid and tube handling coordinated by NMR automation software.

3.3.5. Connecting lab bench and NMR spectrometer

The NMR-based metabolite profiling of biological samples might need automation which includes in addition to the NMR measurement also the time-coordinated preparation of the sample. This needs to be the case if the biological sample has

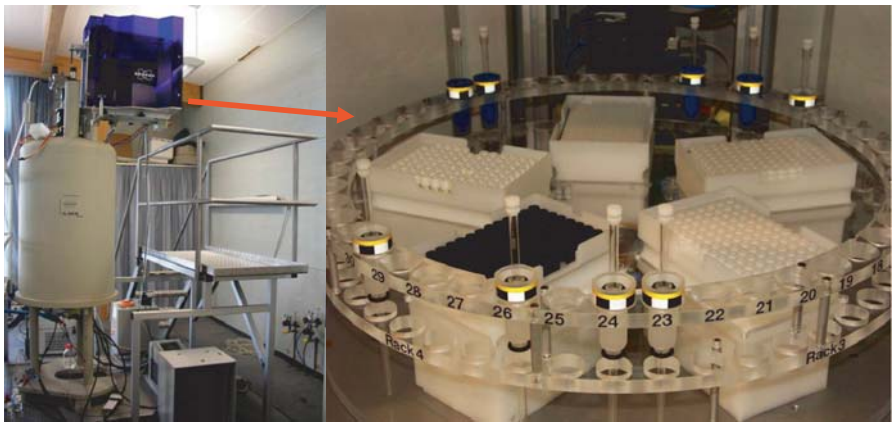

Figure 3.7. A sample changing robot (Sample Jet) allows fast sequential single tube submission under controlled conditions from five positions each holding a 96-well-plate-sized NMR tube rack. The system can thus efficiently handle batches with up to 480 samples. Also manual sample handling with standard NMR tubes and turbines is possible (outer circle).

to be prepared freshly, that is just-in-time for measurement and is not allowed to stand in a waiting queue in order not to deteriorate. In this case, automation should include the following individual steps:

- Just-in-time sample preparation.
- Transfer of the sample to the magnet.
- Setup of the NMR apparatus including locking of the field and shimming.
- Measurement of NMR experiments comprising a set of selected 1D and possibly 2D techniques.
- Back-transfer of the sample to a park position outside the magnet and storage.

Computerized book-keeping is absolutely compulsory in order not to lose track of hundreds or thousands of NMR experimental runs. The need for high throughput and reliability for all steps involved can hardly be over emphasized. A robotic system which has been described and successfully used on thousands of sample in biostructural NMR screening can also be employed in biofluid screening [49, 50]. It consists of a Genesis sample handling robot (Tecan) which prepares the sample by mixing the required components in an NMR tube immediately before it gets automatically transferred to the spectrometer by a Bruker SampleRail system.

3.4. Sample handling

NMR spectroscopy on native biological samples differs in many aspects from conventional NMR spectroscopy on material from synthetic sources. Many parameters must be considered to obtain relevant NMR spectra of good spectral quality. Potential sources of any sort of artifacts have to be consistent and to be avoided by all means during the whole process from collecting the biological sample until the analytical result is provided for further data processing.

3.4.1. Sample collection

Factors, such as sample type, time of collection, containers used, preservatives and other additives, transportation and length of transit time affect the quality of the samples and must be carefully considered before the initial collection stage [51]. In addition the design of the metabolic cages used for animal studies as well as the protocol for sample handling at the animal housing facility influence the sample quality. The metabolic cage for animal housing must prevent feces and food from entering the urine collection container. For urine, cooled sampling units attached to

the metabolic cage are necessary. Microbial degradation[23] during sampling intervals of several hours can be dramatically reduced by such devices. Therefore it is absolutely essential to set up a proper study protocol with clear instructions for all aspects of the sample collection process to ensure a reproducible high quality of samples.

In the following, we focus on the factors affecting the quality of biological samples and some of the provisions that must be made during collection, processing and storage of samples, based on our experience with a focus to urine samples, which are with respect to their variability for obvious biological reasons most demanding.

3.4.1.1. Collection containers

The choice of the containers for sample collection is an extremely important prerequisite for a successful realization of metabolite profiling by any analytical techniques. A wide range of different types of containers for sample collection are available. Examples of spectra of different types of blood sample tubes leached with phosphate buffered water are depicted in Figure 3.8. It has to be pointed out that tubes of

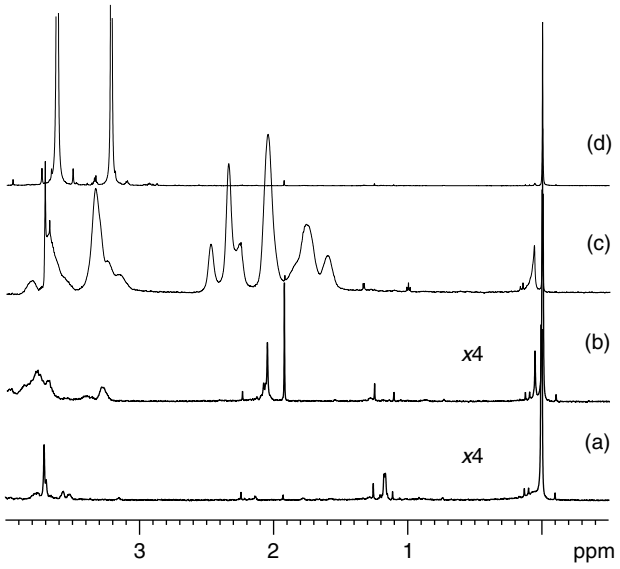

Figure 3.8. ^{1}H NMR spectra of a cross section of sample containers for biofluids of different manufactures. All tubes were leached with buffered water (pH = 7.4). Spectra were originally scaled to same TSP intensity and enlarged as indicated. (a) native PP tube, (b) Li-Heparin coated tube for plasma samples, (c) PP tube coated with a clot activator for serum samples, (d) EDTA coated tube for plasma samples. Spectra for the same tube type but from different manufacturer may vary significantly.

[23] In the collection containers of cages often 1 ml of a 1% solution of NaN_3 solution is provided. It is often overlooked that this sample collection scheme typically adds a concentration of 160 mM Na^+ to urine samples.

the same type, for example clot activated tubes for serum collection of different manufactures or different batches of the same supplier, can show markedly different additive profiles. It can be clearly seen that native tubes without any additives or stabilizers generally show least interfering signals in the ^{1}H NMR spectrum. To test the container (of the same batch that will be used for the study) for possible contaminants before sample collection is therefore strongly recommended. Greatest care has to be taken to select the right sample tube for metabolic profiling studies because analysis of data can fail if contamination by inadequate sample tubes is introduced.

Urine and CSF: For urine or CSF collection no special functionality is needed for the sample container. Here native tubes without any coating, made from glass or polypropylene (PP), are strongly recommended. This is of advantage as more functionality incorporated in the tube holds the risk to render the sample useless for metabolite profiling.

Serum and plasma: Typical sample tubes for blood collection often contain additives or stabilizers that simplify further sample handling or are needed for the preparation of the desired serum or plasma sample. Often the decision whether to collect plasma or serum samples is influenced by practical considerations at the sample collection site. More important for metabolic profiling investigations is the spectroscopic suitability of the finally obtained biofluid. For plasma and serum collection the type of sample container must be carefully evaluated (see Figure 3.8).

Plasma is a cell-free supernatant of anti-coagulated blood and can be obtained fast and in high yields. The risk of uncontrolled and incomplete clotting is low. Therefore plasma is often the favored blood derivative. Tubes containing unspecified anti-coagulants should be avoided (see Figure 3.8) unless they are completely free of any ^{1}H NMR signals. In most cases, Li-Heparin or EDTA (**E**thylene-**d**iamine-**t**etraacetic **a**cid) is used as anti-coagulant. Both types of tubes show signal overlap of endogenous metabolites with EDTA and Heparin resonances in the NMR spectrum. Ideally, per-deuterated EDTA is used as anti-coagulant as this compound is completely free of ^{1}H NMR signals. However, such tubes are not yet commercially available, but can easily be prepared. As a second choice, Li-Heparin tubes are recommended.

Serum is obtained from whole blood by centrifugation after clotting is completed. This process takes about 30 min and may not be exactly controlled at the collection site. The advantage of serum is that no additives are necessary to obtain this biofluid. Native glass tubes can be used for serum collection. The time samples are allowed to clot naturally at room temperature and the cooling chain during preparation and transportation has to be controlled and monitored carefully. Tubes coated with unknown clot activators generally are not recommended for metabolic profiling studies because contamination can be introduced (see Figure 3.8).

3.4.1.2. Stability and storage

The consistency of the metabolite profile across many samples in a study depends on the quality of the biological samples. Not only inappropriate sample containers may lead to unwanted variation but also unequal treatment or storage of individual samples of a study. For example, the study protocol must define maximum allowed storage periods for samples at room temperature, as many endogenous metabolites are sensitive to chemical or microbial degradation. A long total storage time between sample collection and analytical measurement can induce variation in the data. Investigations during the Consortium for Metabonomic Toxicology (COMET) project [52] revealed that biological samples like urine are stable for at least 9 months at −40°C. Slight biochemical changes in tricarboxylic acid (TCA) cycle intermediates were found in urine samples after an 18 months storage period at −40°C. For plasma samples the influence of short- and long-term storage at various temperatures was assessed in detail by Deprez *et al.* [53]. The authors reported no observable changes in plasma NMR spectra after 6 months storage at −80°C.

3.4.1.3. Microdialysis

Microdialysis is an established *in vivo* tool in neuroscience and has been used widely for pharmacological and metabolic profile analysis. The dialysate which is collected for analysis represents the local profile of the extracellular environment of a specific tissue or biofluid. For microdialysis a probe with a hollow fiber dialysis membrane is implanted into the organ or biological matrix. The perfusate, a solution that mimics the physiological composition of the extracellular fluid of the target, is slowly pumped through the probe. Small molecules can diffuse through the membrane and are carried by the dialysate. In most cases the microdialysate is free of large molecules like proteins, but also small molecules bound to proteins are excluded by the membrane [54]. Therefore ^{1}H NMR resonances of endogenous metabolites of microdialysate samples are not obscured by broad lipid or protein signals [55]. This facilitates data interpretation. In addition, microdialysates are metabolically more stable than native biofluids or tissues because enzymes are not able to pass the membrane. A drawback of microdialysis from an analytical point of view is the small volume of dialysates and the low concentration of analytes present in the dialysate. Therefore the solvent is completely removed by vacuum evaporation in most applications [56] and measurements have been performed in miniaturized NMR probes [17, 57].

3.4.2. Sample preparation

Sample preparation must fulfill the following criteria: (i) reproducibility, (ii) robustness, (iii) ease of use, (iv) non-discrimination between samples, (v) reduction of

unwanted variation in the data, due to different pH of samples and so on, (vi) maintenance or improvement of spectral quality regarding resolution and sensitivity, (vii) freedom of analytically visible artifacts.

In contrast to other analytical techniques [58] or clinical chemistry methods, metabolite profiling of biofluids by NMR like urine, plasma or CSF requires in general less sophisticated sample preparation procedures [59]. In most cases, addition of water or buffer to account for pH variation or to reduce viscosity is sufficient as sample preparation before the NMR measurement [60]. This means that one potential source of error, for example due to sample extraction procedures, is absent. However, a homogenous data set is highly desirable for further data processing and data evaluation steps.

3.4.2.1. Concentration and lyophilization

Lyophilization and reconstitution in an appropriate solvent is frequently a practical sample preparation method for samples of biological origin to enhance the analytical sensitivity and spectral quality, or to stop the inherent enzymatic activity of biological samples. For CSF, significant sharpening of small molecule resonances was reported after lyophilization [61]. However, the loss of volatile or unstable components from biofluid samples may occur during the lyophilization process, for example for acetate or formate. Selective deuteration of acidic protons on reconstitution with D_2O may further complicate spectral interpretation. In Figure 3.9 ^1H NMR spectra of human urines are depicted where the effect of lyophilization and concentration was investigated. Aliquots of urine were directly measured after buffering (a), reconstituted in buffer to the original volume (b), and after lyophilizing the sample and concentrating by a factor of 4 (c). The native and lyophilized spectrum show minute but visible differences in metabolite compositions. For example, besides other subtle changes resonances of acetate, dimethylamine and succinate were reduced after lyophilization. Due to higher viscosity the fourfold concentrated sample reveals slightly increased line widths. If sensitivity is not an urgent issue, lyophilization and concentration should be considered with care. Every sample preparation step is a direct intervention into the sample constitution and bears a risk for both loss of metabolites and contamination. It is the power of NMR spectroscopy not to discriminate between analytes and to need only sparse sample preparation steps. This has to be critically compared to the advantages of lyophilization.

3.4.2.2. pH and buffering

The common sample preparation protocol for urine samples includes pH-adjustment with phosphate buffer (e.g. pH 7.4) and the addition of D_2O, TSP and sodium azide to final concentrations of about 5%, 0.3 mM and 1 mM respectively [8]. The pH of different urine samples may vary from 5 to 8, according to the physiological condition in the individual, but usually it lies between 6.5 and 7.5. The

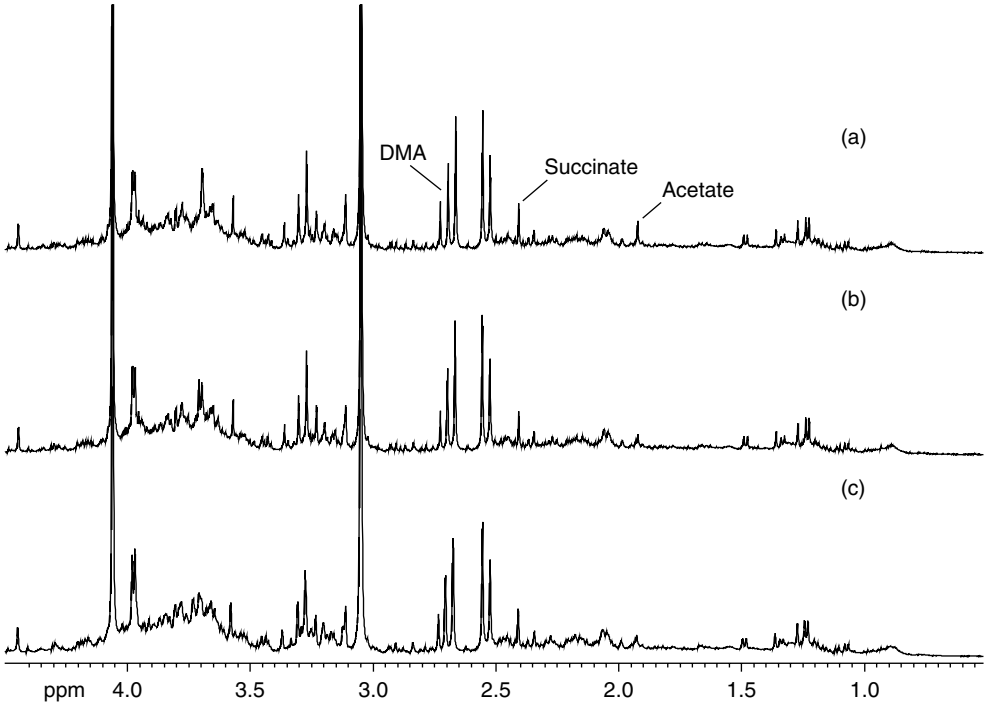

Figure 3.9. Comparison of ^{1}H NMR spectra of: (a) native human urine; (b) lyophilization and reconstitution to original concentration; (c) lyophilization and reconstituted with fourfold concentration. The spectra are scaled to the creatinine intensity.

addition of 100–200 mM phosphate buffer normalizes the pH in most cases to a range of 6.7–7.6, at least for several hours [59]. Nevertheless several small molecular endogenous metabolites present in biofluids show a pronounced pH-dependent chemical shift variation in ^{1}H NMR spectroscopy even after buffering [62]. This is, for example, the case for citrate and histidine. The dependency of the citrate resonances on pH is depicted in Figure 3.10. The remaining chemical shift variation (see also Section 3.2.2.1) between different samples has a direct impact on data reduction procedures and complicates further data evaluation steps. It may thus influence the data interpretation. Therefore, the binning of each spectrum into segments with sufficiently large spectral width was introduced [63] to scope with the shift variation (see also Section 3.6.2).

3.4.2.3. Salt concentration, EDTA

Adjustment of pH of biological fluids alone does not fully remove the chemical shift variation of organic acids like citrate [62], lactate, taurine or others. Further effects related to osmolality of the samples, ionic strength or metal ion composition

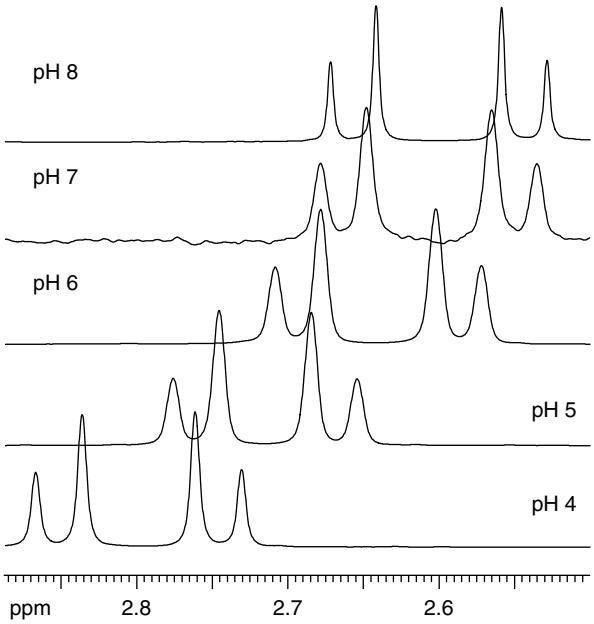

Figure 3.10. Variation of ^{1}H NMR citrate chemical shifts at different pHs ranging from pH 4–8.

significantly contribute to the pH-independent chemical shift variation (see also Section 3.2.2). This is especially the case not only for urine and plasma but also for CSF or saliva [30].

The inorganic metal ion composition of biological fluids, especially of urine, shows a great variance between different individuals and also between different time points of an individual. The major positively charged urinary inorganic constituents are the cations of sodium, potassium, calcium and magnesium. Typical concentrations for these alkali and earth alkali elements in human urine are 224 mmol/d, 70 mmol/d, 5.9 mmol/d and 5.4 mmol/d respectively [64]. Citric acid is known for its avid chemical propensity to bind metal ions like calcium and magnesium, and to a smaller extent also sodium. High-resolution ^{1}H NMR spectroscopy can readily distinguish the various complex/chelate states of endogenous metabolites via the assessments of precise frequency or line widths [30]. In Figure 3.11, the change of the chemical shifts of citric acid in the presence of different sodium concentrations is depicted.

Bivalent calcium and magnesium ions influence the citrate chemical shift more pronouncedly than the sodium cation (see Figure 3.12). Besides pH dependency, the different Ca^{2+} and Mg^{2+} concentrations of biofluids are the major source for the chemical shift variation of citrate and other organic acid anions observed in biofluids. Interestingly, only the chemical shift of one of the two protons of the AB CH_2 spin system of citrate is markedly influenced by added Mg^{2+}, whereas Ca^{2+}

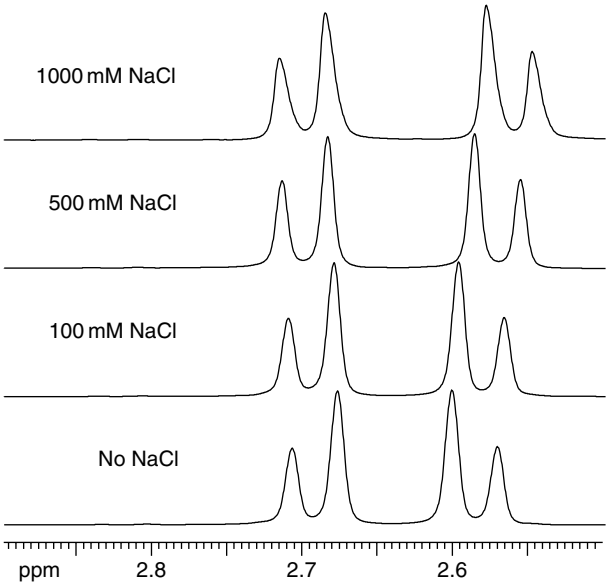

Figure 3.11. Variation of ^{1}H NMR citrate chemical shifts against change of the sodium chloride concentration (ionic strength) referenced to TSP.

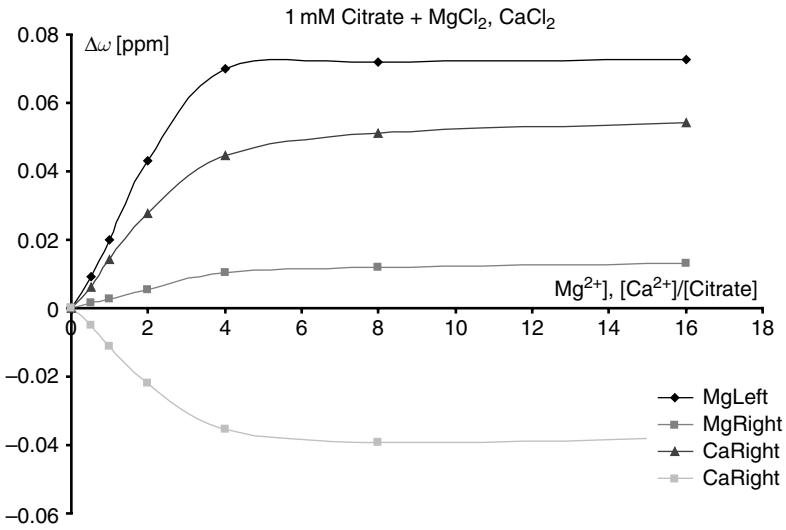

Figure 3.12. Plot of modification of citrate-CH_2 group chemical shift values referenced to TSP ($\Delta\omega$) versus added Ca^{2+} and Mg^{2+} concentrations. Left and Right respectively denote the deshielded and shielded resonances of the AB group of citrate.

changed both resonances of the AB spin system. This may be attributed to different stereochemistry of the Ca^{2+} and Mg^{2+} complexes due to differences in ionic radii (100 pm versus 72 pm [65] respectively).

From an analytical and data evaluation point of view, a homogenous sample set with respect to overall concentration, pH, metal ion composition and osmolality would be highly desirable, but for practical reasons the sample preparation can not account all these sources of variation in biofluids. For example, the adjustment of all samples of a study to the same overall concentration can only be done with respect to the most dilute sample in the series. This would lead to drastically increased experimental acquisition times. As a compromise we propose the addition of per-deuterated EDTA to account for different earth alkali cation composition of biofluids in addition to buffering. This is a simple modification of the sample preparation procedure that fulfills most of the criteria mentioned in Section 3.4.2. The two major sources (pH and bivalent metal ion composition) of variation in chemical shifts can be reduced. The span of chemical shifts for citrate can be dramatically decreased and peak overlap in different bins after data reduction can be avoided. This helps for statistical data analysis and simplifies the binning procedure (see Section 3.6.2). Figure 3.13 shows the effect of addition of 10 mM EDTA-d_{12} to 80 samples of phosphate-buffered control rat urines. It can be clearly seen that not only was the variation in citrate chemical shift minimized but also the potential overlap with a resonance of dimethylamine at $\delta = 2.706$ ppm was prevented. This modified procedure mainly affects the chemical shifts of chelating metabolites like citric acid, taurine or lactic acid. The remaining signal of not fully deuterated EDTA is negligible and does not interfere with further data analysis steps.

If creatinine is taken as a measure for urinary concentration[24] and citrate chemical shifts are plotted against urinary concentration (creatinine) in a regime free of interaction between endogenous components, there should be no correlation visible. However, the situation depicted in Figure 3.14 clearly shows a concentration dependency of citrate chemical shifts in both cases with and without addition of EDTA-d_{12}. The variation of citrate resonance is reduced by a factor of about 2.4 in standard deviation after addition of EDTA-d_{12}. The slope of the trend line indicates remaining interaction modes of the non-chelated citrate with urinary components which is concentration dependent, although less than in absence of EDTA.

Besides the reduction in citrate chemical shift variation, the overall explained variance of a Principal Components Analysis (PCA) model is a measure for the effect of EDTA-d_{12} addition. In Figure 3.15, the convergence of the explained variance of PCA models of 96 human urines with and without addition of EDTA-d_{12} is shown. The

[24] This is valid to a good approximation for samples, where a drug induced change of creatinine signal can be safely excluded.

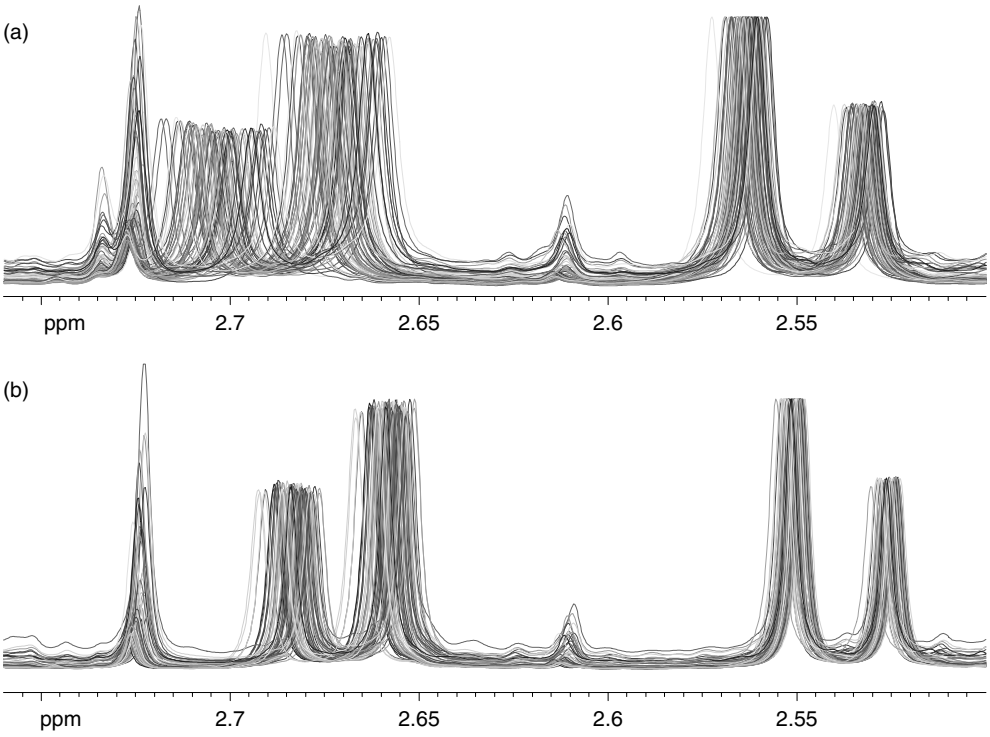

Figure 3.13. (a) Citrate chemical shift region of 80 rat urines buffered with 100 mM phosphate buffer (pH 7.4). (b) Citrate chemical shift region of the same 80 rat urines buffered with 100 mM phosphate buffer and addition of 10 mM EDTA-d_{12}.

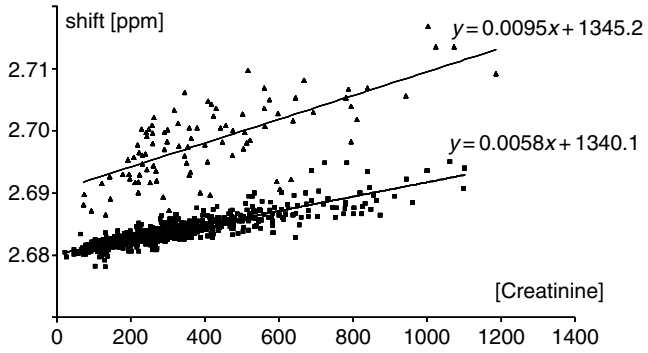

Figure 3.14. Correlation of citrate chemical shift (centre of the low-field doublet) with (lower trace) and without EDTA-d_{12} (upper trace) against urine concentration represented by creatinine intensity. The standard deviation for citrate chemical shifts is reduced from ±3.1 Hz to ±1.3 Hz after addition of EDTA-d_{12}, also the slope of the trend line is smaller.

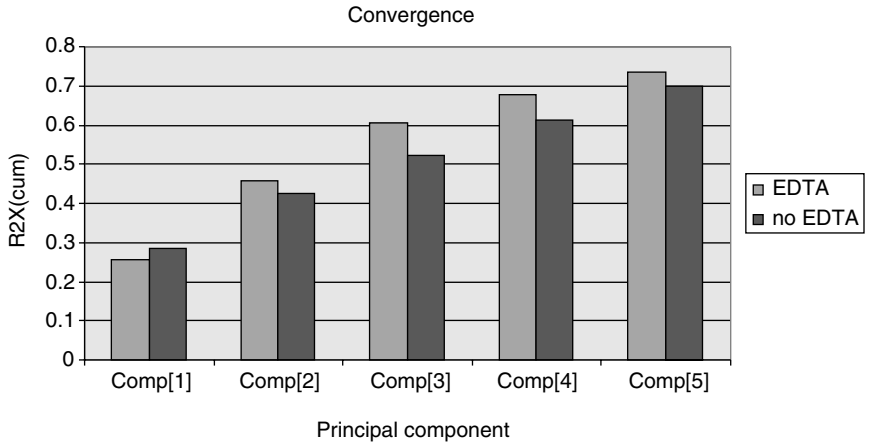

Figure 3.15. Convergence of explained variance of PCA with integral normalized samples with and without EDTA-d_{12} addition.

explained variance increases faster after addition of EDTA-d_{12}, therefore, the beneficial effect of decreased citrate variation is beneficial also for the complete data set. It can be seen that by removing one source of non-dose-dependent variation from the data, the PCA models converge faster and statistical analysis should be facilitated.

3.5. NMR Experiments and their processing

Compared to other applications of NMR, the pulse sequences applied in metabonomics require a lower level of sophistication and the application of all sequences is hampered by the following reason:

All NMR experiments which rely on the detection of hydrogen nuclei (protons) in biological samples face the problem that molecules have to be detected in an aqueous sample at μM or lower concentrations against a background concentration of 110 M water protons. Thus a dynamic range of the receiver and digitizer system of 10^8 would be needed. Modern state of the art receivers allow for 16-bit signal digitization, ensuring a dynamic range below 10^6. Thus a suppression of the water signal of at least 10^2 is required such that a signal of a compound at low concentration "fills at least a single bit". A number of water suppression techniques have been reviewed in the NMR literature [66]. In the following, solvent suppression techniques and related pulse sequences that are applied in metabonomics analysis will be described. The technique used for water suppression is of high importance for the comparability between NMR data. Results obtained from multivariate analysis of NMR spectra taken under different solvent suppression conditions can be erroneously interpreted as biological differences [62] if water suppression artifacts are not carefully excluded.

3.5.1. 1D NMR

Most applications of NMR for metabonomics research rely on 1D NMR experiments. This is dictated by long acquisition times of most multi-dimensional NMR experiments making their application a time-consuming task for studies with large sample arrays. This situation is normally encountered for time- and dose-resolved studies with preclinical or clinical samples. Typical 1D NMR spectra of serum, urine and CSF are shown in Figure 3.16.

3.5.1.1. Water supression

Presaturation: The most widespread solvent suppression in use is the so-called presaturation technique [67]. This technique is applied in an NMR pulse method known as noesy-presat [68]. This pulse sequence is identical to the 1st time increment of the 2D NOESY[25] experiment (Figure 3.17). The water resonance is saturated by

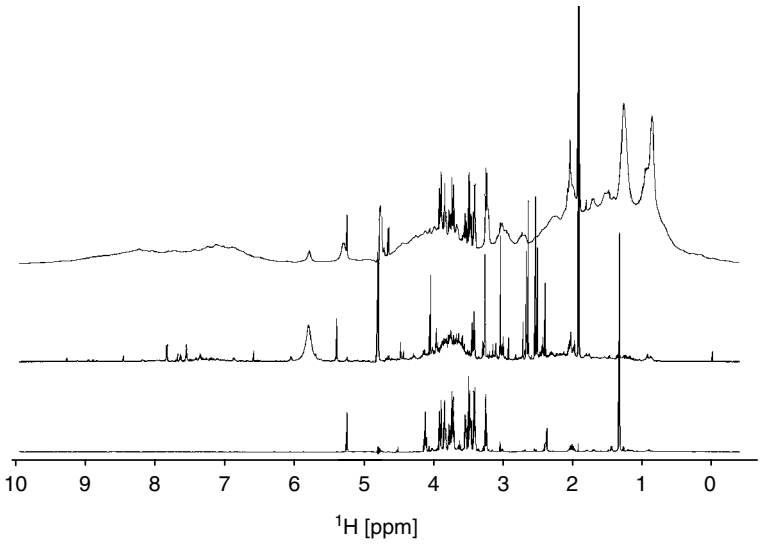

Figure 3.16. 1D ^{1}H NMR spectra from the top of serum, urine and CSF of rat. Very different spectral signatures of the body fluids are clearly visible. The overall shape of the serum spectrum is dominated by a broad signal background having its origin in high-molecular weight compounds (proteins, lipoproteins). The eye-striking feature of urine is a huge urea signal at around 5.9 ppm, whereas the CSF spectrum is dominated by high glucose (3–4 ppm) and lactate (1.2 ppm) content.

[25] **N**uclear **O**verhause**r** **S**pectroscop**y** (NOESY): Here a 2D spectrum is acquired allowing for the determination of distances between protons within a molecule. The NOESY is widely applied in structure determination of proteins by NMR[1]. The technique has found no application in metabonomics.

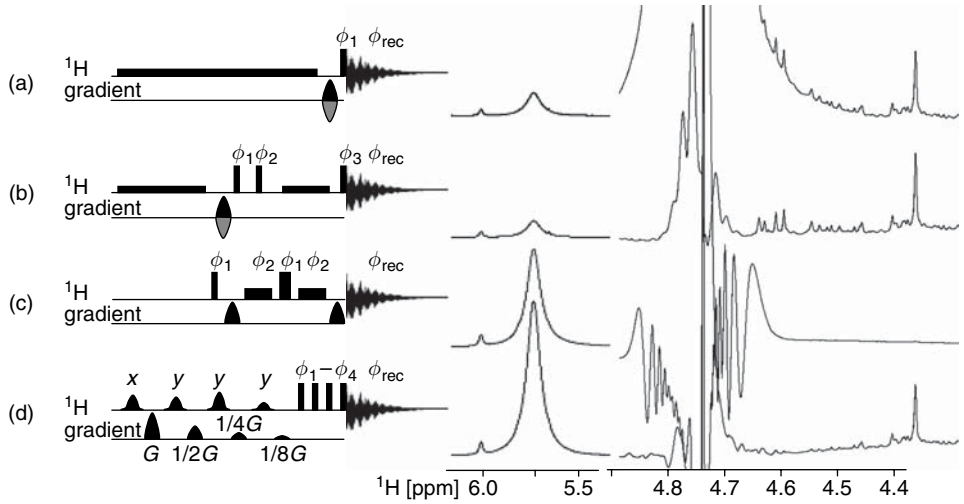

Figure 3.17. Different methods of water suppression: From top to bottom the following methods together with typical results obtained for a urine sample are shown: (a) *standard water presaturation* of 3 s with 50 Hz (phase cycling as follows: $\phi_1 = \phi_{rec} = \{x, -x, -x, x, y, -y, -y, y\}$); (b) *1D noesy-presat* during 3 s and 100 ms mixing time with 50 Hz (phase cycling: $\phi_1 = \{x, x, -x, -x\}$, $\phi_2 = \{16x, 16(-x)\}$, $\phi_3 = \{4x, 4(-x), 4y, 4(-y)\}$, $\phi_{rec} = \{2x, 4(-x), 2x, 2y, 4(-y), 2y, 2(-x), 4x, 2(-x), 2(-y), 4y, 2(-y)\}$); (c) *WATERGATE* with 1 ms selective pulses (phase cycling: $\phi_1 = \phi_{rec} = \{x, y, -x, -y\}$, $\phi_2 = \phi_1 + 180°$); (d) *WET* with 10 ms selective shaped pulses – power levels of pulses compared to a 90° shaped pulse are $+0.87, -1.04, +2.27, -5.05$ dB counting from left to right (phase cycling: $\phi_1 = \{y, 2(-y), y, -x, 2x, -x\}$, $\phi_2 = \{-x, x, -x, x, -y, y, -y, y\}$, $\phi_3 = \{-y, 2y, -y, x, 2(-x), x\}$, $\phi_4 = \{x, -x, x, -x, y, -y, y, -y\}$, $\phi_{rec} = \{2(x, -x), 2(y, -y)\}$). In the pulse schemes (a)–(d) shown to the left, pulses on the top trace represent proton pulses (proton frequency). The heights of the pulse go in parallel with the applied B_1 field strength. The bottom trace shows gradient pulses applied, whereby all gradients are applied with sine shape. The black/grey shaded gradient pulses indicate application with alternating sign of this pulse. The spectral region shown in the middle column shows the size of the urea signal. Urea protons experience chemical exchange with water. Therefore a substantial suppression of this signal is seen for pulse methods which are "slow" on the exchange timescale (presaturation for 3 s in (a) and 2 s in (b) versus selective pulses of a few milliseconds for WATERGATE and WET). Therefore quantification of urea is not possible and the signal is typically excluded from analysis. Note: We have included here the phase cycling scheme that was employed. The addition of scans is performed with different directions of applied pulses. This has to be done in concert with phase cycling of the receiver to obtain coherent addition of signals. Phase cycling is applied for these techniques to remove artifacts due to spectrometer imperfection and relaxation phenomena accruing during the course of the pulse sequence.

frequency selective irradiation during the inter-scan delay and the mixing time of the sequence thus leading to a reduction of the equilibrium population difference of the spin species resonating at the frequency of the weak (<50 Hz) B_1 field applied. The big advantage of this method is given by a chemical shift selectivity

superior to any other technique. If a reliable quantification of signals resonating close to water (e.g. anomeric proton of glucose) is needed, this technique will be the choice. The disadvantage of presaturation comes from the long duration (typically 2s). Therefore hydrogen atoms in chemical exchange with water (OH, NH, NH_2) (see Section 3.2.2.1), protons having an NOE (see Section 3.2.1.3) with a proton under the water line, or which are exposed to a hydration shell will also be saturated. The degree of saturation is given by the rate of the chemical exchange and the strength of the NOE respectively.[26] Any interaction process slow on a 2s timescale can be neglected. The most prominent signal in urine exposed to saturation exchange is urea, which cannot be quantified if water presaturation is applied (see Figure 3.17). This spectral region is excluded normally from metabonomics analysis. The mixing time, together with the intense phase cycling used in noesy-presat, has a positive effect on a flat baseline. This can be explained by a small degree of spatial selectivity[27] encoded in the noesy-presat pulse sequence reducing the influence of protons entering or leaving the sensitive volume of the detection coil during the detection of the FID. It must be noted that a mixing time will falsify the absolute amplitudes of NMR signals detected due to differential T_1, NOE and chemical exchange effects.

Experience in our laboratory shows that for cryogenic probeheads offering higher sensitivity the quality of water suppression achievable by this technique is poor when measuring urine in 5 mm sample tubes. In this case we have successfully used a sign-alternating gradient pulse applied between the end of the presaturation and the first 90° pulse to significantly improve water suppression.

In summary, the noesy-presat technique provides a simple, highly reproducible and robust method for the acquisition of high-quality NMR spectra in aqueous solutions. It has to be kept in mind, however, that absolute value quantification is possible only with limited accuracy.

Gradient pulse based technique: This water suppression technique rests on the application of the so-called gradient pulses. A gradient pulse G_z (typically 1 ms) creates an inhomogeneous magnetic field along the z-axis. Magnetization of amplitude M_0 present in the transverse plane prior to the pulse will be dephased efficiently by G_z, as the frequency of the precession is z-dependent during G_z. Only the integral along the z-direction of the sample given by $M_0 \int_{-z/2}^{+z/2} e^{i \cdot \gamma_x \cdot G_z Z'} dZ' \approx 0$ is detected by the receiver. If the gradient is strong enough (>10 G/cm), the signal is

[26] This phenomenon is referred to as saturation transfer.

[27] Spatial selectivity of B_1 pulses is obtained, as the distribution of the B_1 field strength is not homogeneous over the sample. Therefore, long and/or many B_1 pulses will lead to a reduction of signal amplitude. The region of highest B_1 inhomogeneity is found at the upper and the lower rims of the sample. Therefore contributions of these regions (also prone to high B_0 inhomogeneity and diffusion effects) are suppressed most efficiently by this technique.

dephased by more than 99%.[28] For the WET (water suppression enhanced through T_1 effects) [69] sequence which is used in metabonomics the transverse magnetization of the solvent signal is thus suppressed by the application of a series of weak frequency selective B_1 pulses, interleaved by G_z pulses. The flip-angles (durations) of the selective pulses and the strengths of the gradient pulses were optimized to achieve high-frequency selectivity, robustness and a flat baseline. The performance of the sequence is further improved by the application of a composite 90° excitation pulse as the last pulse taking advantage of its higher spatial selectivity [27]. Effects from chemical exchange and NOE are less serious for WET, as the overall duration of the sequence is more than one magnitude of order shorter compared to the duration of presaturation pulses described in above. The drawback is a reduced performance in respect to frequency selectivity, having its origin in the higher B_1 field strength typically used. All other parameters are essentially the same as in noesy-presat. The absence of the mixing time allows for an accurate quantification here. As a consequence, data taken with WET and noesy-presat cannot be analyzed statistically in a combined data set but must be treated separately.

3.5.1.2. Editing techniques

Relaxation editing: Rigid molecules (e.g. proteins) with high molecular weight experience a fast transverse relaxation (see Section 3.2.1). This is exploited in metabonomics for the removal of signals of high-molecular-weight compounds. Thereby a 40–100 ms relaxation delay [70, 71] is included prior to data acquisition. J-coupling effects are most efficiently removed by use of the so-called CPMG (named after its inventors **C**arr **P**urcell **M**eiboom and **G**ill) [72] pulse sequence with a refocusing delay set to 1 ms. This technique is especially attractive for blood serum or plasma samples having high content of proteins (see Figure 3.18). For applications on soft tissues the technique was successfully applied using a HR-MAS probe-head [73].

Diffusion editing: The principle of gradient pulses was introduced in Section 3.5.1. The application of a pair of z-direction gradient pulses of opposite sign will result in a dephasing of transverse magnetization followed by a rephasing. No signal is lost. This only holds true if the magnetization in a certain z-plane of the sample has not moved to another z-plane between the application of the pair of gradient pulses. If a delay is inserted between the two gradients, rephasing is only partially possible due to Brownian motion and depending on the molecular weight of the solute molecule. This phenomenon is used to remove signals from "small" molecules

[28] Another important application of gradient pulses is the selection of coherence pathways. This technique allows, for example in HSQC type experiments, for the selection of signals of protons bound to ^{13}C, and artefacts due to protons bound to ^{12}C are removed. For details, the reader is referred to citations in Section 3.5.2.2.

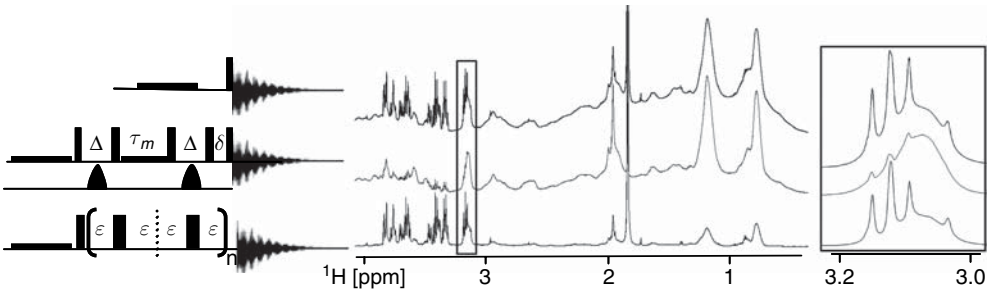

Figure 3.18. Diffusion and T_2 weighted filtration of a serum ^{1}H NMR spectrum: The top row shows a 1D presaturation spectrum together with a boxed expansion. The middle row shows a spectrum and expansion with diffusion-weighted editing. A delay τ_m of 300 ms serves to reduce signals from low-molecular-weight compounds (see triplet in expansion). The bottom row shows a CPMG-based T_2 weighted result. Here, n was set to 150 ($150^*4\varepsilon = 200$ ms) which reduces the broad lipoprotein hump in the background of the triplet substantially (see expansion).

which experience a high mobility [31]. To overcome fast signal decay due to T_2 relaxation, the magnetization, dephased by the first gradient, is aligned along the z-axis and flipped back to the transverse plane prior to the 2nd gradient after 50–300 ms. This technique is referred to as PGSE (**P**ulsed **G**radient **S**pin **E**cho) [74], whereby bipolar gradients [75] up to 100 G/cm are used. An extension to 2D by incrementation of the gradient strength is known by the acronym DOSY [76] (**D**iffusion **O**rdered **S**pectroscop**y**). This technique allows "the sorting" of the compounds in a mixture according to their translation diffusion; applications for metabonomics research including tissues can be found in the literature [32, 77].

3.5.1.3. ^{13}C spectroscopy

The chemical shift dispersion of ^{13}C spins is higher by approximately a factor of 20 when compared to protons. In addition, all signals in a ^{13}C spectrum appear as singlets if ^{1}H decoupling is applied[15]. Homonuclear ^{13}C–^{13}C scalar coupling is absent due to the high dilution of ^{13}C nuclei (only 1.11% of all carbon nuclei carry a spin). Thus the reduction in spectral complexity combined with substantially reduced spectral overlap would make the ^{13}C nucleus an ideal candidate for metabolic profiling. However, this is prohibited by the low intrinsic sensitivity of ^{13}C detection given by $S_{13C} = S_{1H} \cdot \gamma_{13C}^3 \cdot 0.011 \approx S_{1H} \cdot 3.4 \cdot 10^{-4}$. Thus the acquisition of a ^{13}C 1D NMR spectrum that is identical in signal-to-noise to a ^{1}H spectrum of the same sample will need $\sim 10^8$ times as long. Compromises in spectral amplitude are unpreventable and highest sensitivity equipment (^{13}C cryogenic probehead together with 900 MHz) is absolutely compulsory. An additional gain in signal-to-noise compared to ^{1}H is of course seen as no spin multiplicities reduce the amplitude. The situation changes slightly if sensitivity enhancement by polarization transfer from protons (INEPT [78]) is used (see also 3.5.2.2).

In the literature there is one example of a metabonomics analysis by use of ^{13}C spectra [79]. However, general use of ^{13}C detection can only be recommended if further substantial improvements in sensitivity of the detection of NMR signals will be achieved. What is of interest is the enrichment of ^{13}C in biofluids or tissues by use of ^{13}C-labeled tracer molecules in biological studies. With this labeling approach a new window into metabolic analysis and diagnostics may be opened also in higher organisms and metabolic pathway analysis including fluxes can be performed with acceptable sensitivity [80].

3.5.2. 2D NMR

It was mentioned before that a widespread use of 2D NMR in metabonomics is limited to small sample arrays by the long acquisition times needed. Nevertheless examples of these techniques can be found in the literature and are applied in our laboratory. The 2D techniques unfold their full power if a *de novo* structure determination of a newly found metabolite in a biological sample is needed. We will not go into the process of structure determination here. The reader is referred to the expert literature [81].

3.5.2.1. Homonuclear 2D NMR

J resolved (J-Res): The most simple homonuclear 2D NMR experiment is the so-called J-resolved experiment [82]. Here J-coupling information is separated into a 2nd dimension from the chemical shifts. The experiment was successfully applied to many body fluids including urine [83], CSF[61], seminal fluid [84] and blood plasma [71]. An attractive feature of the 2D data matrix obtained is the calculation of a so-called chemical shift spectrum. It is obtained as a tilted projection of the full 2D matrix. All signals (except those with strong coupling, e.g. citrate) appear as singlets positioned at the center of their respective multiplets in the standard 1D counterpart (see Figure 3.19). Thus the complexity of the spectrum is substantially reduced. The integral of signals in chemical shift spectra is strongly influenced by T_2 relaxation. A straightforward way to use this data for absolute quantification of concentration is not possible, but comparison with a spectral database of metabolites recorded under identical experimental conditions might be promising for relative quantification. Typical experimental settings require the acquisition of 4–8 k complex data points along t_2 for 32–128 t_1 increments. An interscan delay of 2 s allows for T_1 relaxation. Exponential or shifted sine-bell windowing is required along t_2. Along t_1 a sine or squared sine multiplication together with data summation can be recommended.

Correlation Spectroscopy (COSY) [85]: In this type of spectroscopy a cross-peak is obtained if two spins are connected by a homonuclear J-coupling (over

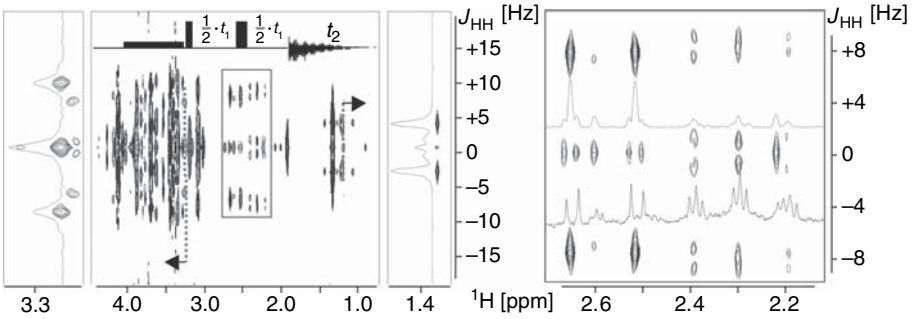

Figure 3.19. J-coupling resolved spectroscopy of rat CSF: Aliphatic region of a tilted 2D J-Res spectrum. On the panels left and right of the overview plot, expanded regions together with 1D traces are shown. The position where these 1D traces where taken are indicated in the overview plot by green dotted lines and black arrows. Along f_1 detailed information on J-coupling patterns can be extracted from the 1D traces. The red boxed region is shown in expansion on the right. Here, for comparison, the corresponding region of the 1D spectrum (black) together with the projection of the 2D dataset (red) is inserted. This demonstrates that only chemical shift information is retained in the 1D projection calculated, leading to a substantial simplification of the spectra. The pulse sequence for 2D *J*-Res is shown.

3–5 bonds). Thus information on spin-system topologies can be extracted. The cross-peaks contain fine-structure allowing for the determination of the values of active and passive J-couplings. The latter is physically not contributing to the origin of the cross-peak. These values are of importance for structure elucidation of unknown metabolites. In metabonomics also COSY45 [85] using a 45° mixing pulse and Double-Quantum COSY [86, 87] are applied. As the magnitude of the cross-peak is related to the maximum length of t_1 and t_2, the acquisition of at least 512 times 4 k data points along the respective dimensions can be recommended. Sine or squared sine window functions along both time-domain dimensions are standard. Body fluids for which this experiment has been applied together with suited references are listed in the *J*-Res section above.

Total Correlation Spectroscopy (TOCSY) [88]: In contrast to COSY described above, cross signals appear here also between pairs of spins which have a common coupling to a third or fourth (and so on) spin. This is achieved by multistep transfer of magnetization over many spins. Chemical shift information on the complete spin-system may thus be obtained for any of the members of the spin-system (see Figure 3.20). The much higher information content of the 2D TOCSY map can cause serious signal overlap on cross-peaks.[29] The size of the cross-peak is determined

[29] It must be noted that overlap is reduced compared to COSY due to more favourable line-shapes of TOCSY cross-peaks.

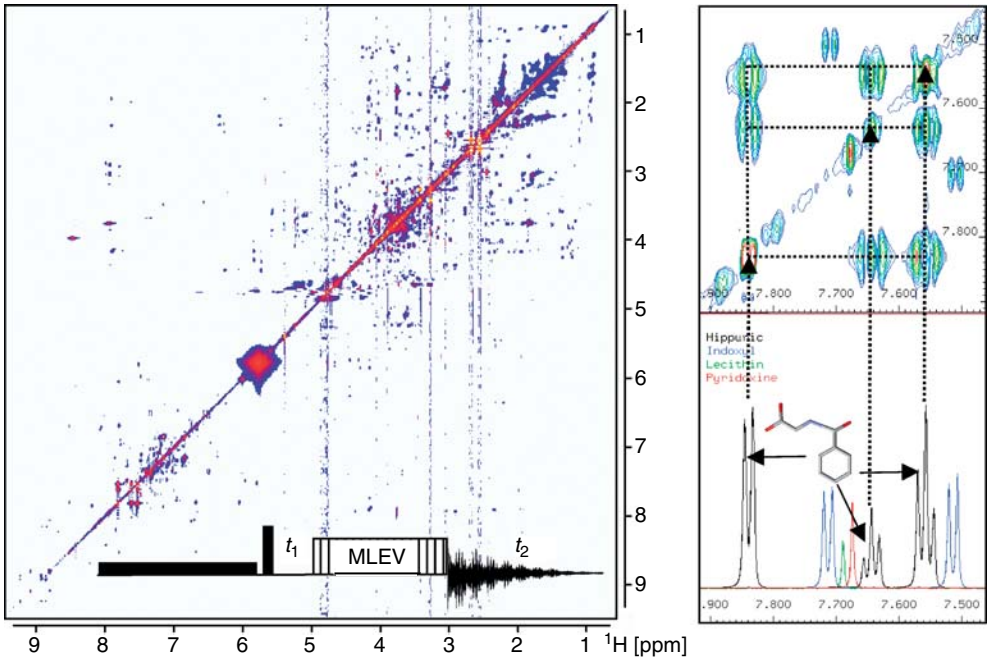

Figure 3.20. Total Correlation spectroscopy (TOCSY): The left panel shows an overview TOCSY spectrum taken from urine of one of the authors. Spin systems of metabolites can be identified as rows of cross-signals correlating the chemical shifts of all spins involved in a spin system topology. The insert (left) shows a schematic drawing of the pulse sequence. The MLEV mixing sequence serves to transfer magnetisation along J-coupled networks. The right panel shows an expanded region correlating protons of aromatic moieties. As an example the spin system of hippurate is identified (right). It is immediately clear which signals belong to the aromatic protons of this molecule. Another spin system also easily assigned in the same spectral region is that of indoxyl. Note: lecithin and pyridoxine, which only show singlet signals in the 1D, do not cause off-diagonal cross-signals in TOCSY. Accurate chemical shift measurements to distinguish both metabolites are mandatory. For comparison the 1D spectra of all four compounds are shown at the bottom (right).

by the length of a so-called mixing sequence. In metabonomics [89] the Malcolm Levitt's CPD sequence (MLEV) mixing sequence with duration of 80–300 ms is applied. Smaller t_1/t_2 data-matrices (e.g. 256*2 k data points) are possible here making the application of squared cosine shaped window functions attractive. Also applications to tissues employing HR-MAS are found in the literature [90, 73]. A combination of TOCSY with diffusion editing (300 ms diffusion delay) has also been reported [32].

3.5.2.2. Heteronuclear correlation experiments

In contrast to the direct 1D-based observation of ^{13}C information, the indirect measurement via J-coupled ^{1}H spins offers several advantages: (1) the sensitivity of

these class of experiments depends on the higher gyromagnetic ratio of the protons. As a consequence, ^{13}C spins are observed with "only" 100 times reduced sensitivity compared to ^{1}H spectroscopy; (2) the chemical shift of the carbons is correlated with the chemical shift of J-coupled protons. Several experiments found in metabonomics make use of these advantages.

Heteronuclear Single Quantum Coherence (HSQC) [71, 91, 92]: The chemical shift of ^{13}C is correlated via the large one-bond coupling constant $^{1}J_{CH}$ to the ^{1}H chemical shift of the directly bound proton (see Figure 3.21). Typical experimental settings are a relaxation delay of 1s and a refocusing delay of 4.2 ms. In total, 1024 data points with 200–300 scans for each of up to 512 t_1 increments are acquired. The spectral width is set to 140–180 ppm along the ^{13}C dimension [93]. The standard for processing is a squared cosine window function applied along both dimensions. It must be noted, however, that the acquisition time for such an experiment will be in the range of half a day making the application attractive only for selected samples.

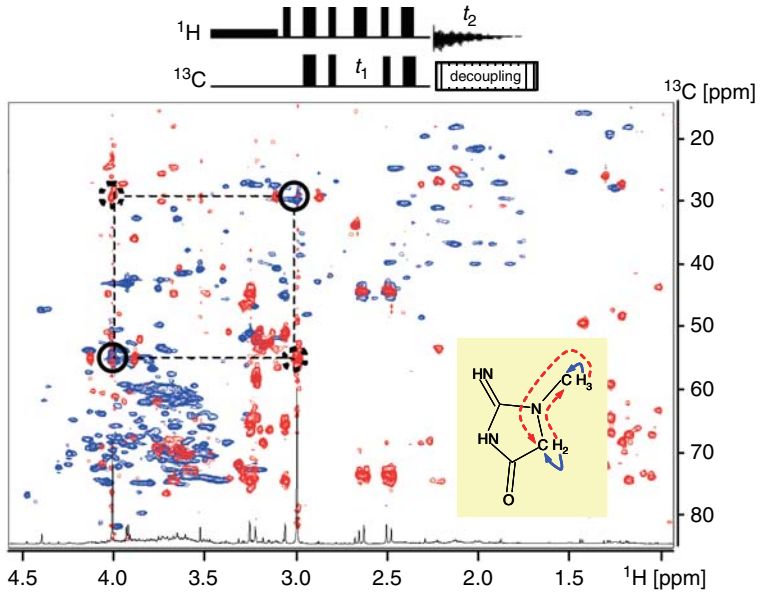

Figure 3.21. The schematic drawing of the HSQC sequence (HMBC not shown) is shown at the top. The 2D spectrum shows a superposition of a HSQC spectrum (blue) and a HMBC spectrum (red) of urine of one of the authors; only the aliphatic region is shown. Highlighted are the correlation signals for the CH_2 and CH_3 groups of creatinine. The full circles show the direct $^{1}J_{HC}$ cross-peaks of the CH_3 and CH_2 group, respectively, in the HSQC spectrum (blue); the dotted circles highlight the long range $^{3}J_{HC}$ cross-peak observed in the HMBC spectrum of the two groups. The direct and long range correlations are also indicated in the structure (insert) with full blue and dotted red arrows, respectively. Many other very informative correlations are seen in the two 2D spectra. The acquisition of these two data sets took one weekend of experimental measuring time.

Heteronuclear Multiple Quantum Coherence (HMQC) [94]: The information content obtained from this experiment is identical to that from HSQC. Applications in metabonomics [79] reported 800 scans and 128 t_1 increments. A short 500 ms interscan relaxation is described in literature [95]. The same holds true for tissue-based work employing HR-MAS [90, 73]. A single example of the application of ^{1}H-^{31}P HMQC-TOCSY [96] can also be found. Here the ^{1}H spin-system information, obtained via TOCSY transfer, is correlated with the chemical shift of phosphorous. This technique is not limited in sensitivity due to the high natural abundance of the NMR active phosphorous isotope.

Heteronuclear Multiple Bond Correlation (HMBC) [97]: With this experiment ^{13}C nuclei having a small multiple bond coupling $^{n}J_{CH}$ are correlated via their chemical shifts to ^{1}H (see Figure 3.21). For each proton several cross signals related to coupled ^{13}C nuclei appear in the spectrum. The application in metabonomics [98] is hampered by the long experimental time. About 128–512 FIDs encompassing 220 ppm ^{13}C spectral width and up to 400 scans per increment are required. Along t_2, between 2 k and 32 k data points can be recorded. A squared sine-shaped window function is recommended for processing. Typically a delay of 70 ms was used for transfer of magnetization. The HR-MAS applications are described in literature [89, 90].

3.6. Data pre-processing

Data pre-processing can be seen as an intermediate step between recording of raw spectra and applying data analysis and modeling methods. Data pre-processing transforms the data in a way that subsequent analyses and modeling are easier, more robust and more accurate. For NMR spectra from metabonomic studies, pre-processing methods are usually used to reduce variances and influences, which are not of interest or which might interfere with data analysis. These variances can be caused by measurements, by biology or by combined effects. Among these effects, NMR typical effects such as incorrect phases of the spectrum or varying baselines, animal study specific effects such as varying overall concentrations of biofluids and combined effects such as shifting peak locations in the spectrum play a role. Data pre-processing methods, which account for effects typical for NMR measurement and which are routinely applied to NMR spectra in many fields, are discussed in Section 3.2.1.2. This section covers data pre-processing methods, which are more specific for metabonomics. Data pre-processing for metabonomic studies usually accounts for three effects discussed in the following sections, whereby not all pre-processing steps are mandatory.

3.6.1. Exclusion regions

Usually, the first step of data pre-processing is an exclusion of spectral regions, which contain non-reproducible information or which do not contain information about metabolites. Typically, the spectral width of recording NMR spectra is wider than chemical shifts populated by endogenous metabolites. On the one hand, a minimum spectral width is predefined by a symmetrical range around the water resonance when using water suppression techniques. On the other hand, a broad spectral width simplifies NMR pre-processing methods such as phasing and baseline corrections. For the data analysis, spectral regions not populated by endogenous metabolites are not useful. As these regions are sensitive to spectral artifacts such as inadequate phasing, exclusion is beneficial. Therefore, the spectrum outside the window of 0.2–10 ppm is usually excluded. Other parts of the spectrum, which are usually excluded, are resonances of the solvent water. Although special suppression techniques for water resonances are used, the remaining signal of water, which varies due to different pre-saturation conditions, dominates the spectrum between 4.6 ppm and 5.0 ppm rendering the analyses of signals of metabolites lying below the water resonances impossible. For spectra of urine, the signal of urea, which is very close to water, is the most dominating peak besides that of water, as urea is the most concentrated metabolite in urine. Yet, the urea peak is not quantitative, as the protons exchange with water and consequently the peak intensity varies with the quality of water suppression (see Figure 3.17). As this peak has a high intensity and a broad range between 5.4 and 6.0 ppm, and as this peak varies, it highly influences methods of multivariate data analysis unless the peak is excluded.

In many publications on the analysis of urine samples, the exclusion regions of water and of urea are combined and the complete spectral region between 4.50 and 5.98 ppm is excluded. Although this combination of exclusion areas also excludes spectral regions with signals of metabolites of interest (for example the resonance of the α-anomeric proton of glucose at 5.24 ppm), it is the *de facto* standard in the literature. Consequently, only the spectral range between 0.2 and 4.5 ppm and between 5.98 and 10 ppm is analyzed in the case of urine samples. For plasma samples and extracts of tissue the spectral ranges from 0.2 to 4.6 ppm and from 5.0 to 10 ppm are used, whereby additional exclusion regions might be necessary depending on resonances of solvents for tissue extracts.

3.6.2. Peak shifts and binning

One of the major challenges for automated data analyses and peak assignments for spectra of biofluids from metabonomic studies is the effect of shifting peaks (see Section 3.4.2). As the matrix of biofluids and, in particular, urine is highly varying, the local environment of protons also varies. Thereby many parameters play a role

such as changes of the pH, changes of the salt concentration, overall dilution of the sample, relative concentrations of specific ions, relative concentration of specific metabolites and many more. All these parameters can influence the shifts of peaks, whereby not all peaks are affected and different peaks are affected to a different extent even when belonging to the same metabolite. In the left panel of Figure 3.22 an example with ^{1}H NMR spectra of 30 samples from a human metabonomic study is shown. The clipping between 3.35 ppm and 3.49 ppm shows one singlet belonging to para-hydroxyphenylacetate and one triplet belonging to taurine. The position of the singlet is very stable at 3.455 ppm whereas the position of the taurine triplet varies between 3.407 and 3.435 ppm. As the pH of the samples was constant due to buffer added, the shifts of the taurine triplet are caused by different concentrations of salts, of metal ions, of metabolites, or by varying dilutions of the samples.

In a recent publication [99], it has been shown that peak shifts can be beneficial for classifying groups of samples, if the locations of peaks are systematically different between the various groups of samples. Yet, it is not possible to trace these differences back to a single parameter such as systematic differences of pH, as many parameters interact with peak shifts. In addition, the interpretation of classifications and separations based on shifts of peaks is difficult, as the peak shifts are usually overlaid by changes of the intensity of peaks due to changes of concentrations of metabolites. Consequently, the loadings of decomposition methods (PCA, partial least squares (PLS), O-PLS) show distorted shapes of peaks, which need thorough manual interpretations. Therefore, a reduction of peak shifts either in silico or by special sample preparation methods (see Section 3.4.2.3) is most beneficial for the majority of applications. The mathematical approaches of reducing the influence of peak shifts can be classified into two categories. The first widespread category reduces the resolution of spectra (by the so-called binning or bucketing procedures), whereas methods belonging to the second category try to align peaks.

Equidistant binning: The most popular methods of reducing the influence of shifting peaks are the so-called binning or bucketing methods, which reduce the resolution of spectra.[30] Thereby the spectra are integrated within small spectral regions, which are called “bins” or “buckets”. Subsequent data analysis procedures, which are applied to the binned spectra, are not influenced by peak shifts, as long as these shifts remain within the borders of the corresponding bins. The vast majority of ^{1}H-NMR metabonomics literature uses an equidistant binning of 0.04 ppm. This means that the spectrum is split into evenly spaced integral regions with a spectral width of 0.04 ppm. In addition, it is common practice to sum up bins covering the citrate doublets (two bins per doublet) into super-bins, as the shifts of the two citrate

[30] This reduction of resolution is not identical to that obtained by a lower B_0 field strength. A multiplet with a total width of 24 Hz covers 0.04 ppm at 600 MHz. The ppm-width doubles at 300 MHz.

doublets span more than one bin width each. With respect to exclusion areas the spectra are typically reduced to 206 bins visualized in Figure 3.23.

The major drawback of equidistant binning is the non-flexibility of the boundaries. If a peak crosses the border between two bins, this peak shift can significantly influence the data analysis. For example, when analyzing binned spectra by PCA or PLS, peak shifts result in loadings with opposite signs. If the spectra shown in Figure 3.22 are binned by the wide-spread equidistant binning method with a bin width of 0.04 ppm, the signal of the triplet of taurine is distributed among two bins (see Figure 3.22, left). The intensity assigned to each of the two bins highly depends on the exact location of the triplet. Thus, significant variance is added to the two bins, which is not related to the concentrations of taurine and para-hydroxyphenylacetate. Therefore the analysis of the concentrations of these metabolites on the basis of these bins is hampered.

Non-equidistant binning: To prevent peaks being cut by the boundaries of bins, binning methods have been proposed, which are based on non-equidistant spacing of bins. In the ideal case, the borders of the bins are adjusted in a way that bins cover only complete peaks including all possible locations of the peaks. This means that the bin width depends on the width of the peak shape and on the shift width of the peak. An example is shown on the right side of Figure 3.22. None of the peaks is cut by the

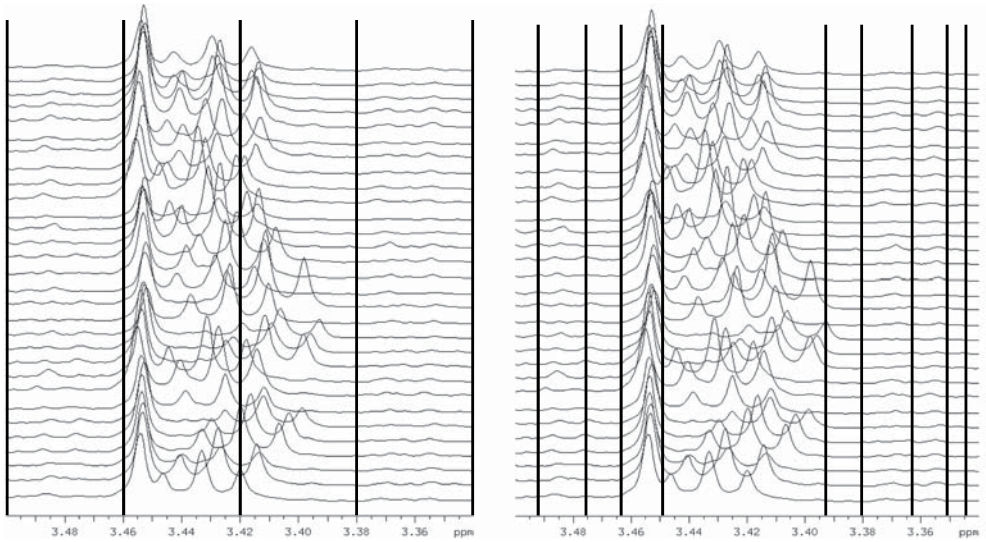

Figure 3.22. Two different binning methods applied to a set of spectra. On the left side the boundaries of the widely used equidistant binning with a bin width of 0.04 ppm are shown. The right side shows the boundaries of a non-equidistant binning method, which locates the boundaries of the bins into minima of mean spectra.

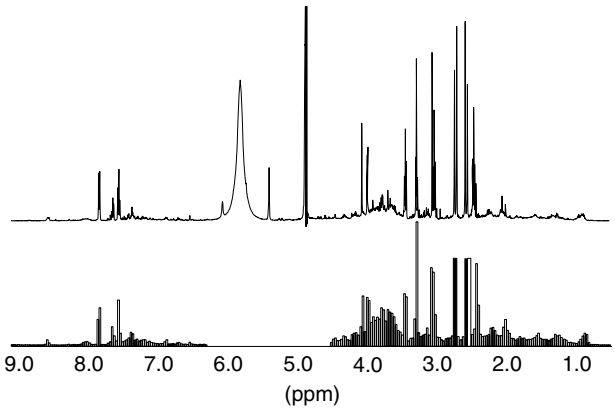

Figure 3.23. Spectrum before and after equidistant binning with a bin width of 0.04 ppm. In addition, two super bins for the citrate doublets were created and the spectral area between 4.5 and 5.98 ppm was excluded. A vast majority of ^{1}H NMR metabonomics literature uses this form of reduced spectra for multivariate data analyses.

bin boundaries. The width of the bin covering the peak of para-hydroxyphenylacetate is narrow, as the peak does not shift. On the other side, the triplet of taurine is quite wide and shows extensive shifts. Therefore, the corresponding bin is very wide.

The binning shown on the right side of Figure 3.22 is based on positioning the borders of the bins on local minima of the mean spectrum of the set of control spectra of the study. The considerable number of spectra results in a very smooth spectrum. Therefore, the local minima separate well between shifting peaks while conserving each peak in a bin. For static peaks, the hyperfine structure is split into separate bins. Overall, 604 bins were generated for the spectral range of 0–10 ppm with the water peak excluded.

The unequal bin width also has some fundamental consequences for the subsequent data analysis. For the example discussed above the bin width varied between 0.0014 and 0.46 ppm (broad peak of urea). If no further processing of the spectrum is performed, data points belonging to different bins influence the subsequent data analysis to a different extent. For example, a data point of a very small bin has a higher influence than a data point of a very wide bin. If the subsequent data analysis is to represent the data points of the raw spectra as adequate as possible, it might be appropriate to scale the integral of the bins by the inverse of the bin width. On the other side, each bin represents one single signal in the ideal case. Therefore the integral of a bin represents the "proton concentration" and with it the concentration of the metabolite behind the signal. Thus, no further scaling is needed, if the data analysis is to be based on the concentrations of metabolites and changes of concentrations of metabolites.

A combination of equidistant binning and non-equidistant binning has been proposed and applied in metabonomics studies recently [100–102]. This method starts with the traditional bucket size of 0.04 ppm and allows the adjustment of the borders by 50% resulting in a bin width between 0.02 and 0.06 ppm. The adjustment is based on finding local minima of a sum of spectra or of a skyline projection of spectra.

Peak alignment: The alignment of peaks is an alternative to reducing the spectrum to account for peak shifts. A very promising approach has been proposed recently [37]. Thereby the spectra are segmented and the segments are shifted, stretched and shrunken to maximize the correlation coefficient between the segments of a spectrum and a target spectrum. Overlapping connections after alignment are deleted and missing connections are interpolated. The optimal parameters are determined either by a genetic algorithm [37] or by a grid-search [103]. Although first results are very promising, further validation has to be performed as eliminations of peaks, changes of peak shapes and erroneous alignments cannot be excluded. Another approach for the alignment of NMR spectra of complex mixtures with the same strength and drawback has been proposed by Stoyanova [38]. In this approach peak shifts are identified by comparing a set of typical derivative shapes (e.g. derivatives of Lorentzian doublets with various coupling constants) with the loadings of the second principal component of the set of spectra. Identified regions with peak shifts are aligned to a certain average frequency, whereby gaps are filled with zeros.

3.6.3. Normalization

A crucial step in data pre-processing of spectra from metabonomic studies is the so-called normalization. The normalization step tries to account for variations of the overall concentrations of samples. Especially for urine, these variations of the overall concentrations of samples are very distinctive. Urine of animals and humans can typically be diluted by a factor of 4–5, whereby the dilution due to food deprivation or due to drug effects can exceed a factor of 10. In addition, sample preparation steps, such as addition of buffer, can further contribute to dilution of urine. In contrast to changes of the overall concentration of urine, metabonomic responses and fluxes mainly influence only a few analytes in urine. Thereby concentrations of these few metabolites change specifically. These changes are visible as relative changes of concentrations of the few metabolites related to the concentrations of all other metabolites, which represent the overall concentration of urine. Usually these relative specific changes are of interest in metabonomic studies. Therefore, a normalization step, which compensates for differences of the overall concentration, is crucial, as variations of overall concentrations obscure specific changes of metabolites. Besides compensating for changes of the overall concentration of samples, normalization can also be necessary due to technical reasons. If spectra are recorded using different

number of scans or if spectra are recorded with different devices, the absolute values of the spectra are different rendering a joint analysis of the spectra without prior normalization impossible.

All normalization procedures scale complete spectra in a way that these spectra represent the same overall concentration. The scaling factor of each spectrum corresponds to the dilution factor of the corresponding sample. In general, most normalization methods (also in other scientific fields) are special formulations of the general equation:

$$I(i) = \frac{I^{\text{old}}(i)}{\sum_k \left[\int_{j_k^l}^{j_k^u} (I(x))^n dx \right]^{\frac{1}{n}}} \tag{3.6}$$

Hereby $I^{\text{old}}(i)$ and $I(i)$ are the intensities of the variable i (spectral feature, wavelength, bin, chemical shift) before and after normalization, k is an index of the spectral regions used for normalization, j_k^l and j_k^u are the lower and upper borders of the spectral region k, for which the power n of the intensities $I(x)$ are integrated.

For the integral normalization, the power n is set to 1. Special variants of integral normalization, such as creatinine normalization in metabonomics, use only parts of the spectral range. Thereby j_k^l and j_k^u mark the borders of the k parts (e.g. left and right borders of the 2 creatinine peaks). The vector length normalization, which has not been used in the field of metabonomics up to now, uses a power n of 2. In the next sections, the normalization procedures used in the field of metabonomics are discussed in more detail.

Integral normalization: This method is the *de facto* standard of normalizing NMR spectra of biofluids. Integral normalization assumes that the integrals of spectra are mainly a function of the overall concentrations of samples. A linear concentration series of urine should result in a linear series of integrals of the corresponding spectra. Influences of changes of individual concentrations of single analytes are assumed to be small compared to changes due to varying overall concentrations of urine. In addition, specific down-regulations of metabolites should balance up-regulations of metabolites to a certain extent. The integral normalization procedure divides each signal or bin of a spectrum by the integral of the spectrum or part of it. Virtually all publications dealing with ^{1}H-NMR metabonomics of urine samples exclude the spectral range of urea and water resonances from the integration and multiply each variable additionally by a factor of 100 ending up in a total integral of 100 for each spectrum [37, 104–108].

In Figure 3.24, two NMR spectra of rats from the same study are shown before normalization and after normalization. The spectrum shown in black is diluted

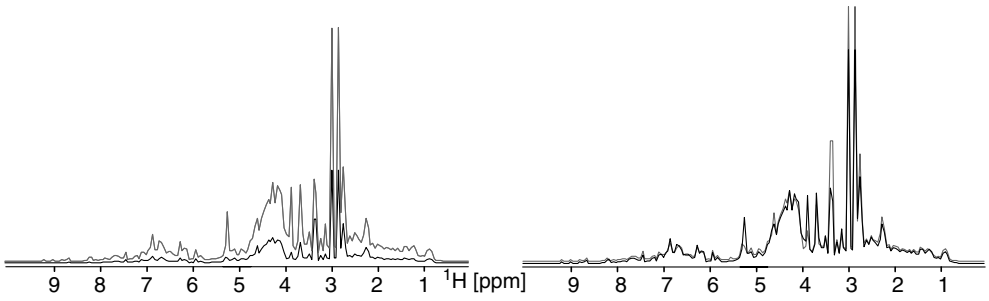

Figure 3.24. Two NMR urine spectra before normalization (left side) and after normalization (right side). The black spectrum is diluted by a factor of 3 compared to the grey spectrum. After normalization it is obvious that only very few differences of relative concentrations of metabolites between both spectra exist.

by a factor of 3 compared to the grey spectrum. Therefore the spectra look very different before normalization. The right side of the figure shows the two spectra after normalization. Only very few differences are visible. These differences are based on differences of relative concentrations of metabolites (specific differences of the metabolic profiles) and are typically of interest. Before normalization these differences were covered by the dilution effect.

Although integral normalization is the *de facto* standard for NMR measurements for metabolic profiling, the robustness, which is crucial for analysing highly varying biofluids, is the main weakness of the method. As soon as the integral of a spectrum is dominated not only by the overall concentration but also by specific changes of metabolites, the integral normalization does not scale the corresponding spectrum correctly. In Figure 3.25, an example of a sample is shown, which contains extreme

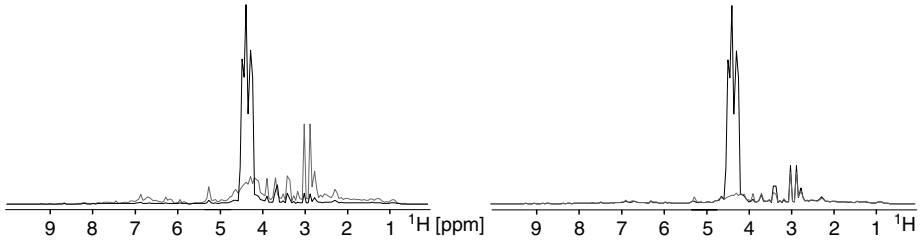

Figure 3.25. Two NMR urine spectra after integral normalization (left side) an optimal normalization (right side, quotient normalization). The black spectrum contains massive amounts of glucose due to a glucosuria of the animal. As the peaks of glucose account for 75% of the integral of the spectrum, the integral normalization method has incorrectly downscaled the spectrum by a factor of 0.25. An optimal normalization shown on the right side reveals that besides the extreme amounts of glucose hardly any differences between the two samples exist.

amounts of glucose due to glucosuria of the corresponding animal. As the peaks of glucose account for 75% of the integral of the spectrum, the integral normalization has downscaled the complete spectrum by a factor of 0.25 compared to an optimal normalization. Besides the presence of extreme amounts of endogenous metabolites, drug metabolites, for which the signals were not excluded or replaced, can also dramatically mislead the integral normalization.

Creatinine Normalisation: This method is a special version of the integral normalization originating from clinical chemistry. For the investigation of urine of humans and animals in clinical chemistry, it is a common procedure to normalize concentrations of analytes by the concentration of creatinine [111, 112–114]. The assumption is a constant excretion of creatinine into urine (often referred to as creatinine clearance). Thus, creatinine is an indicator of the concentration of urine. As the concentration of creatinine can be determined by its peaks in the spectra (peaks at 3.05 and 4.05 ppm), it has been proposed in early publications to normalize spectra by the concentration of creatinine [113].

Yet, the practical application of the creatinine normalization is faced by technical and biological difficulties. From the technical point of view, metabolites with overlapping peaks can interfere with the determination of the creatinine concentration (e.g. creatine at 3.04 ppm). In addition, the chemical shift of creatinine at 4.05 ppm depends on the pH of samples rendering a sophisticated peak picking algorithm necessary. Biological challenges for creatinine normalization are changes of the concentrations of creatinine due to metabonomic responses, which has been shown in several studies [114]. In that case, the normalization by creatinine is worthless. As at the step of normalization of spectra a possible increase of the creatinine level due to metabonomic responses is usually not yet known, a creatinine-based normalization is not of general use in metabonomics.

Quotient normalization: Recently, a new approach of normalization has been proposed [115], which is based on the assumption that changes in concentrations of single analytes only influence parts of the spectra, whereas changes of overall concentrations of samples influence the complete spectrum. Instead of the assumption of a constant total integral of the integral normalization, a most probable quotient between a spectrum and a reference spectrum is calculated as normalization factor. This most probable quotient for a specific spectrum can be derived from the distribution of each bin (variable, spectral value, data point) of a spectrum divided by the corresponding bin of a reference spectrum. In Figure 3.26, two histograms of the distributions for the quotients of two spectra each are shown. The histogram on the left side represents the quotients of the 205 bins of the black and grey spectrum shown in the left part of Figure 3.24. It is clear that the distribution of the quotients is rather sharp. The most probable quotient is located around 0.33, which corresponds to a threefold dilution of the black spectrum. The histogram on the right

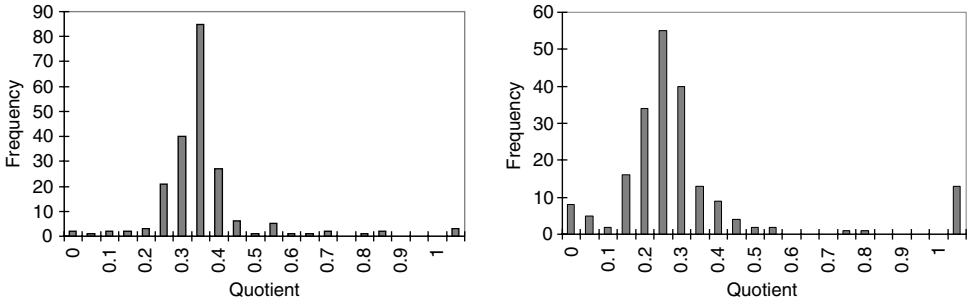

Figure 3.26. Histograms for the distributions of the quotient of two spectra each. The left histogram represents the quotients of the two spectra each shown in the left part of Figure 3.24. The dilution factor of 3 of the black spectrum is represented as most probable quotient around 0.33. The right histogram represents the quotients of the two spectra shown in the left part of Figure 3.25. The most probable quotient is around 0.25. The dominance of the glucose peaks is only visible as a rather small bar for quotients higher than 1 (last bar on the right side).

side of Figure 3.26 represents the quotients of the two spectra shown in the left part of Figure 3.25. The distribution shows that the most probable quotient is around 0.25. This corresponds to the incorrect downscaling of the black spectrum by the integral normalization, as the glucose peak contributes 75% to the total integral. The dominance of the glucose peak hardly influences the distribution and the most probable quotient. It is only visible as rather small bar on the right side, which represents quotients higher than 1.

The quotient normalization is based on the robustness of the distribution of quotients in contrast to integrals. The most probable quotient, which is needed for the normalization, can be estimated in several ways. Using the median as estimation of the most probable quotient has been proven to be a very robust and very exact method. For the calculation of the quotients, a reference spectrum is needed in contrast to integral normalization. The reference spectrum can be a single spectrum of the study, a "golden" reference spectrum from a database, or a calculated median or mean spectrum on the basis of all spectra of the study or on the basis of a subset of the study. It has been shown that the choice of reference spectrum is uncritical, and a median spectrum of control animals seems to be the most robust reference spectrum for studies with only few animals. It is also recommended to perform an integral normalization prior to quotient normalization to scale different studies to the same absolute magnitude. The algorithm of the quotient normalization can be summarized as follows:

1. Perform an integral normalization (typically a constant integral of 100 is used)
2. Choose/calculate a reference spectrum (best approach: calculate median spectrum of control samples).

3. Calculate the quotients of all variables of interest of the test spectrum with those of the reference spectrum.
4. Calculate the median of these quotients.
5. Divide all variables of the test spectrum by this median.

The quotient normalization can be applied to raw spectra or to binned spectra, but variables not containing signals should be excluded from calculating the most probable quotient if possible.

The robustness of the quotient normalization can be seen in Figure 3.25. The integral normalization is heavily hampered by massive amounts of glucose in the sample represented by the black spectrum. Therefore the spectrum shown in black is incorrectly downscaled (left part of the figure). On the right side of Figure 3.25, the results of the quotient normalization are shown. It is clear that the quotient normalization is not influenced by the glucose peak, but performs an ideal normalization: The quotient normalization reveals the high similarity of both spectra representing nearly identical relative concentration levels of most metabolites except glucose. This similarity remains hidden when using the integral normalization.

The quotient normalization performs better than the integral normalization not only in the case of very intense signals due to strong metabolic responses or due to drug metabolites, but also in the case of "normal" spectra of control animals showing only small normal metabolic variations. For example, 4023 spectra from control animals and pre-dosed animals were normalized by the total integral normalization procedure and by the quotient normalization procedure. As all animals had not been dosed, the creatinine levels of the samples were not influenced by metabonomic changes and consequently represent a good measure for the overall concentration of samples. The variations of the creatinine levels, which can be seen as performance indicator for the normalization, are 7.6% for the integral normalization and 6.7% for the quotient normalization. This means that even when looking at control animals, small specific variations due to metabonomic fluxes negatively impact the integral normalization.

A drawback of the quotient normalization compared to integral normalization is the absence of a fixed independent absolute reference value (e.g. a total integral of 100). This renders a re-normalization of a combined data set necessary, if two data sets which were normalized separately are merged or compared. This re-normalization can be prevented if a common reference spectrum is used for the separate normalizations of various data sets.

3.7. Outlook

The intrinsic properties of complex biological samples will continue to be a big challenge for NMR in metabolic analysis:

- Metabonomics fundamentally rests on large sample arrays. Perfect control of all spectral parameters (including chemical shifts) in mixtures cannot be achieved. As a consequence, assigning corresponding variables across different samples of one and the same study to the same molecular structure is of paramount importance. NMR will have to deal with "imperfect" samples although improvements in sample preparation are certainly possible.
- Currently there is a severe "assignment gap" – many NMR peaks in body fluids are not assigned to metabolites. Recourse to pattern-only analysis without knowing their most prominent molecular features is unsatisfactory for mechanistical biochemical interpretation.
- In principle, using NMR spectroscopy submicrogram quantities can be quantified. However, this is possible in practice only if signals of high amplitude do not overlap with small signals. This prerequisite is not fulfilled for biological samples, where concentration ranges of metabolites can span several orders of magnitude. Large changes of compounds that are low in concentration can only be detected if signal overlap is avoided. Here liquid chromatography (LC)-based separation techniques might come into play combined with NMR. It might well be that the fundamental "dogma" of metabonomics "no sample preparation needed" will soon have to be revisited.

Thus from the NMR spectroscopic point of view, complete profiling of biological mixtures will remain difficult for very fundamental technology reasons.

The ongoing development of NMR is addressing some of these problems:

- The trend to higher field-strength will improve the analysis of crowded spectra. More metabolites will become visible without overlap of lines. The 1 GHz magnet is within reach. Nevertheless the dynamic range problem in signal intensity combined with signal overlap will always cause difficulties in the detailed analysis of metabolites that are low in concentration.
- The introduction of cryogenic probeheads has improved the sensitivity of detection within the last 10 years by a factor of 4–5. Now it is a fact that for aqueous biological samples the "sample noise" is one limiting factor of NMR sensitivity. It seems unlikely that NMR devices will be available in the mid-term future which will allow the detection of compounds which are orders of magnitude lower in concentration than those analyzed today. We estimate that less than 100 metabolites in urine provide more than 95% of the total signal amplitude, and this will be true at any given sensitivity.
- The low concentration and the high range to be detected prevent the application of multi-dimensional NMR techniques, which could resolve the overlap problem for large sample arrays. Due to the extensive measuring times needed for diluted samples, these methods can be employed only for very focused questions like the *de novo* determination or verification of the structure of an unknown endogenous

or xenobiotic metabolite. There is no obvious gap in the availability of NMR pulse sequences. Nevertheless improvement in details (see Section 3.5.1.1) is needed to adjust the pulse sequence to the given sample and problem.

Having said that, we foresee the two following main developments in NMR methods that will open new avenues in metabonomic analysis:

- Currently different vendors of NMR spectrometer and industrial research groups are building up spectral databases [116] of metabolites known to be present in biological samples at concentrations visible in NMR. The aim of these efforts is to develop an analysis tool for *a priori* assignment of spectra of biofluids to individual metabolites. The analysis rests today almost exclusively on multivariate analyses of spectral features and patterns, in order to identify and extract significant peaks which are *a posteriori* assigned. The *a priori* assignment would open up a window to concentration values of hundreds of known metabolites [117]. Different software companies [116, 118, 119] are working on algorithms to allow extraction of this information by interactive analysis of 1D NMR spectra.
- A second development expected to have impact in the field is the concerted analysis of data acquired with different analytical methods on the same sample (NMR, GC-MS, LC-MS) [120]. The richness of information to be obtained by this approach cannot be overestimated.

We expect that both developments will have a substantial impact on the proliferation of metabolite profiling methods in wide fields of discovery and applied medical research.

References

[1] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley and Sons, New York, 1986.
[2] D.G. Gadian, *NMR and its application to living systems* Oxford University Press, Oxford, 1995.
[3] F.F. Brown, I.D. Camplell, P.W. Kuchel, D.C. Rabenstein, *FEBS Lett*. 82 (1977) 12.
[4] J.R. Bales, D.P. Higham, I. Howe, J.K. Nicholson, P.J. Sadler, *Clin. Chem.* 30 (1984) 426.
[5] J.K. Nicholson, I.D. Wilson, *Prog. NMR Spectrosc.* 21 (1989) 449.
[6] Bell, J.D. *MRS in biology and medicine*, Eds J.D. de Certaines, W.M.J. Bovee, F. Bodo, Pergamon Press, Oxford.
[7] J.K. Nicholson, J. Conelly, J.C. Lindon, E. Holmes *Nature Reviews*, *Drug Disc*. 1 (2002) 153.
[8] J.C. Lindon, J.N. Nicholson, E. Holmes, H. Antii, E. Bollard, H. Keun, O. Beckonert, T. Ebbels, M. Reily, D. Robertson, G. Stevens, P. Luke, A. Breau, G. Cantor, R. Bible, U. Niederhauser, G. Schlotterbeck, H. Senn, U. Sidelmann, A. Laursen, A. Tymiak, B. Car, L. Lehman, J.-M. Colet, C. Thomas, *Tox. Appl. Pharm.* 187 (2003) 137.
[9] A.R. Fernie, R.N. Tethewey, A.J. Krotzky, L.W. Willmitzer, *Nature Reviews, Molec. Cell Biol.* 5 (2002) 1.

[10] R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, Oxford, 1990.
[11] M. Goldman, *Quantum Description of High-Resolution NMR in Liquids*, Oxford University Press, Oxford, 1991.
[12] J.L. Sudmeier, S.E. Anderson, J.S. Frye, *Conc. Magn. Reson.* 2 (1990) 197.
[13] Q.N. Van, G.N. Chmurny, T.D. Veenstra, *Biochem. Biophys. Res. Comm.* 301 (2003) 952.
[14] P.T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Oxford University Press, Oxford, 1991.
[15] P.F. Flynn, D.L. Mattiello, H.D.W. Hill, A.J. Wand, *J. Am. Chem. Soc.* 122 (2000) 4824.
[16] D. Moskau, *Concepts Magn. Reson.* 15 (2002) 164.
[17] G. Schlotterbeck, A. Ross, R. Hochstrasser, H. Senn, T. Kühn, D. Markek, O. Schett, *Anal. Chem.* 74 (2002) 4464.
[18] D.L. Olson, T.L. Peck, A.G. Webb, R.L. Magin, J.V. Sweedler, *Science* 270, (1995) 1967.
[19] T. Horiuchi, M. Takahashi, J. Kikuchi, S. Yokoyama, H. Maeda, *J. Magn. Reson.* 174 (2005) 34.
[20] J.C. Hoch, A.S. Stern , *NMR Data Processing*, John Wiley & Sons Inc. New York, 1997.
[21] D.D. Traficante, M. Rajabzadeh, *Conc. Magn. Reson.* 12 (2000) 83.
[22] V. Silvestre, S. Goupry, M. Trierweiler, R. Robins, and S. Akoka, *Anal. Chem.* 73 (2001) 1862.
[23] R. Freeman, E. Kupce, *NMR in Biomed.* 10 (1997) 272.
[24] I. Solomon, *Phys. Rev.* 99 (1955) 559.
[25] W.P. Aue, E. Bartholdi, R.R. Ernst, *J. Chem. Phys.* 64 (1976) 2229.
[26] A. Ross, G. Schlotterbeck, W. Klaus, H. Senn, *J. Biomol. NMR* 16 (2000) 139.
[27] O. Cloarec, M.-E. Dumas, A. Craig, R.H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J.C. Lindon, E. Holmes, J. Nicholson, *Anal. Chem.* 77 (2005) 1282.
[28] H.M. McConnell, *J. Chem. Phys* 28 (1958) 430.
[29] L.Y. Lian, G.C.K. Roberts, *NMR of Macromolecules – A Practical Approach*, IRL Press, Oxford, 1993.
[30] C.J.L. Silwood, M. Grootveld, E. Lynch, *J. Biol. Inorg. Chem.* 7 (2002) 46.
[31] E.O. Stejskal, J.E.J. Tanner, *Chem. Phys.* 42 (1965) 288.
[32] M. Liu, J.K. Nicholson, J.C. Lindon, *Anal. Chem.* 68 (1996) 3370.
[33] J.D. Otvos, E.J. Jeyarajah, D.W. Bennett, *Clin. Chem.* 37 (1991) 377.
[34] M. Pellechia, D.S. Sems, K. Wüthrich, *Nature Rev. Drug Disc* 1 (2002) 211.
[35] P.J. Hajduk, T. Gerfin, J.M. Boehlen, M. Häberli, D. Marek, S.W. Fesik, *J. Med. Chem.* 42 (1999) 2315.
[36] J.C. Lindon, E. Holmes, J.K. Nicholson, *Prog. in NMR Spectrosc.* 45 (2004) 109.
[37] J. Forshed, I. Schuppe-Koistinen, S.P. Jacobssen, *Anal. Chim. Acta* 487 (2003) 189.
[38] R. Stoyanova, A.W. Nicholls, J.K. Nicholson, J.C. Lindon, T.R. Brown, *J. Magn. Reson.* 170 (2004) 329.
[39] G.C. Lee, D.L. Woodruff, *Anal. Chim. Acta.* 513 (2004) 413.
[40] T.L. Peck, R.L. Magin, P.C. Lauterbur, *J. Magn. Reson.* 108 (1995) 114.
[41] G.E. Martin, R.C. Crouch, A.P. Zens, *Magn. Reson. Chem.* 36 (1998) 551.
[42] W.L. Fitch, G. Detrer, C.P. Holmes, J.N. Shoolery, P.A. Keifer, *J. Org. Chem.* 59 (1994) 7955.
[43] D.L. Olson, J.A. Norcross, M. O'Neil-Johnson, P.F. Molitor, D.J. Detlefsen, A.G. Wilson, T.L. Peck, *Anal. Chem.* 76 (2004) 2966.
[44] H. Kocacs, D. Moskau, M. Spraul, *Prog. NMR Spectrosc.* 46 (2005) 131.
[45] A. Geiger, M. Holz, *J. Phys. E: Sci. Instrum.* 13 (1980) 697.
[46] C.G. Wade, R.D. Johnson, S.B. Philson, J. Strouse, F.J. McEnroe, *Anal. Chem.* 61 (1989) 107A.

[47] M. Spraul, M. Hofmann, M. Ackermann, A.W. Nicholls, S.J.P. Damment, J. N. Halselden, J.P. Shockcor, J.N. Nicholson, J.C. Lindon, *Anal. Commun.* 34 (1997) 339.
[48] P.A. Keifer, *Drug Disc. Today* 2 (1997) 468.
[49] A. Ross, H. Senn, *Drug Disc. Today* 6 (2001) 583.
[50] A. Ross, H. Senn, *Curr. Top. Med. Chem.* 3 (2003) 55.
[51] N.T. Holland, M.T. Smith, B. Eskenazi, M. Bastaki, *Mutat. Res.* 543 (2003) 217.
[52] COMET Consortium internal presentation O. Beckonert, 2002.
[53] S. Deprez, B.C. Sweatman, S.C. Connor, J.N. Haselden, C.J. Waterfield, *J. Pharm. Biomed. Anal.* 30 (2002) 1297.
[54] D.K. Hansen, M.I. Davies, S.M. Lunte, C.E. Lunte, *J. Pharm. Sci.* 88 (1999) 14.
[55] K.E. Price, S.S. Vandaveer, C.E. Lunte, C.K. Larrive, *J. Pharm. Biomed. Anal.* 38 (2005) 904.
[56] P. Khandelwal, C.E. Beyer, Q. Lin, L.E. Schechter, A.C. Bach, *Anal. Chem.* 76 (2004) 4123.
[57] P. Khandelwal, C.E. Beyer, Q. Lin, P. McGonigle, L.E. Schechter, A.C. Bach, *J. Neurosci. Methods* 133 (2004) 181.
[58] O. Fiehn, *Plant Molec. Biol.* 48 (2002) 155.
[59] J.C. Lindon, J.K. Nicholson, E. Holmes, J.R. Everett, *Concepts Mag. Reson.* 289 (2000) 12.
[60] H.C. Keun, T.M.D. Ebbels, H. Antti, M.E. Bollard, O. Beckonert, G. Schlotterbeck, H. Senn, U. Niederhauser, E. Holmes, J.C. Lindon, J.K. Nicholson, *Chem. Res. Toxicol.* 15 (2002) 1380.
[61] B.C. Sweatman, R.D. Farrant, E. Holmes, F.Y. Ghauri, J.K. Nicholson, J.C. Lindon, *J. Pharm. Biomed. Anal.* 11 (1993) 651.
[62] B.C.M. Potts, A.J. Deese, G.J Stevens, M.D. Reily, D.G. Robertson, J. Theiss, *J. Pharm. Biomed. Anal.* 26 (2001) 463.
[63] R.D. Farrant, J.C. Lindon, E. Rahr, B.C. Sweatman, *J. Pharm. Biomed. Anal.* 10 (1992) 141.
[64] Ciba-Geigy, Eds, *Wissenschaftliche Tabellen Geigy*, vol. 1, 8th edn, Ciba-Geigy, Basel, 1980.
[65] N.N Greenwood, A. Earnshaw, *Chemistry of the elements*, Pergamon Press, Oxford, 1986.
[66] W.S Prince, *Ann. Rep. NMR Spectr.* 38 (1999) 289.
[67] D.I. Hoult , *J. Magn. Reson* 21 (1976) 337.
[68] A. Kumar, R.R. Ernst, K. Wüthrich, *Biochem. Biophys. Res. Commun.* 95 (1980) 1.
[69] R.J. Ogg, P.B. Kingsley, J.S. Taylor, *J. Magn. Reson.* 104B (1994) 1.
[70] E.M. Lenz, J. Bright, I.D. Wilson, S.R. Morgan, A.F.P. Nash, *J. Pharm. Biomed. Anal.* 33 (2003) 1103.
[71] J.K. Nicholson, P.J.D. Foxall, M. Spraul, R.D. Farrant, J.C. Lindon, *Anal. Chem.* 67 (1995) 793.
[72] S. Meiboom, and D. Gill, *Rev. Sci. Instrum.* 29 (1958) 688.
[73] M.E. Bollard, S. Garrod, E. Holmes, J.C. Lindon, E. Humpfer, M. Spraul, and J.K. Nicholson, *Magn. Reson. Med.* 42 (2000) 201.
[74] S.J. Gibbs, and C.S. Johnson Jr, *J. Magn. Reson.* 93 (1991) 395.
[75] G. Wider, V. Dötsch, K. Wüthrich, *J. Magn. Reson.* 108A (1994) 255.
[76] K.F. Morris, C.S. Johnson Jr, *J. Am. Chem. Soc.* 114 (1992) 3139.
[77] J.L. Griffin, H.J. Williams, E. Sang, and J.K. Nicholson, *Magn. Reson. Med.* 46 (2001) 249.
[78] G.A. Morris, R. Freeman, *J. Am. Chem. Soc.* 101 (1979) 770.
[79] H.C. Keun, O. Beckonert, J.L. Griffin, C. Richter, D. Moskau, J.C. Lindon, J.K. Nicholson, *Anal. Chem.* 74 (2002) 4588.
[80] J.G. Jones, M.A. Solomon, A.D. Sherry, F.M.H. Jeffrey, C.R. Malloy, *Am. J. Physiol.* 275 (1998) E843.
[81] H. Günther, *NMR Spectroscopy,* Wiley & Sons Inc. New York, 1992.
[82] W.P. Aue, J. Karhan, R.R. Ernst, *J. Chem. Phys.* 64 (1976) 4226.
[83] E. Holmes, P.J.D. Foxall, M. Spraul, R.D. Farrant, J.K. Nicholson, J.C. Lindon, *J. Pharm. Biomed. Anal.* 15 (1997) 1647.

[84] M.J. Lynch, J. Masters, J.P Pryor, J.C. Lindon, M. Spraul, P.J.D. Foxall, J.K. Nicholson, *J. Pharm. Biomed. Anal.* 12 (1994) 19.
[85] A. Bax, and R. Freeman, *J. Magn. Reson.* 44 (1981) 542.
[86] A. Derome, and M. Williamson, *J. Magn. Reson.* 88 (1990) 177.
[87] B. Ancian, I. Bourgeois, J.-F. Dauphin, and A.A. Shaw, *J. Magn. Reson.* 125A (1997) 348.
[88] A. Bax, and D.G. Davis, *J. Magn. Reson.* 65 (1985) 355.
[89] A.W. Nicholls, E. Holmes, J.C. Lindon, J.P. Shockcor, R.D. Farrant, J.N. Haselden, S.J.P. Damment, C.J. Waterfield, J.K. Nicholson, *Chem. Res. Toxicol.* 14 (2001) 975.
[90] S. Garrod, E. Humpfer, M. Spraul, S.C. Connor, S. Polley, J. Connelly, J.C. Lindon, J.K. Nicholson, E. Holmes, *Magn. Reson. Med.* 41 (1999) 1108.
[91] A.G. Palmer III, J. Cavanagh, P.E. Wright, M. Rance, *J. Magn. Reson.* 93 (1991) 151.
[92] L.E. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* 114 (1992) 10663.
[93] S. Ringeissen, S.C. Connor, H.R. Brown, B.C. Sweatman, M.P. Hodson, S.P. Kenny, R.I. Haworth, P. Mcgill, M.A. Price, M.C. Aylott, D.J. Nunez, J.N. Haselden, C.J. Waterfield, *Biomarkers* 8 (2003) 240.
[94] R.E. Hurd, B.K. John, *J. Magn. Reson* 91 (1991) 648.
[95] A.W. Nicholls, J.K. Nicholson, J.K. Haselden, C.J. Waterfield, *Biomarkers* 5 (2000) 410.
[96] M. Coen, S.U. Rüpp, J.C. Lindon, J.K. Nicholson, F. Pognan, E. Lenz, I.D. Wilson, *J. Pharm Biomed. Anal.* 35 (2004) 93.
[97] A. Bax, M.F. Summers, *J. Am. Chem. Soc.* 108 (1986) 2093.
[98] M.-E. Dumas, C. Canlet, F. André, J. Vercauteren, A. Paris, *Anal. Chem.* 74 (2002) 2261.
[99] O. Cloarec, M.E. Dumas, J. Trygg, A. Craig, R.H. Barton, J.C. Lindon, J.K. Nicholson E. Holmes, *Anal. Chem.* 77 (2005) 517.
[100] B. Lefebvre, S. Golotvin, L. Schoenbachler, R. Beger, P. Price, J. Megyesi, R. Safirstein, Intelligent Bucketing for Metabonomics – Part 1, Poster, http://www.acdlabs.com/download/publ/2004/enc04/intelbucket.pdf.
[101] B. Lefebvre, R. Sasaki, S. Golotvin and A. Nicholls, Intelligent Bucketing for Metabonomics – Part 2, Poster, http://www.acdlabs.com/download/publ/2004/intelbucket2.pdf.
[102] L. Schnackenberg, R.D. Beger, Y. Dragan, *Metabolomics*, 1 (2005) 87.
[103] G. Lee and D.L. Woodruff, *Anal. Chim. Acta* 513 (2004) 413.
[104] H.C. Keun, T.M.D. Ebbels, H. Antti, M.E. Bollard, O. Beckonert, E. Holmes, J.C. Lindon and J.K. Nicholson, *Anal. Chim. Acta* 490 (2003) 265.
[105] H. Antti, M.E. Bollard, T. Ebbels, H. Keun, J.C. Lindon, J.K. Nicholson and E. Holmes, *J. Chemometrics* 16 (2002) 461.
[106] J.T. Brindle, J.K. Nicholson, P.M. Schofield, D.J. Grainger, E. Holmes, *Analyst* 128 (2003) 32.
[107] O. Beckonert, M.E. Bollard, T.M.D. Ebbels, H.C. Keun, H. Antti, E. Holmes, J.C. Lindon, J.K. Nicholson, *Anal. Chim. Acta* 490 (2003) 3.
[108] T. Ebbels, H. Keun, O. Beckonert, H. Antti, M. Bollard, E. Holmes, J. Lindon, J. Nicholson, *Anal. Chim. Acta* 490 (2003) 109.
[109] The Merck Veterinary Manual, http://www.merckvetmanual.com/mvm/index.jsp?cfile=htm/bc/150204.htm.
[110] P. Jatlow, S. McKee, S.Ó. Malley, *Clin. Chem.* 49 (2003) 1932.
[111] Sterlin Reference Laboratories, http://www.sterlingreflabs.com/thccrearatios.html.
[112] G. Fauler, H.J. Leis, E. Huber, C. Schellauf, R. Kerbl, C. Urban and H. Gleispach, *J. Mass Spectrom.* 32 (1997) 507.
[113] E. Holmes, P.J.D. Foxall, J.K. Nicholson, G.H. Neild, S. M. Brown, C.R. Beddell, B.C. Sweatman, E. Rahr, J.C. Lindon, M. Spraul, P. Neidig, *Anal. Biochem.* 220 (1994) 284.
[114] J.P. Shockor and E. Holmes, *Curr. Top. Med. Chem.* 2 (2002) 35–51.

[115] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, *European Patent Application* 05006476.5 (24.03.2005).
[116] AMIX Metabolite Database, Bruker BioSpin AG, Karlsruhe, Germany.
[117] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, *Anal. Chem.* 78 (2006) 3551.
[118] S.W. Provencher, *NMR Biomed.* 14 (2001) 260.
[119] Chenomx Eclipse, Chenomx Inc. Edmonton, Canada.
[120] C.D. Eads, I. Noda, J. *Am. Chem. Soc.* 124 (2002) 1111.