Processing and analysis of NMR data
Impurity determination and metabolic profiling

Jenny Forshed

Doctoral Thesis
Department of Analytical Chemistry
Stockholm University
2005
Processing and analysis of NMR data
Impurity determination and metabolic profiling

Doctoral Thesis, 2005

Jenny Forshed
jenny.forshed@anchem.su.se
Department of Analytical Chemistry
Stockholm University
SE-106 91 Stockholm

Academic dissertation for the Degree of Doctor of Philosophy in Analytical Chemistry at Stockholm University to be publicly defended on Thursday 1 December 2005 at 13:00 in Magnélisalen, Kemiska övningslaboratoriet, Svante Arrhenius v. 10-12.

© Jenny Forshed 2005, pp 1-95.
Cover by Mina, Folke and Jenny Forshed
Akademitryck AB, Edsbruk, Sweden
till Mina, Folke och Östen
Abstract

This thesis describes the use of nuclear magnetic resonance (NMR) spectroscopy as an analytical tool. The theory of NMR spectroscopy in general and quantitative NMR spectrometry (qNMR) in particular is described and the instrumental properties and parameter setups for qNMR measurements are discussed. Examples of qNMR are presented by impurity determination of pharmaceutical compounds and analysis of urine samples from rats fed with either water or a drug (metabolic profiling). The instrumental parameter setup of qNMR and traditional data pre-treatments are examined. Spectral smoothing by convolution with a triangular function, which is an unusual application in this context, was shown to be successful regarding the sensitivity and robustness of the method in paper II. In addition, papers III and IV comprise the field of peak alignment, especially designed for ¹H-NMR spectra of urine samples. This is an important preprocessing tool when multivariate analysis is to be applied. A novel peak alignment method was developed and compared to the traditional bucketing approach and a conceptually different alignment method.

Univariate, multivariate, linear and nonlinear data analyses were applied to qNMR data. In papers I–II, calibration models were created to examine the potential of qNMR for these applications. The data analysis in papers III–VI was mainly explorative. The potential of data fusion and data correlation was examined in order to increase the possibilities of analysing the highly complex samples from metabolic profiling (papers V–VI). Data from LC/MS analysis of the same samples were used with the ¹H-NMR data in different ways. Correlation analyses between the ¹H-NMR data and the drug metabolites identified from the LC/MS data were also performed. In this process, data fusion proved to be a valuable tool.
Processing and analysis of NMR data
Denna avhandling beskriver hur NMR (nuclear magnetic resonance) spektroskopi kan användas för kvantifiering i analytisk kemi. Teorin för NMR-spektroskopi i allmänhet och kvantitativ NMR (qNMR) i synnerhet är beskriven och instrumentella förutsättningar och parameterinställningar för qNMR diskuteras. Förreningsanalys av två olika läkemedelskomponenter och analys av urin från råttor som matats med vatten eller läkemedel (metaboliska profiler) ges som exempel på qNMR. I de två första artiklarna utvärderades instrumentella parameterinställningar för qNMR och betydelsen av traditionell databehandling. Utjämning av spektra (”spectral smoothing”) med hjälp av faltning med en triangulär funktion visade sig förbättra både känslighet och robusthet hos qNMR-metoden i artikel II, dock syns få sådana applikationer i litteraturen. Artiklarna III–IV omfattar topplinjering, att anpassa analytiska signaler uteftes en rät linje, anpassat för ¹H-NMR data från urinprov. Detta är en viktig förprocessning av data vid multivariat analys. En ny metod för topplinjering utvecklades och jämfördes med den traditionella (integrering av förutbestämda, lika stora segment) och en konceptuell topplinjningsmetod.

The work on this thesis has been carried out in the borderland between the fields of \textit{nuclear magnetic resonance} (NMR) spectroscopy – mostly from a structure elucidation perspective, \textit{analytical chemistry} – aiming for low detection limits and high accuracy and precision, and \textit{chemometrics} – including the large field of multivariate data analysis. These three fields of science do not often come together in the same department, and even less often in the same person. This PhD thesis, however, aims to combine these fields of knowledge in order to create opportunities for novel research in the field of quantitative NMR spectrometry (qNMR).

As I come from a background in analytical chemistry and chemometrics, the work of this thesis started with the basic theories and practice of NMR spectrometry, which was then a new technique to me. The theory and background sections in this thesis were written more as a handbook for myself at that time. The thesis also contains a summary of the six papers I have completed during my PhD studies at the Department of Analytical Chemistry at Stockholm University. The laboratory work was done at AstraZeneca in Södertälje, which provided all the laboratory materials, chemicals and people with expertise in NMR spectroscopy.
Processing and analysis of NMR data
## Contents

List of papers .................................................. 13
Abbreviations and notations .................................. 15
1  Introduction .................................................. 19
2  Nuclear magnetic resonance theory ...................... 21
   Spin ......................................................... 21
   Applying a magnetic field .................................. 22
   The laboratory frame ...................................... 24
   The rotating frame of reference ......................... 25
   NMR data acquisition .................................... 30
   NMR data processing ..................................... 33
3  NMR for quantitative analysis ............................. 37
   Background ............................................... 37
   qNMR methods ............................................ 39
   The sensitivity of NMR spectrometry .................. 39
   Accuracy and precision of qNMR ......................... 43
   Discussion ............................................... 44
4  qNMR applications .......................................... 45
   Impurity determination .................................. 45
   Metabolic analyses ....................................... 50
5  Chemometrics ................................................ 53
   Principal component analysis ............................ 53
   Calibration ............................................... 55
   Data preprocessing ....................................... 59
   Validation ............................................... 60
6  Peak alignment ............................................. 63
   Profile alignment or peak alignment .................... 65
   Target for alignment ..................................... 67
   Optimisation .............................................. 67
   NMR data compression ................................... 70
Discussion .......................................................... 70
7  $^1$H-NMR + LC/MS ............................................. 71
   Data fusion ...................................................... 72
   Data correlation ............................................... 75
   Discussion .................................................... 77
8  Conclusions ...................................................... 79
9  Future perspectives ............................................... 81
   Acknowledgements .......................................... 83
   References .................................................... 85
List of papers

I. *NMR and Bayesian regularized neural network regression for impurity determination of 4-aminophenol*
Forshed, J., Andersson, F.O. and Jacobsson, S.P.,

*The author is responsible for the experimental work and for writing the paper.*

II. *Quantification of aldehyde impurities in poloxamer by ¹H-NMR spectrometry*
Forshed, J., Erlandsson, B. and Jacobsson, S.P.

*The author is responsible for the experimental work, the calculations and the writing of the paper.*

III. *Peak alignment of NMR signals by means of a genetic algorithm*
Forshed, J., Schuppe-Koistinen, I. and Jacobsson, S.P.

*The author is responsible for the experimental work, developing the algorithm and writing the paper.*

IV. *A comparison of methods for alignment of NMR peaks in the context of cluster analysis*
Forshed, J., Torgrip, R.J.O., Åberg, K.M., Karlberg, B., Lindberg, J. and Jacobsson, S.P.

*The author is responsible for the comparison idea, developing the algorithm for the segmentwise method and writing the main part of the paper.*
V. *Evaluation of different techniques for fusion of LC/MS and 1H-NMR data*
Forshed, J., Idborg, H. and Jacobsson, S.P.

*The author is, together with H. Idborg, responsible for the ideas, implementing the code, and writing the paper.*

VI. *Enhanced multivariate analysis by correlation scaling and fusion of LC/MS and 1H-NMR data.*
Forshed, J., Stolt, R., Idborg, H. and Jacobsson, S.P.

*The author is responsible for the major ideas (together with H. Idborg), some of the computations and writing the paper.*
Abbreviations and notations

Abbreviations

$^1$H-NMR  proton-NMR
ADC  analogue-to-digital converter
CC  correlation coefficient
COW  correlation optimised warping
DFT  discrete Fourier transform
dsp  digital signal processing
DTW  dynamic time warping
FID  free induction decay = $S(t)$
FT  Fourier transform
FW  fuzzy warping
GA  genetic algorithm
GC  gas chromatography
IR  infrared spectroscopy
LC  liquid chromatography
LOD  limit of detection
LS  least-squares regression
LW  local warping
MLR  multiple linear regression
MS  mass spectrometry
NMR  nuclear magnetic resonance
NN  neural network
N-PLS  N-way PLS
OPA  outer product analysis
PAGA  peak alignment by means of a genetic algorithm
PARAFAC  parallel factor analysis
PARS  peak alignment by reduced set mapping
PC  principal component
PCA  principal component analysis
PCR  principal component regression
PLF  partial linear fit
PLS  partial least squares or projection to latent structures
PLS-DA  PLS-discriminant analysis
ppm  parts per million (the units of the horizontal scale in an NMR spectrum)
PTW  parametric time warping
PWA  piecewise alignment
qNMR  quantitative NMR
RF  radio frequency
RMS  root-mean-square
RMSE  root-mean-square error
RMSEP  root-mean-square error of prediction
S/N  signal-to-noise ratio
TMS  tetramethyl silane
TMSP  (sodium 3-(trimethylsilyl)propionate-2,2,3,3-d₄)

Notations

Generally the algebraic notation is used as follows: small letters represent scalars, small bold letters represent vectors, and bold capitals represent data matrices.

\( A(\omega) \)  absorption spectra
\( at \)  acquisition time
\( b \)  regression vector
\( b \)  receiver bandwidth
\( B_0 \)  external magnetic field
\( B_1 \)  RF pulse
\( D(\omega) \)  dispersion spectra
\( E \)  residual matrix
\( E \)  energy
\( f \)  filling factor
\( F(\omega) \)  the spectrum as a function of frequency
\( F_{\text{sampling}} \)  sampling frequency
\( F_{\text{signal}} \)  the highest frequency signal
\( b \)  Planck’s constant \((= 6.62618 \times 10^{-34} \text{ Js})\)
\( I \)  spin quantum number
\( I(\omega) \)  imaginary part of spectrum
\( k \)  the Boltzmann constant \((= 1.380 \times 10^{-23} \text{ J/K})\)
\( m \)  magnetic quantum states
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>net magnetisation</td>
</tr>
<tr>
<td>$N$</td>
<td>number of detectable nuclei (magnetically equivalent)</td>
</tr>
<tr>
<td>$n$</td>
<td>the noise figure of the amplifier</td>
</tr>
<tr>
<td>$nt$</td>
<td>the number of transients (scans)</td>
</tr>
<tr>
<td>$p$</td>
<td>angular moment</td>
</tr>
<tr>
<td>$P$</td>
<td>loading matrix</td>
</tr>
<tr>
<td>$Q$</td>
<td>the quality factor of the RF coil</td>
</tr>
<tr>
<td>$R(\omega)$</td>
<td>real part of spectrum</td>
</tr>
<tr>
<td>$S$</td>
<td>spectrum to be aligned</td>
</tr>
<tr>
<td>$S(t)$</td>
<td>the NMR signal as a function of time (= FID)</td>
</tr>
<tr>
<td>$rd$</td>
<td>relaxation delay</td>
</tr>
<tr>
<td>$rt$</td>
<td>the pulse repetition time</td>
</tr>
<tr>
<td>$T$</td>
<td>sample temperature</td>
</tr>
<tr>
<td>$T$</td>
<td>scores matrix</td>
</tr>
<tr>
<td>$T_1$</td>
<td>spin-lattice relaxation time constant</td>
</tr>
<tr>
<td>$T_2$</td>
<td>spin-spin relaxation time constant</td>
</tr>
<tr>
<td>$V_s$</td>
<td>volume of the sample</td>
</tr>
<tr>
<td>$w$</td>
<td>weights</td>
</tr>
<tr>
<td>$\beta$</td>
<td>pulse angle</td>
</tr>
<tr>
<td>$\mu$</td>
<td>magnetic moment</td>
</tr>
<tr>
<td>$\nu_0$</td>
<td>the Larmor frequency [Hz]</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>phase error</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>the Larmor frequency [rad/s]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>the magnetogyratic ratio</td>
</tr>
<tr>
<td>$\tau$</td>
<td>pulse width [$\mu$s]</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>wavelength</td>
</tr>
</tbody>
</table>
Processing and analysis of NMR data
1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for determination of molecular structures as well as chemical and physical properties. It allows the detection of the most common atoms in organic compounds, i.e. carbon and hydrogen. Experience of quantitative determinations by NMR spectrometry (qNMR) is still, however, fairly limited.

One limiting factor of NMR spectroscopy as a quantitative tool is that it is considered to be a relatively insensitive technique. However, recent developments of the technique have led to fairly robust instrumentation and low limits of detection in the picomole range have been reported. Consequently, NMR spectroscopy has attracted a growing interest and has become a viable analytical technique.

The scope of this thesis

The aim of this thesis has been to address the possibilities and limitations in qNMR and to make a contribution to this field of research. While the requirements for an accurate and precise analysis with qNMR has been investigated in the literature, there is a large demand for further investigation.

In analytical chemistry reference is frequently made to what is known as the analytical chain:

Sampling → sample preparation → instrumental analysis → data analysis.

This work includes all the steps in this chain, with the main emphasis on the last one. Linear and nonlinear, univariate and multivariate calibration methods have been employed, as well as principal component cluster analysis. The substantial need for methods to correct peak shifts, especially when analysing biological samples, were realised and explored. In response to this need, such methods
were developed. Furthermore, techniques for data fusion and data correlation of $^1$H-NMR and LC/MS data from metabolic samples have been investigated and evaluated.

This thesis will cover some basic theory of NMR spectroscopy, to an extent comprehensive enough to enable one to understand the significance of qNMR and be able to use it. This will be followed by the special case of analysing low-level components by qNMR in general. $^1$H-NMR spectrometry for impurity determination (papers I–II) and urine analysis (papers III–VI) will be described in more detail.

The final chapters will include the chemometric methods that have been used in the papers. Separate chapters have also been devoted to peak alignment, which has been dealt with in papers III and IV, and data fusion, covered in papers V and VI.

My experiences, results and conclusions from the work in the papers will be addressed in the chapters when considered relevant.
2 Nuclear magnetic resonance theory

This chapter includes the NMR theory needed for understanding quantitative measurements by one-dimensional NMR spectrometry. Some quantum mechanics and classical mechanics will be addressed to facilitate the understanding of the theories. More extensive descriptions of the NMR theory can be found in the references.4,9-17

History

The theoretical bases of NMR spectroscopy were proposed by W. Pauli as early as 1924. However, it was not until 1946 that Bloch and Purcell independently showed that nuclei absorb electromagnetic radiation in a strong magnetic field.18,19 They shared the Nobel prize in Physics for this work in 1952. In 1953 the first high-resolution NMR spectrometer was presented (a continuous wave instrument), although it was not until about 1970 that the first Fourier transform (FT) NMR instrument was available on the market. At least eight Nobel prizes in physics and two in chemistry have been awarded to scientists for their work in the field of magnetic resonance.18,17

Spin

Atomic nuclei have a spin quantum number, I. If I differs from zero, the nucleus possess a magnetic moment (μ) that may interact with an external magnetic field:

\[ \mu = \gamma p, \]  

where γ is the magnetogyric ratio and p is the angular momentum. γ will differ between nuclei and may be positive or negative. The possible values of p are quantised and depend on I. The interrelation between the nuclear spin and μ leads to 2I+1 observable magnetic quantum states (m)...

1Also denoted as gyromagnetic
Applying a magnetic field

The fact that nuclei are magnetic allows us to study them by means of NMR spectroscopy. A nucleus with \( I \neq 0 \) will interact with a magnetic field and can absorb or emit electromagnetic radiation at certain resonance conditions. Some of the most important nuclei in organic chemistry and biochemistry (\(^{1}\text{H}, ^{13}\text{C}, ^{19}\text{F}\) and \(^{31}\text{P}\)) have \( I = 1/2 \) and therefore two allowed magnetic quantum states: \( m = +1/2 \) and \( -1/2 \). The \(^{1}\text{H}\) nuclei will be the model in the following theoretical descriptions.

In the absence of an external magnetic field, the nuclear spins are disordered. Applying a magnetic field \((B_0)\) will make the \(^{1}\text{H}\) nuclei orient with a small access in the lower energy state \((m = +1/2)\) according to the Bolzmann distribution:

\[
\frac{N_{-1/2}}{N_{+1/2}} = e^{(\frac{\Delta E}{kT})},
\]

where \( k \) is the Bolzmann constant, \( T \) is the absolute temperature and \( N \) represents the population at the different stages. The excess of nuclei in the lower energy state is small (64 nuclei per million (for \(^{1}\text{H}\)) in a 400 MHz instrument). The difference in energy \((\Delta E)\) is proportional to the strength of the magnetic field according to:

\[
\Delta E = \gamma b \frac{B_0}{2\pi},
\]

where \( b \) is Planck’s constant and \( B_0 \) the external/applied magnetic field. This creates the basic prerequisites for NMR spectrometry.

Energy transitions

The transition between lower and upper energy levels in NMR spectroscopy is analogous to absorption in other spectroscopic methods, although wavelengths and frequencies differ (Figure 1). However, the discrete energy levels between which the energy transitions take place are created artificially in NMR spectroscopy, by placing the nuclei in a magnetic field.
Transitions between energy states are brought about by the absorption or emission of electromagnetic radiation of a frequency $\nu_0$ with the corresponding energy:

$$\Delta E = h\nu_0$$ (4)

From equations (3) and (4) it follows that the frequency of radiation required for a transition is:

$$\nu_0 = \frac{\gamma B_0}{2\pi}$$ (5)

Applied energy with the frequency $\nu_0$ will induce both upward and downward spin transitions of nuclei. As long as there are a greater number of nuclei in the lower energy state, the energy absorption will exceed the emission and a signal can be observed. However, continuously applied energy of the “right” frequency will cause the spin populations to become equal and no signal will be observed, due to competing relaxation processes. This phenomenon is called saturation and may be utilised when a specific signal (for example, water) needs to be suppressed, e.g. when urine samples are to be analysed.
The laboratory frame

To illustrate the measurement of energy transitions, a classical description of the behaviour of a charged particle in a magnetic field is illuminative. A coordinate system with $B_0$ along the z axis, and the x and y axes describing the horizontal plane, is introduced to facilitate the following theories. This is called the laboratory frame of reference.

Larmor frequency

When a magnetic field ($B_0$) is applied to a nucleus with a magnetic moment ($I \neq 0$), it starts to precess about the z axis because of the gyroscopic effect, Figure 2.

\[ \omega_0 = \gamma B_0 \]  

$\omega_0$ [rad/s] is equivalent to $\nu_0$ [Hz] (equation (5)). The frequency, $\nu_0$, for the $^1$H nucleus is actually what is referred to when the strength of an NMR magnet is denoted in MHz.

Net magnetisation

The ensemble of spins in a sample generates an average magnetisation or net magnetisation ($M_0$) in the direction of the z axis upon interaction with a magnetic field ($B_0$), Figure 3. The individual spins may point in any direction in the xy plane, so their component in the xy plane evens out.

---

† As when a toy gyroscope starts to precess/wobble because of the gravity force.
$M_0$ is the sum of all the spins and is directly proportional to the population difference between the low and high energy levels, $N_{\frac{1}{2}}^+ - N_{\frac{1}{2}}^-$. Although the magnetic moment ($\mu$) in each nucleus is quantised, $M_0$ is continuous since it reflects the whole population of nuclei.

**The rotating frame of reference**

To further facilitate the illustration of energy absorption and to describe the single pulse experiment, we change the angle of view. When studying the nuclei in the laboratory coordinate system, the laboratory frame of reference, each nucleus is precessing at the Larmor frequency ($\nu_0$ or $\omega_0$) as described above. By subtracting the laboratory frame we can define a new *rotating frame of reference* in which the precession around the z axis is cancelled out. Moving our view into this rotating coordinate system makes it easier to illustrate the observed energy transitions. This is further described below in “Analogue filtering”.

**RF pulse**

When a radio frequency (RF) pulse ($B_1$) at the Larmor frequency is applied perpendicular to $B_0$, we achieve a resonant condition and the system absorbs energy ($\Delta E$). The populations of $N_{\frac{1}{2}}^+$ and $N_{\frac{1}{2}}^-$ are changed according to the Bolzmann equation (2) and $M_0$ changes direction. Application of the RF pulse is equivalent to application of a magnetic field along the x or y axis. This rotates $M_0$ away from the z axis into the xy plane to give $M_{xy}$, Figure 4.
Figure 4: The direction of the net magnetisation ($M_0$) is changed with the angle $\beta$ when an RF pulse ($B_1$) is applied to give $M_{xy}$.

The changed angle ($\beta$) of $M_0$ depends on the field strength of the RF pulse ($B_1$) and the duration $\tau$ (typically $\mu$s) according to:

$$\beta = \gamma B_1 \tau$$  \hspace{1cm} (7)

**Relaxation**

After applying the RF pulse, the system will relax back and Bolzmann equilibrium (2) will be re-established. The relaxation can be described by two different processes, the disappearance of $M_{xy}$ and the appearance of $M_z$ until equilibrium is reached and $M_{xy} = 0$ and $M_z = M_0$, as illustrated in Figure 5.

Figure 5: The directions of $T_1$ and $T_2$ are shown in a schematic representation of the relaxation of $M_{xy}$ back to $M_0$. 

---

Processing and analysis of NMR data
The process restoring $M_z$ is called the spin-lattice (or longitudinal) relaxation and is defined by the relaxation time constant $T_1$. The energy of the excited state dissipates into molecular vibrations, rotations etc, until $M_0$ has returned to its original value on the z axis. The “lattice” denotes the surroundings of the molecule, hence $T_1$ will depend on, for example, the viscosity of the sample (low mobility will give a high $T_1$). The loss of $M_{xy}$ is due to intrinsic molecular properties and to instrumental imperfections. When excluding the contributions from instrumental imperfections, we may define the spin-spin (or transverse) relaxation time constant $T_2$, which includes interactions between neighbouring nuclei. This results in loss of xy magnetisation until no preferred orientation in the transverse plane remains. When we include the contribution of instrumental imperfections to the loss of xy magnetisation, we define $T_2^*$. $T_2^*$ determines the spectral resolution (peak width) and depends on, for example, inhomogeneities in the magnetic field. $T_2^*$ is always less than $T_2$ and $T_2$ is always less than or equal to $T_1$. For many liquids, $T_1$ and $T_2$ are about the same and in the order of seconds.

In contrast to other spectroscopic methods, it is not possible to accurately measure the absolute values of transition frequencies. Instead, the difference between two resonance frequencies is measured. Each nucleus, rotating with its own resonance frequency $\omega_i$ differs from the Larmor frequency $\omega_0$ by a few parts per million (ppm) in a span of about 10 ppm for the $^1\text{H}$-nuclei. This relaxing magnetisation induces a measurable current in the xy plane, the free induction decay (FID), Figure 6.
The FID is composed of exponentially decaying signals, expressed in simple form by

\[
\text{FID} = M e^{-\gamma B_0 t} = \sum_{i=1}^{K} \cos(\omega_i t + \phi) e^{-\gamma B_0 t}, \quad 0 < t < a_t, \tag{8}
\]

where \(\omega_i\) represents the angular frequency of the \(i\)th of \(K\) magnetically different nuclei in the same sample. \(\phi\) is the phase error, which will be corrected for later, \(t\) is the time and \(a_t\) is the acquisition time.
Quadrature detection

To distinguish between vectors rotating faster or slower than the carrier frequency, the FID is sampled in two directions perpendicular to each other (Figure 6). This is called quadrature detection and results in two signals with phases shifted 90°. These two signals are subsequently treated as one real $R(t)$ and one imaginary part $I(t)$:

$$I(t) = \sum_{i=1}^{K} \sin (\omega_i t + \varphi) e^{-\gamma T_2}$$  \hspace{1cm} (9)

$$R(t) = \sum_{i=1}^{K} \cos (\omega_i t + \varphi) e^{-\gamma T_2}.$$  \hspace{1cm} (10)

One signal of the FID may be illustrated as a combination of equations (9) and (10):

$$S(t) = R(t) + iI(t) = \cos (\omega t + \varphi) + i \sin (\omega t + \varphi).$$  \hspace{1cm} (11)

Chemical shifts

Resonance frequencies from the same isotopes will differ depending on different molecular (magnetic) surroundings, which give the chemical shifts. Therefore, nuclei with different surroundings result in peaks at different frequencies in an NMR spectrum. In $^1$H-NMR spectroscopy TMS (tetramethyl silane) or TMSP (sodium 3-(trimethylsilyl)propionate-2,2,3,3-d$_4$) are usually set as the reference value of 0 ppm (Figure 7). This relative scale implies that each peak will appear at the same chemical shift, regardless of the strength of the magnetic field.

Coupling constants

The energy levels of a nucleus may also be affected by the spin state of nuclei nearby, in which case the nuclei are said to couple to each other. This will give split peaks and the magnitude of this separation depends on the strength of the coupling interaction, the coupling constant (Figure 7). These coupling patterns are crucial for the determination of chemical structures but will not be described further here.
Figure 7: A sketch of a $^1$H-NMR spectrum of ethane (CH$_3$-CH$_3$) showing the chemical shift scale ($\delta$) and the coupling constant (J). TMS assigns the reference value $\delta=0$.

### NMR data acquisition

#### Locking and shimming

The frequency of a resonance in the sample (usually deuterium in the solvent) is compared to a fixed frequency derived from the master clock used to generate the spectrometer frequency. Any deviation due, for example, to small changes in the magnetic field is converted into a correction to $B_0$. This is called field frequency locking.$^4,21$

The lack of homogeneity in the magnetic field is compensated for by spinning the sample and/or by shimming. Shimming is the automatic or manual adjustment of certain shim coils or magnetic gradients until the instrumental contributions to the observed line width are minimised.$^4,21$

#### Analogue filtering

We are interested in spectral frequencies ($\omega_i$) relative to some reference frequency ($\omega_0$), as stated above. The NMR signal from the laboratory frame of reference is demodulated by subtracting the reference frequency ($\omega_0$) by a bandpass filter, leaving a signal which then is sampled and digitised. This is comparable with how FM radio stations broadcast in the MHz range. The audio signal desired is superimposed on a high-frequency carrier signal, and this carrier frequency is tuned into the radio receiver and subsequently subtracted to leave a signal consisting of frequencies audible to the human ear. $\omega_0$ is adjusted by tuning (to find the right frequency) and matching (impedance).$^{21}$
Analogue to digital conversion

The FID collected by the detection coils in the xy plane (Figure 6) is converted to digital signals. An analogue-to-digital converter (ADC) is characterised by its sampling rate and dynamic range. The Nyquist sampling theorem\(^{22}\) says that the rate of sampling \(F_{\text{sampling}}\) must be at least twice the frequency of the highest frequency signal \(F_{\text{signal}}\) that is to be detected, i.e. at least two points per cycle of the highest frequency signal according to

\[
F_{\text{sampling}} \geq 2 F_{\text{signal}}
\]  \hspace{1cm} (12)

The dynamic range of the ADC gives the limit of the smallest signals that are measurable in the presence of a strong signal and may be a source of error/noise when there are large concentration differences.\(^{1,22}\) This is the case, for example, in impurity determinations and biological samples containing a substantial amount of water. A typical dynamic range is 16 bits, which corresponds to an intensity resolution of \(1/2^{16}\). It is important to set the receiver gain of the instrument so that the dynamic range is fully utilised, but to avoid overflow, which results in a truncated FID. Theoretically, the height of the smallest detectable peak is then \(1/2^{16-1} = 30\) µmol per mole of the main peak. This is also shown to be the practical limit in paper I, where the influence of the dynamic range is further discussed.

Digital filtering

Digital signal processing (DSP) or digital filtering represents processing for extending or enhancing the spectrometer capabilities, particularly when analysing small signals in the presence of large ones. The general steps involved are oversampling, filtering and downsampling.\(^{23}\)

**Oversampling** is a technique borrowed from audio technology and has proved to be useful in NMR technology. The oversampling rate describes how many times \(q\) greater the actual sampling rate is compared to the Nyquist sampling rate (equation (12)). Theoretically, an oversampling factor of \(q\) gives a gain in the effective dynamic range of the ADC by \(\log_2(q)\) bit. For instance, with an oversampling of 8, a gain of 3 bits in ADC resolution is obtained.\(^{22}\)
After the oversampled data is acquired, a digital filter is applied to the FID to reduce noise or extraneous signals outside the final spectral width desired. This consists of signals with frequencies higher than the Nyquist frequency, which will be “undersampled” and consequently appear at the wrong frequency in the spectrum.

Downsampling of the filtered data is then done to reduce the data size, which is generally reset to the values of the selected spectral width and the Nyquist sampling frequency.

Signal averaging

In the pulsed NMR technique, the responses from multiple experimental repetitions may be co-added, a technique known as signal averaging. Like infrared (IR) spectroscopy, NMR spectroscopy fits the definition of a “detector noise limited” measurement, which means that the process of signal averaging results in an increase of the signal-to-noise ratio by the square root of the number of repetitions.\(^{24}\)

\[
\frac{S}{N} \propto \sqrt{nt}, \tag{13}
\]

where \(\frac{S}{N}\) is the signal-to-noise ratio and \(nt\) is the number of transients (scans).

Important parameters to be set for a repeated pulse experiment are the acquisition time \((at)\) during which the FID is collected and the relaxation delay \((rd)\), see Figure 8. The choice of \(at\) and \(rd\) is influenced by the relaxation processes \(T_2^*\) and \(T_1\).

Figure 8: An example of two repeated RF pulses \((p)\) from the transmitter (upper) with acquisition time \((at)\), during which the FID is collected, and relaxation delay \((rd)\) of the receiver (lower).
NMR data processing

Zero filling
The spectral resolution could be enhanced by zero filling (zero padding) of the FID. The information content in the FID is not increased but the distribution between the real and imaginary parts of the spectrum is changed. A rule of thumb is to zero fill by at least double the number of data points to regain the spectral resolution after Fourier transformation, when the imaginary part of the spectrum is eliminated.

Apodisation
More information from the spectrum may be obtained if the FID is multiplied by a suitable apodisation function before Fourier transformation. The Lorenzian line shapes in spectra may be changed by multiplying the exponentially decaying FID by a function. This will affect the spectral resolution as well as S/N.

A negative exponential function will emphasise the early part of the FID with the strongest signal and suppress the late part, which consists mostly of noise. Such a function would increase S/N but also give line broadening since it would simulate faster relaxation ($T_2^*$). However, in general, broader peaks will be covered with more data points which help to obtain more precise integral values. $e^{-bt}$ is a so-called matched filter (compare with equation (8)) where $b$ is called the line-broadening factor. The matched filter is evaluated in papers I–II.

Besides the negative exponential function, an abundance of various functions have been used together with the FID. Examples are gaussian, lorenzian-to-gaussian transformation, sine bell and trapezoidal function. No single apodisation is optimal for all signals in spectra and different functions will enhance different phenomena, such as S/N or resolution.

Convolution
Furthermore, convolution or spectral smoothing could be performed on the spectrum in various ways. To give one example, a one-dimensional convolution with a triangular function means
that the spectra are filtered by a triangular window. The convolution theorem says that convolving two vectors is equal to multiplying their Fourier transforms. Thus, the convolution with a triangular function is the same as multiplying the FID by the Fourier transform of a triangular function (an exponentially decaying cosine curve), which also enhances the very first part of the FID. This has been tried and evaluated in paper II.

**Fourier transformation**

One aim of NMR data processing is to calculate the frequencies ($\omega_i$) from equation (8) and obtain a spectrum. The most usual method for converting the NMR signal from the time domain into a frequency domain is Fourier transformation (FT). Fourier-transforming the signal $S(t)$ from equation (11) gives a function of frequency, $F(\omega)$, also consisting of one real and one imaginary part:

$$F(\omega) = R(\omega) + i I(\omega)$$  \hspace{1cm} (14)

Since we have discrete values from the ADC, a discrete FT (DFT) must be applied. DFT requires linear sampling, data sampling at regular time intervals until the signal is close to zero, and truncation of the FID will lead to artefacts.

Hodgkinson and Hore have pointed out possible ways in which nonlinear sampling of data can be more efficient. A frequency spectrum could then be obtained by alternative data processing procedures such as least squares methods (LS), linear prediction (LP), the Bayesian method or the maximum entropy method (MEM). These alternative methods are recommended for quantitative problems in the literature, although only a few have been employed.
Phase correction

The FID $S(t)$ as well as the Fourier-transformed spectrum $F(\omega)$ has been described as a signal consisting of one real part $R$ and one imaginary part $I$, equations (11) and (14). Due to phase shifts, $R(\omega)$ and $I(\omega)$ consist not just of the absorption spectra $A(\omega)$ and the dispersion spectra $D(\omega)$ respectively, but a linear combination of the two according to:

\begin{align}
R(\omega) &= A(\omega)\cos(\varphi) + D(\omega)\sin(\varphi) \\
I(\omega) &= A(\omega)\sin(\varphi) - D(\omega)\cos(\varphi)
\end{align}

(15)

where $\varphi$ represents the phase error. A linear combination of $R(\omega)$ and $I(\omega)$ will give the pure absorption and dispersion spectra, according to

\begin{align}
A(\omega) &= R(\omega)\cos(\varphi) + I(\omega)\sin(\varphi) \\
D(\omega) &= R(\omega)\sin(\varphi) - I(\omega)\cos(\varphi)
\end{align}

(16)

This is obtained by phase correction, which is most often performed manually. Since the phase correction is frequency dependent, $\varphi$ is defined as:

$$\varphi(\omega) = \varphi_0 + \omega \varphi_1$$

(17)

where the constant phase error, $\varphi_0$, arises from the inability of the spectrometer to detect the signals from the two detector coils in the exact $x$ and $y$ directions. Frequency-dependent linear phase errors, $\varphi_1$, arise because the detectors cannot detect the transverse magnetisation immediately after the RF pulse due to the risk of pulse breakthrough and the RF pulse being slightly off the Larmor frequency of every nucleus.\(^4\)
Processing and analysis of NMR data
3 NMR for quantitative analysis

Peak area determination is very common in routine proton NMR spectrometry. Accurately calculating integrals to determine the relative number of protons contributing to particular peaks is straightforward. For this purpose, an integral accuracy of 10–15% is satisfying and easily achieved. When NMR spectrometry is applied as a quantitative analytical tool, an accuracy of 1–2% or better is desirable. In this thesis NMR spectrometry for quantitative applications is denoted qNMR; however, this is not a generally accepted convention. This chapter will review the possibilities of, discuss the requirements for and hopefully give a deeper understanding of the applicability of qNMR.

Background

Early research in the qNMR field was done, among others, by Muller and Goldenson, Hollis, and Jungnickel and Forbes and by 1976 qNMR in pharmaceutical research had been reviewed. Several texts have discussed this subject in the last two decades and the notation QNMR was introduced in 1998.

The use of NMR spectrometry as a quantitative technique has been rather limited despite its many advantages, which include the following: Since it is able to detect small differences in a molecular structure, it is a very selective technique. It is also directly applicable to most samples, little or no sample preparation being required. For instance, samples with a complex background matrix such as urine, plasma, petrol, oil, beer or egg yolk extracts can be analysed virtually directly. It is also a non-destructive tool of analysis. One special advantage of NMR spectrometry is its property of being a primary method of measurement, a method with the highest metrological qualities whose operations can be completely described...
and understood. Thus the results can be accepted without reference to an independently determined standard of the same material but can be determined directly from the physical context. In theory, the peak intensity of each NMR signal exactly reflects the molar ratio of the nuclei present (i.e. the molar response is 1 for all nuclei of the same isotope), making the technique conceptually and technically simple for quantification.

Because organic compounds typically generate several peaks, the complexity of the spectra, especially for $^1$H-NMR spectrometry, may cause problems as well as bringing an advantage. A discouraging fact in qNMR analysis may be the large amount of parameters that have to be set. The most important ones are accounted for in this chapter. Compared to other analytical techniques, the major disadvantage of qNMR is its high limit of detection.

**qNMR vs other analytical techniques**

Despite its limitations, qNMR may rival chromatography in sensitivity, speed precision and accuracy, while also avoiding the need for a reference standard of each analyte. Examples where qNMR is comparable and even more accurate and precise than the standard liquid chromatography (LC) method are easily found in the literature. Other advantages of qNMR over LC include time for development and validation, robustness and analysis time (i.e. cost), together with the possibility of getting a good overall picture of all types of organic compounds in the sample.

Compared to mass spectrometry (MS), where individual events could be counted, NMR spectrometry has S/N ratios that are lower by many orders of magnitude. However, it does have some advantages over MS: normally, no separation (chromatographic or other) is required, no expensive, authentic reference samples are necessary and, in many cases, it is quicker and easier to perform.

There are no general rules for predicting whether analytical problems can be solved by NMR spectrometry. Pauli et al. demonstrated the strong demand for non-chromatographic alternatives in the quality assessment of reference compounds. The area of impurity determination (papers I–II) is suitable for qNMR since the technique is very selective. Favourable application areas of qNMR are also where it can replace complicated and time-consuming sample preparation and analysing methods, one example being found in paper II.
qNMR methods

Two principal methods may be used for qNMR: the absolute method with reference to an internal standard compound (a primary method) and the calibration method with reference to a calibration model.

(i) The absolute method

The intensity of a signal ($A$) is proportional to the total number of nuclei generating it ($N$). Comparing two signals, where one works as an internal standard ($i$), gives the relationship:

$$\frac{A_i}{A_u} = \frac{N_i}{N_u} \Leftrightarrow N_s = \frac{A_i N_i}{A_u},$$

which makes it easy to determine the number of nuclei generating the other signal ($s$) without any calibration. However, in order for equation (18) to be valid, the nuclei $s$ and $i$ have to be relaxed to the same extent or, easier to fulfil, fully relaxed.

(ii) The calibration method

NMR spectrometry can also be used without utilising the advantage of the equal molar response for all nuclei. However, this requires a calibration based on the internal standard technique. The equal molar response does not then have to be valid, and incomplete relaxations or peak interferences may be included in the model. The focus of the analysis will then be to optimise the S/N ratio for the particular quantification task. Calibration qNMR is used (more or less) in all the papers in this thesis, although it has been presented in the literature\textsuperscript{43,44,47,52-55} to a lesser extent than the absolute method.

The sensitivity of NMR spectrometry

The low sensitivity of NMR spectrometry\textsuperscript{26} is due particularly to the very small differences in energy that are measured. Low-frequency energies are also more difficult to detect than high-frequency ones. Moreover, only the small differences in population between the energy states (equation (2)) are observed in the experiment.
However, since the nuclei have a relatively long lifetime in the excited state (compared to other spectroscopic techniques), the resonance is well defined, resulting in narrow peaks, according to Heisenberg’s uncertainty principle.$^{4,13}$

New technical developments in the NMR instrumental field will provide greater sensitivity in the analyses. Imminent developments include higher magnetic field strengths. Also NMR probes with cooled transmitter/receiver coils and preamplifiers will increase the sensitivity due to reduction of electronic thermal noise. Probes cooled by cryogenic liquids such as liquid helium are called cryoprobes and will increase the S/N ratio approximately fourfold or more.$^{36}$ References$^{1,57}$ are also made to the possibilities of using microcoil technology in order to reach a limit of detection (LOD) down to picomole ($10^{-12}$) level.

**Signal-to-noise ratio**

The achievable signal-to-noise ratio (S/N) of an NMR single pulse experiment is a function of a number of parameters. These are summarised in equation (19)$^{58,59}$ and in Table 1, where the time parameters are also included.

$$S/N = \gamma N I (I + 1) \left( \frac{B_0}{T} \right)^{\frac{3}{2}} f \left( \frac{Q V}{b} \right)^{\frac{1}{2}} \frac{1}{n}$$  \hspace{1cm} (19)

Instrumental parameters determined by the available equipment are the filling factor ($f$), the magnetic field strength ($B_0$), receiver bandwidth ($b$), the quality factor of the RF coil ($Q$) and the noise figure of the amplifier ($n$). The filling factor is a measure of the fraction of the coil volume occupied by the sample and could be increased by fixing the RF coil directly onto the sample cell. This construction, however, makes it impossible to spin the sample to get improved magnetic field homogeneity. On the other hand, the magnetic field homogeneity in such a probe is superior to those with an unfixed coil. According to equation (19) an increase in magnetic field strength will increase S/N by a power of $3/2$, i.e. changing a magnet of 500 MHz into 600 MHz would give an increase in S/N of 30%. Instrumental factors affecting sensitivity are further discussed by Hoult.$^{60}$

$^1$To measure an energy difference accurately, one needs a long time. If the system relaxes rapidly, the time available is small and hence $\Delta E (=h\nu)$ is poorly defined and the observed NMR signals are wide.
Table 1: A summary and categorisation of the factors influencing NMR spectrometry S/N.\textsuperscript{16,59}

<table>
<thead>
<tr>
<th>Categorisation</th>
<th>Parameter</th>
<th>Possibility of S/N enhancement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Probe design</strong></td>
<td><strong>Parameter</strong></td>
<td><strong>Possibility of S/N enhancement</strong></td>
</tr>
<tr>
<td></td>
<td>( f ) filling factor</td>
<td>RF coil fixed onto the sample cell</td>
</tr>
<tr>
<td></td>
<td>( b ) receiver bandwidth</td>
<td>decrease</td>
</tr>
<tr>
<td></td>
<td>( Q ) factor of the RF coil</td>
<td>increase</td>
</tr>
<tr>
<td></td>
<td>( n ) figure of the amplifier</td>
<td>decreased by cooled probe</td>
</tr>
<tr>
<td></td>
<td>( V_s ) volume of the sample</td>
<td>a large coil volume</td>
</tr>
<tr>
<td><strong>Adjustable laboratory parameters</strong></td>
<td><strong>Parameter</strong></td>
<td><strong>Possibility of S/N enhancement</strong></td>
</tr>
<tr>
<td></td>
<td>( N ) number of detectable nuclei</td>
<td>increased sample concentration</td>
</tr>
<tr>
<td></td>
<td>( T ) temperature</td>
<td>decrease</td>
</tr>
<tr>
<td><strong>Instrument parameter</strong></td>
<td><strong>Parameter</strong></td>
<td><strong>Possibility of S/N enhancement</strong></td>
</tr>
<tr>
<td></td>
<td>( B_0 ) magnetic field strength</td>
<td>increase</td>
</tr>
<tr>
<td></td>
<td>( \gamma ) gyromagnetic ratio</td>
<td>a nucleus with a high ( \gamma ), e.g. (^1)H or (^19)F</td>
</tr>
<tr>
<td></td>
<td>( I ) spin quantum number</td>
<td>a nucleus with a high ( I )</td>
</tr>
<tr>
<td></td>
<td>( T_1 ) spin-lattice relaxation time</td>
<td>decrease by e.g. temperature or relaxation reagents</td>
</tr>
<tr>
<td></td>
<td>( T_2^* ) experimental spin-spin relaxation time</td>
<td>increase by shimming ( \Rightarrow ) narrower peaks</td>
</tr>
<tr>
<td></td>
<td>( n ) number of scans</td>
<td>increase</td>
</tr>
</tbody>
</table>

An adjustable instrumental parameter is the sample temperature (\( T \)). Lowering the temperature will give a larger polarisation (larger \( M_0 \)) according to the Boltzmann equation (2), giving increased sensitivity. However, lowering the temperature may reduce \( T_2^* \) and lead to a loss of S/N due to larger line widths.

The volume of the sample (\( V_s \)) and the number of detectable nuclei (\( N \)) are factors that may be limited by the available amount of sample and solubility, but should be as high as possible. Too high a concentration, however, may risk incomplete dissolution or cause
ADC overflow, which results in signal truncation or downscaling and hence a decrease in the ADC dynamic range. A sample with a very high concentration can also give line broadening due to intermolecular interactions. The number of nuclei detected also depends on the number of structurally equivalent nuclei that give rise to one signal from the studied molecule.

The gyromagnetic ratio ($\gamma$) and the spin quantum number ($I$) are given parameters for the chosen nucleus. If possible, the nuclei of protons are the ones to measure since $^1$H is a frequently occurring isotope (99.99% natural abundance) and has a high gyromagnetic ratio (26.75 compared to 6.73 for $^{13}$C). All in all, $^1$H-NMR is some 5,700 times more sensitive than $^{13}$C NMR spectrometry.9

Furthermore, the splitting of the signal, depending on the structural and chemical surroundings of the proton, affects S/N. To give an example, a singlet peak has a much higher S/N than a doublet peak, although they come from the same number of nuclei and thus have the same total area. If possible, this should be considered when choosing one peak for quantification.

As discussed earlier, various time parameters will also affect the S/N ratio in the resulting spectra: the number of transients (pulse experiments) in signal averaging (equation (13)), acquisition time (determined by $T_2^*$) and pulse repetition time (determined by $T_1$). These are also included in Table 1.

For the best S/N ratio, there is an optimum combination of the tip angle ($\beta$), the $T_1$ relaxation time and the time between pulses, given by Ernst’s equation:61

$$\cos(\beta) = e^{-\pi/\tau_1}$$  \hspace{1cm} (20)

where $\tau$ is the pulse repetition time. This equation is valid for one nucleus with the spin-lattice relaxation $T_1$ and does not give the conditions for the best accuracy of the peak integral that are required for the absolute method (i).

The simplest, and most frequently mentioned, way to ensure that the spin system is in equilibrium between pulses is to wait for 5 times the longest $T_1$ after a 90° pulse before repulsing. This corresponds to a relaxation of 99.3% and will give a maximum 0.7% error in the integral accuracy. According to Trafficante,62 a pulse angle of
83° and a pulse repetition period of 4.5 times the longest $T_1$ give
the optimum recovery after a pulse regarding accurate integrations.
The settings of these parameters for quantification are discussed in
several references.\textsuperscript{4,5,9,61,65,64}

\section*{Accuracy and precision of qNMR}

The signal peaks in an NMR spectrum are generally characterised
by four attributes: chemical shift, multiplicity, line width and relative
intensity. A well-resolved, narrow, single peak with a high S/N ratio
has the best chance of giving an accurate and precise quantification.
Several studies have discussed the uncertainties and errors in
qNMR,\textsuperscript{2-9} and accuracy and precision better than 1\% have been
reported.\textsuperscript{2,3,7,25} The main contributions to the experimental error
have been localised in these studies to the sample preparation and
the analyst.\textsuperscript{2,7}

\section*{Accuracy}

The systematic errors presented will mainly affect the accuracy of the
absolute method (i) of qNMR, where no calibration is performed.

If the \textit{acquisition time} is shorter than the time it takes for all nuclei
to relax completely, the molar response will differ from unity. To
avoid systematic errors in the peak areas, the \textit{relaxation delay} must
also be long enough to allow complete relaxation in z-led ($T_1$) for all
nuclei. Also vertical truncation of the FID, due to \textit{ADC overflow}, will
introduce distortions of spectra.

Resonances ($\omega_i$) with different distances from the excitation pulse
($B_1$) may be differently excited and detected due to \textit{off resonance effects}
and non-uniform responses to the RF pulse. If the bandpass filter
width prior to the ADC is too narrow, the lines at the ends of the
spectrum may be reduced in intensity. In this respect it is beneficial
to use peaks close to each other for quantification.

The \textit{integration range} for Lorenzian NMR peaks is limited, especially in
a crowded spectrum. An integration error smaller than 1\% requires
integration limits of $\pm 24$ times the peak width.\textsuperscript{25} However, if the
integration limits are consistent between peaks, errors or less than
0.5\% are easily achieved.\textsuperscript{3,7,40}
Precision

Random errors will affect both the precision of the absolute (i) and the calibration method (ii) for qNMR.

Point-to-point noise due to spectral resolution is a random error source. If the spectral resolution exceeds 0.4 of the peak width, the maximum error of the integral will be 0.1%. The spectral resolution may be enhanced by zero filling. Furthermore, the intensity resolution will be limited when a small peak is detected in the presence of a strong peak, depending on the dynamic range of the ADC, which will give random errors.

Automatically phased peaks are rarely acceptable with available methods, so manual phasing with a vertical expansion of the peaks is still the recommended method for the best results. Since perfect phasing of NMR spectra then relies on a subjective judgment of the goodness of the shape of signals, this will be a source of error. Baseline and phase anomalies are thoroughly discussed in the literature.

Discussion

Very few articles describe the calibration method (ii) in the context of qNMR. By using this method, the parameter determination ensuring complete relaxation is avoided. This is a great advantage over the absolute method. The drawbacks are that the calibration samples have to be prepared and that all the sample components are required as standards. The advantage of the absolute method (i) is that it is very straightforward and that no calibration is needed.
4 qNMR applications

This chapter will describe qNMR applications for impurity determination and metabolic profiling. The sample matrices, NMR instrumental parameters and applied data processing from the papers will be briefly illustrated.

Impurity determination

The presence of unwanted chemicals, e.g. residuals from synthesis and storage, even in small amounts may affect the efficiency and safety of pharmaceutical products. This is why the different pharmacopoeias prescribe limits for the allowable levels of impurities present in the active pharmaceutical ingredients and formulations. Impurity profiling – the identification and quantification of impurities – is of great importance in the pharmaceutical industry.65

qNMR is a very useful technique for impurity determinations due to the absolute response factor, selectivity and the reproducibility of chemical shifts. Impurities with similar chemical structures may readily be both identified and quantified by NMR spectrometry, making it a very cost-effective alternative compared to other analytical methods that require the acquisition of separate standards of the target analyte and possible impurities.

The use of 1H-NMR spectrometry for impurity (or low-level) determination has been reported in several papers1,7,45,48,52,55,66-72 and in papers I–II in this thesis. Organic components have been reliably quantified below the 0.1% level7,70 and some papers have reported LODs lower than 100 µg/g.49,52,67,70,papers I–II The main aim in papers I–II was to investigate the potential of the qNMR technique for impurity determination in these specific cases. There were no limits of materials, i.e. no problem of making calibration samples, and no attempt was made to perform an absolute determination.
Paracetamol samples (paper I)

Paracetamol (N-(4-hydroxyphenyl)-acetamide) is a drug having an analgesic, antipyretic and antiphlogistic action that is widely employed in therapeutics. The main impurity in paracetamol preparations is 4-aminophenol, an intermediate in the synthesis of paracetamol (Figure 9), which may also be formed during storage. The 4-aminophenol content in paracetamol is limited to 50 µg/g by the European, United States, British and German pharmacopoeias.73

![Figure 9: Structures of paracetamol (left) and 4-aminophenol (right).](image)

The Roman numbers is referred in the legend of Figure 10.

In paper I, the potential of determining 4-aminophenol in paracetamol by qNMR was explored. Mixtures with different concentrations of the impurity were dissolved in DMSO-d<sub>6</sub>. The pulse angle and repetition time chosen were 83° and 4.5*T<sub>1</sub> respectively, the optimal settings for best peak integral accuracy according to Trafficante.62 To expand the vertical dynamic range, the oversampling factor was set to 20. The study is fully described in paper I and part of a spectrum is shown in Figure 10.
Figure 10: A part of an NMR spectrum of paracetamol holding 1 mg/g of the impurity 4-aminophenol. The peak marked p+4ap comprises a $^{13}$C-satellite signal (0.55% of the main peak intensity) from two aromatic protons in paracetamol (marked I in Figure 9) and the signal from the corresponding signals in 4-aminophenol (II). The peak marked 4ap arises from 4-aminophenol (III).

**Poloxamer samples (paper II)**

Poloxamer materials are synthetic copolymers of ethylene oxide and propylene oxide (Figure 11) and were first synthesised in 1954 by Lundsted and Ile. Poloxamers form a thermoreversible gel, a low-viscous liquid at low temperatures and a gel at body temperature, which has been used for pharmaceutical formulations.

Acetaldehyde and propionaldehyde (Figure 11) are well-known impurities in poloxamers. They are formed from the monomers ethylene oxide and propylene oxide, which are the building blocks of the poloxamers. The maximum allowed amounts of acetaldehyde and propionaldehyde in the poloxamer are 80 and 100 µg/g, respectively, for some medical applications.
The LC method for the determination of acetaldehyde and propionaldehyde in poloxamer in use today involves a very time-consuming work-up procedure. By means of qNMR the poloxamer samples can be analysed only by dissolving the samples in acidic water. Due to the volatile aldehydes and because the poloxamer forms a gel at temperatures above room temperature, the poloxamer samples were analysed at 275 K. The viscosity of the samples was also the reason for running the experiments in a non-spinning mode. The oversampling factor was set to 16 to expand the vertical dynamic range (chapter 2, “Digital filtering”). The remainder of the parameter setup varied according to Table 2. An NMR spectrum is shown in Figure 12.

Table 2: NMR instrumental parameter setups that were tried for determination of acetaldehyde and propionaldehyde in poloxamer.

<table>
<thead>
<tr>
<th>Shimming</th>
<th>Gradient</th>
<th>Gradient</th>
<th>Manual</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulse angle</td>
<td>27°</td>
<td>27°</td>
<td>78°</td>
<td>78°</td>
</tr>
<tr>
<td>Recycle time (s)</td>
<td>6 (3.3·T_{1aa})</td>
<td>4 (2.1·T_{1aa})</td>
<td>9 (5·T_{1aa})</td>
<td>14 (7.8·T_{1aa})</td>
</tr>
<tr>
<td>Acquisition time (s)</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Number of scans</td>
<td>64</td>
<td>256 / 64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Tot. exp. time (min)</td>
<td>6.4</td>
<td>17 / 4.3</td>
<td>9.6</td>
<td>14.9</td>
</tr>
</tbody>
</table>

$T_{1aa}$: Spin-lattice relaxation time constant for acetaldehyde (1.8 s)
Figure 12: A spectral segment including signals from the CH$_3$ groups from the substances poloxamer at 1.2 ppm, acetaldehyde (aa) at 2.2 ppm (176 μg/g poloxamer) and propionaldehyde (pa) at 0.89 ppm (175 μg/g poloxamer).

**Zero filling and line broadening (papers I–II)**

Zero filling and line broadening were performed to a varied extent previous to a Fourier transform in both paper I (Table 3) and paper II. The evaluation showed no improvement in the quantitative results due to zero filling, although line broadening improved the quantifications to some extent in paper II.

<table>
<thead>
<tr>
<th>Zero filling†</th>
<th>Wavelet compression</th>
<th>Line broadening</th>
<th>RMSEP (μg 4ap/g pa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>no</td>
<td>no</td>
<td>35 24 15 27 25</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>21 59 32 38 37</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>62 59 30 46 49</td>
</tr>
</tbody>
</table>

† to double the number of data points
4ap: 4-aminophenol
pa: paracetamol

Table 3: Root-mean-square error of prediction (RMSEP) (equation (31)) of different test sets from NN calibration models including partly different data sets (A-D), and with different data preprocessing (see paper I for details).
Spectral smoothing (paper II)

A one-dimensional convolution of spectra with a triangular function (of height 1 and width 41) was performed in paper II. The evaluation of this coarse smoothing of spectra improved both repeatability, reproducibility and the LOD.

Metabolic analyses

NMR spectrometry is very attractive for quantitative and qualitative determinations of small molecules in biological fluids since it requires no or very little sample preparation. Drugs, toxic substances and metabolites may be analysed without prior assumptions, e.g. to measure the biological response to some xenobiotics. Already in the early eighties NMR spectrometry was used for the analysis of serum, plasma and urine.77-79 This field of research in combination with high resolution 1H-NMR was reviewed by Nicholson and Wilson in 1989.42

Although the terminology in this area of research is still evolving, metabolic analyses may be divided into four general areas.80-83

• Metabolite target analysis: The quantification of specific metabolites.

• Metabolite/metabolic profiling: Quantitative or qualitative determination of a group of related compounds or of specific metabolic pathways.

• Metabolomics/metabonomics: Comprehensive analysis of all metabolites at a specified time and given environmental conditions.

• Metabolic fingerprinting: Sample classification by global analysis of all present components.

In drug research, the term metabolic profiling is often used to describe the metabolic fate of an administrated drug.83 Metabolic profiling makes it possible to generate information about in vivo multiorgan functions. This is an effective preclinical screening method. This type of analysis may also provide one key to “systems biology”: the combination of genomic, proteomic, and metabolomic data.80,81,84
By means of urine sampling, information can be semi-continuously collected from the same animal without euthanasia. The $^1$H-NMR analysis of urine samples takes a few minutes and the sample preparation includes only buffering and centrifugation. Urine sample analysis is expected to reflect a large number of biochemical processes in the body, making it a potential diagnostic tool.\textsuperscript{81,84}

The improved capabilities of the NMR instruments due to the development of technology makes it possible to detect thousands of signals arising from hundreds of endogenous molecules in spectra from $^1$H-NMR analysis of biological fluids such as urine or plasma. Although these signals often overlap, this is a useful tool for metabolic fingerprinting. When analysing these complex samples, the assignment of all peaks is seldom accomplished. However, even unidentified signals can be used as fingerprints. Techniques for multivariate data analysis and pattern recognition then help to give qualitative and quantitative interpretation out of complex spectra.\textsuperscript{46,47,54,56,72,80,82,84-97}

A lot of biological variations will affect the quality of urine sample $^1$H-NMR spectra. For example, the ionic strength varies considerably in urine samples and may affect tuning and matching (chapter 2, “NMR data acquisition”).\textsuperscript{81} Water, which is present in a very high concentration,\textsuperscript{42} has to be reduced by solvent suppression during the NMR experiment. The concentration of the urine must also be compensated for in the analyses.

**Rat urine samples (papers III–VI)**

The rat urine samples studied in papers III–VI were received from an animal study at AstraZeneca. The study included two groups of six rats each: one control group orally dosed with water and one group dosed with citalopram (positive control for phospholipidosis). The rats were dosed once a day for 14 consecutive days and urine samples were collected on days -5 (pre-dose, two occasions), 1, 3, 7, 10 and 14. The study is fully described in paper III.

The samples were buffered and centrifuged prior to $^1$H-NMR analysis on a Bruker DRX600 instrument working at 600.13 MHz. Each sample was automatically shimmed by gradient shimming and suppression of the water signal was achieved using presaturation during relaxation delay and mixing time with a shaped pulse for
selective saturation. The acquisition time was 4.68 s and the total pulse recycle time was 7.71 s, 64 scans per sample using a spectral width of 7002.8 Hz. These parameters were chosen on the basis of previous work analysing urine samples with NMR spectrometry.\textsuperscript{92,95} The total NMR acquisition time for one sample was 8.6 min. An example of an NMR spectrum is shown in Figure 13.

![Figure 13: 1H-NMR spectrum of urine from a rat dosed with citalopram for 7 days.](image)

Phase correction and peak alignment (papers I–VI)

Spectra were manually phased and referred to the TMS peak on the NMR instrument and further processed in MATLAB.\textsuperscript{98} In paper I, the spectral areas of interest (6.4-6.55 ppm) were cut out, automatically phased and corrected laterally by means of a genetic algorithm until the peaks were aligned. In papers II–VI each spectrum was manually phased (again) in MATLAB and peak-aligned segmentwise as described in papers III and IV. The subject of peak alignment is thoroughly discussed in chapter 6.
Data from a chemical analysis consist of information and noise. To get high accuracy and precision for a chemical analysis, we need knowledge about both of these components. The properties of the instrumental noise of NMR spectrometry are discussed in chapter 3 above. The noise also has chemical and/or biological contributions.

The desired and relevant information is often just one part of the total information. When it comes to data analysis, this is what we want to extract. This can be performed by chemometrics, which comprises the application of multivariate statistics, mathematics and computational methods and also includes how to represent and display this information. A philosophical article about the subject has been written by Svante Wold. More theoretical references to the subject of chemometrics can be found in books.

**Principal component analysis**

Principal component analysis (PCA) is a well-established bilinear projection method for interpreting multivariate data. Pearson is considered to be the first person to have described PCA as we know it today, although Cauchy derived it as early as 1829. PCA denotes the extraction of eigenvectors from the variance-covariance (var-cov) matrix to obtain a number of uncorrelated (orthogonal) new variables called latent variables or principal components (PCs).

\[
\text{cov}(\mathbf{X}) = \begin{bmatrix}
\text{var}(x_1) & \ldots & \text{cov}(x_i, x_p) \\
\vdots & \ddots & \vdots \\
\text{cov}(x_p, x_1) & \ldots & \text{var}(x_p)
\end{bmatrix}
\]  

(21)

for variable \(i, j = 1 \ldots p\) in the data matrix \(\mathbf{X}\) of the dimensions \(n \times p\).

\[†\text{From the German } eigen \text{ meaning “inherent, characteristic”}\]
PCA gives a new representation of data in a new, rotated coordinate system. Every PC is optimised in regard to its ability to explain the direction of *maximum variance* in the original data and each PC explains variation uncorrelated with the other PCs. The data are approximated by the model:

\[ X = TP' + E, \]

where \( X \) is the original \( n \times p \) set of data, typically \( n \) samples and \( p \) variables. \( P' \) is an \( m \times p \) matrix of variable coefficients (loadings) that give the relative contributions of every variable in calculating the object scores, \( T \), which is an \( n \times m \) matrix, of the coordinate values on the PCs. \( E \) is an \( n \times p \) matrix containing the residuals not explained by the model using \( m \) PCs. The score value representation (scores plot) will show a clustering of similar samples, as sketched in Figure 14. The loadings represent the variable direction, describing the influence of the variables on the scores, and are a valuable interpretation tool.

![Figure 14: A geometrical interpretation of a PCA model. The original data matrix \( X \) consist of \( p \) variables and eight samples. The first three variables (v) are represented in the coordinate system to the left, where the first two principal component vectors, PC1 and PC2, are also shown.](image)

PCA as a projection technique in combination with NMR spectra is frequently employed in systems biology and in the pharmaceutical industry. Examples are found in papers III–VI and several references.\(^{82,86,90-97,106}\)
Parallel factor analysis

Parallel factor analysis (PARAFAC) is a multiway decomposition method, a generalisation of PCA to higher order arrays.\textsuperscript{107,108} $\mathbf{X}$ is decomposed in the loading matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, illustrated by:

$$\mathbf{X}_{npq} = \mathbf{A}_{nm} \mathbf{B}_{pm} \mathbf{C}_{qm} + \mathbf{E}_{npq}$$

where $n$, $p$ and $q$ represent the dimensions of the data matrices, $m$ denotes the number of PARAFAC components and $\mathbf{E}$ is the residual matrix in this three dimensional example.

PARAFAC was applied to three- and four-dimensional data in paper VI and has the advantage of being able to retain the time-related structure of this data. Compared to bilinear models with unfolded data, multiway models are generally more robust and easier to interpret, provided that all dimensions hold information.\textsuperscript{107} Theoretically, the PARAFAC model has a unique solution compared to bilinear methods, which solutions suffer from rotational freedom. However, scaling and centring are not as straightforward in the multiway case.

Calibration

Since qNMR has the property of being a primary method of measurement, no calibration is needed as long as we have only random noise. However, if we are uncertain about the randomness in the noise or we do not wait for full relaxation, the equal molar ratio between peaks will not be valid and a calibration is required (as discussed above in chapter 3).

The simplest form of calibration is to measure one sample for which we have all the information, typically concentration. From this sample we get a relationship (b) between the signal ($\mathbf{x}$) and the information ($\mathbf{y}$), from which an unknown sample ($\mathbf{y}_i$) could be predicted when we have new measured data ($\mathbf{x}_i$):

$$\mathbf{y}_i = b \cdot \mathbf{x}_i \quad (22)$$

This is suitable when we have samples to determine in a known, rather limited, concentration range.
To cover a larger concentration interval, several samples \((x)\) are required in a calibration model:

\[
y = b \cdot x + e
\]  
(23)

e will include both measurement and model error, and \(b\) is estimated by least-squares regression (24) and is called the regression vector.

\[
\hat{b} (LS) = (x'x)^{-1}(x'y),
\]  
(24)

where \(x = (x_1, x_2, \ldots, x_i, \ldots, x_n)'\) and \(y = (y_1, y_2, \ldots, y_i, \ldots, y_n)'\).

qNMR calibration modelling including area determinations, i.e. univariate calibration, is exemplified in paper II and references.43,44,53

NMR peak area determinations are easily performed with consistency if the peaks are singlets and have no other peaks in their immediate surroundings. However, if the peaks suffer from interferences or if singlets have to be compared with multiplets, integration is a more difficult task. A very complex spectrum, such as the NMR spectra of biofluids, may also lead to difficulties in sorting out the important signals. To exploit the full information content of NMR data, including hundreds of overlapping signals, multivariate data analysis methods are preferred.

The use of methods for multivariate calibration methods have increased with the desire to solve increasingly difficult problems and complex data. Naturally this goes hand-in-hand with the increased computer performance. A simple method for multivariate regression is multiple linear regression (MLR), with the estimate of \(b\), according to:

\[
\hat{b} (MLR) = (XX)^{-1}(X'y)
\]  
(25)

where \(X = (x_1, x_2, \ldots, x_i, \ldots, x_n)\), in accordance with equation (24) above.

MLR may also be performed on PCA scores \((T)\) on \(y\), which is called principal component regression (PCR). \(b\) is then estimated as:

\[
\hat{b} (PCR) = (T'T)^{-1}(T'y)
\]  
(26)

The advantage of PCR over MLR is that it involves no limit in the number of variables, since in MLR the number of variables must be less than the number of samples.
Partial least squares

When the maximum variance in $X$ does not correlate with the maximum variance in $y$, PCR will give an unnecessarily complex and unstable model for calibration. In such a case, it may be more appropriate to use partial least squares or projection to latent structures (PLS). The objective function of PLS is to find maximum covariance between $y$ and $X$, in contrast to PCR, where the maximum variance in $X$ is found.

PLS was first invented by Herman Wold in the 1960s\textsuperscript{109,110} and was later modified and extended by Svante Wold et al. in 1983,\textsuperscript{111} when it also was named partial least squares. This method for regression has been established by its success in a wide variety of problems.\textsuperscript{112-114} The details of PLS regression have been thoroughly described in the literature, and only a short summary follows here.

A vector of weights, $w$, describing the maximum covariance between $X$ and $y$ is defined according to:

$$\hat{w} = \frac{X'y}{|X'y|}$$\hspace{1cm}(27)

Scores values $t$ are obtained when $X$ is projected onto $w$

$$t = Xw,$$\hspace{1cm}(28)

and the regression coefficient vector is then determined by

$$\hat{b} = t'y/(t't)$$\hspace{1cm}(29)

After estimation of the x-loadings vector

$$p = X't/(t't),$$\hspace{1cm}(30)

the $tp'$ matrix is subtracted from $X$ and the $bt$ matrix is subtracted from $y$, and the procedure is repeated from (27) to extract additional PLS components.

PLS applied to NMR data

PLS has previously been applied to NMR data\textsuperscript{46,47,54,55,85,87} and in the course of the work on all the papers in this thesis, both for creating calibration models and for explorative analysis. In paper I, PLS did not perform as well as a calibration model as the applied neural network, and was therefore not described. In paper II, the univariate analysis based on integration of peaks performed as well
as the PLS models. In order to chose as simple a model as possible, the PLS models were rejected. In paper III, PLS was rejected in favour of the unsupervised PCA model to avoid the risk of guiding the latent variables in the direction of a false $y$ variable. Day of sampling was tried as $y$, although the kinetics of the drug-induced metabolic changes were unknown. Papers IV–VI include PLS, both with $y$ as a variable of time and a measured $y$ variable (LC/MS), and as a discriminant analysis (DA) method.\textsuperscript{112} There is, however, a risk, especially with the PLS-DA models, in such analyses. The PLS model may appear to perform really well if only one or a few, e.g. misaligned peaks or noise, discriminate between the samples, though the actual model performance is poor. This risk must be minimised by studying loadings and by a proper validation.

N-PLS was developed by Bro 1996\textsuperscript{115} as an extension of PLS to higher order data. It was used in paper VI, making it possible to utilise the structure of original data and avoid unfolding.

**Nonlinear regression**

Linear multivariate regression methods like PCR and PLS may to some extent explain nonlinearities if a sufficient number of latent variables are selected or the $x$ and/or $y$ block are transformed or expanded. However, a better alternative may be to use a nonlinear model. There are examples of nonlinear PLS models where the inner relation between $x$ and $y$ scores is defined as nonlinear (a quadratic polynomial).\textsuperscript{116} A more common way of modelling nonlinear data is probably neural networks (NN), which was employed in paper I, where a nonlinear relationship was observed. The first application of NN to NMR data was demonstrated in 1989 by Thomsen and Meyer,\textsuperscript{117} followed by, among others, Amendolia et al.\textsuperscript{118} A short description will be given here, and further reading about NN can be found in the literature.\textsuperscript{119}

A multilayer feed-forward NN is composed of a highly interconnected mesh of nonlinear and linear computing elements. The fundamental processing element of an NN is the neuron. Neurons are connected by links, and there is a coefficient (or a weight) associated with each link. The neurons receive external inputs (first layer) or inputs from other neurons and perform a weighted sum of these inputs. The neurons process the resulting signal with a transfer function and
then produce an output passed on as an input to other neurons or as an output from the entire model. This is illustrated in Figure 15.

![Figure 15](image)

**Figure 15**: A sketch of a neural network with three input data, two neurons in the hidden layer and one output data. Each line represents a weighting and each circle is a neuron defining a transfer function.

A process known as training or learning establishes the values of the weights. Output vectors ($\hat{y}$) are calculated from a set of input vectors ($x$) for which $y$ is known a priori. The weights are modified based on the difference $y - \hat{y}$. As this process is repeated, the weights gradually converge to values that transform each input pattern to an output pattern close to its target ($\hat{y}$). Any type of variation in the inputs may be accounted for by including them in the training process.

**Data preprocessing**

A common denominator of all of the multivariate methods is the necessity of proper data preprocessing prior to data analysis. Several approaches are available that deal with undesired spectral artefacts, e.g. multiplicative scatter correction (MSC)\(^{120}\) for baseline correction, orthogonal scatter correction (OSC)\(^{86,121}\) for removal of unwanted variances and corrections for relative intensity variations such as mean-centring ($x_j - x_{j\text{mean}}$, for variable $j$ and sample $i$), unit variance scaling ($x_j / x_{j\text{std}}$, where $x_{j\text{std}}$ is the standard deviation of $x_j$), and normalisation (which could be performed in various ways). Mean-centring is performed prior to multivariate analysis to centre the new data representation on the origin of coordinates since we are generally not interested in the mean level of the variables, but the variation around the mean. Unit variance scaling is generally not suitable for spectral data since noise is scaled to be as important as peaks, which may be risky, especially when PLS is applied due to the risk of overfitting.
Normalisation is actually worth a chapter of its own, but will only briefly be mentioned here. It must be performed with care since normalising to the wrong standard could introduce erroneous trends into data. In papers III–VI, spectra are normalised to equal area in order to deal with concentration differences in the urine samples. Also, normalisation with reference to one peak as an internal standard (e.g. the TMSD\textsuperscript{55} or creatinine peak) appears in the literature.

Furthermore, peak alignment is an important preprocessing tool in multivariate analysis of NMR data. It is employed in all the papers in this thesis and the subject is discussed in detail in chapter 6.

Validation

One may make an explorative model to search for trends or groupings in data and use no validation at all, as is done in papers III, IV and VI in this thesis. This is, however, a somewhat “quick and dirty” approach. A proper validation of a model (explorative or for prediction) is preferred. The purpose of the validation is to derive estimates of the accuracy of the validity or predictions from the model.

An appropriate method of validation is, prior to any data analysis, to choose a (representative of the population) test set to be put aside. The remaining samples, the training set, are then used to determine the model parameters, e.g. the number of components in the model, variable selection, number of learning cycles in a NN training, as well as preprocessing and scaling of data. The test set is then predicted or projected onto the optimised model to estimate the generalisation error. This procedure was used in papers I, II and V.

Model selection

The optimisation of a model could be guided by cross validation.\textsuperscript{122} The data set (here the training set) is divided into a number of subsets. Each subset is predicted onto a model built from the other subsets. When all the subsets have been predicted, the generalisation error of the model can be estimated. The subsets must be picked with care, duplicates must be in the same subset, and the subset must not represent all samples from one single corner of the sample domain. An example of this type of optimisation is shown in paper V.
If the number of subsets equals the number of samples, this is called leave-one-out cross validation. This could give a good estimate of the error of prediction if the samples are too few to comprise a test set and the samples are well spread in the domain, as in paper II. In paper I, an “internal test set”, here called a validation set, is used for the optimisation of the neural net parameters. Additional validation methods for optimisation can be found in the literature.\textsuperscript{102}

**Objective function**

We may have different performance objectives for the model depending on the situation. Most often, the error of prediction is to be minimised. Root-mean-square error of prediction (RMSEP), equation (31), is a suitable and common measure (papers I–II).

$$\text{RMSEP} = \sqrt{\frac{1}{J} \sum_{i=1}^{J} (\hat{y}_i - y_i)^2},$$  \hspace{1cm} (31)

where $\hat{y}$ is the predicted concentration, $y$ is the measured concentration and $J$ represents the number of predictions.

Another measure of quality may be a measure of class separation.\textsuperscript{123} In papers IV–VI we have used a measure developed by M. Åberg,\textsuperscript{124} illustrated in Figure 16. In the scores vectors space, each class was approximated by a bivariate normal probability distribution. The boundary between a pair of classes was defined as the hypersurface on which the probability densities of the two classes were equal. The measure of class separation was calculated as the minimum Mahalanobis distance\textsuperscript{125} between the class boundary and the class centres.

![Figure 16: A geometrical interpretation of the measure of class separation where x and o represent two classes, the line between them represents equal probability density and the arrow shows the measure of class separation.](image-url)
6 Peak alignment

NMR peak area and position give detailed chemical and quantitative information. Peak shifts and differences in peak shapes may, however, also arise from variations in experimental conditions such as temperature variations and inhomogeneities in the applied magnetic field or differences in the background matrix of the sample.\textsuperscript{81,87} The pH of the sample is a major source of variation in peak positions,\textsuperscript{126} and even if the samples are buffered, small pH differences will be detected.\textsuperscript{42,92,95,127} In biofluid samples the background matrix also reflects an individual variation depending on the metabolism.\textsuperscript{82,128} These factors are, however, most often not what we want to measure.

From an ideal multivariate analysis point of view, any NMR peak originating from identical nuclei should have an identical position and shape between samples. If this is not the case, a multivariate model will be unnecessarily complex and more difficult to interpret. Consequently, the usefulness of the results achieved will decrease. Examples of problems and artefacts originating from misaligned NMR signals have previously been discussed in papers III–IV and in references.\textsuperscript{87,89,126,129,130}

Peak alignment in NMR spectrometry

To overcome this NMR peak shift problem in combination with pattern recognition techniques, different approaches have previously been suggested. Cloarec et al.\textsuperscript{88} have recently presented the use of orthogonal projection to latent structures (O-PLS)\textsuperscript{131} discriminant analysis to handle peak shift problems. The predominant method in the literature is, however, integration of predetermined spectral segments, i.e. bucketing, typically 0.04 ppm wide.\textsuperscript{86,91,92,95,96} The bucketing approach deals with the peak shift problem but ruins
the spectral resolution of the acquired data, typically the spectral resolution decreasing from 6,500 data points/ppm to 28 data points/ppm. This makes multivariate detection of small variances (peaks) impossible if they are confounded in the same bucket as larger variances. Hence, the interpretation of a bucketed data model will not reveal the specific peaks responsible for the clustering, as is shown, for example, in paper III and by Cloarec et al.88

Also, more refined alignment methods (than bucketing) for NMR data have been reported. One example is partial linear fit (PLF), a method outlined by Vogels et al. 1996.47 PLF automatically picks out segments in an NMR spectrum of size $d$ and shifts them $s$ points left and right. Every possible and relevant combination of $d$ and $s$ is tried until the sum of squared differences between the spectrum and the target for alignment is minimised. Another approach, outlined by Brown and Stoyanova 1996,129 performs automatic removal of both frequency and phase shifts in NMR spectra by using PCA to determine the misalignment in a single peak appearing across a series of spectra. This method has been extended and applied to in vivo NMR spectra,132 and has been further developed and applied to high-resolution spectral data.130

As a contribution to this field of research, a method of peak alignment by means of a genetic algorithm (PAGA) was developed in Paper III. It is an automated segment picking method including both linear interpolation and shifting of segments, optimised by a genetic algorithm with the correlation coefficient as objective function. This algorithm was speeded up by Lee and Woodruff 133 by replacing the genetic algorithm by a beam search algorithm, and was further analysed and compared to peak alignment by reduced set mapping (PARS) in paper IV.

PARS was developed by Torgrip et al.134 The method relies on peak detection and converts spectra into sparse vectors consisting of zeros and peak intensities corresponding to peak maxima. This peak alignment is performed by a fast tree-search algorithm with early pruning of non-optimal alignment solutions.
Peak alignment for other analytical techniques

There are also several examples of alignment of first-order data, which have not yet been applied to NMR data. Some of them will be mentioned here.

Dynamic time warping (DTW) is an application of dynamic programming which has been used for aligning profiles from industrial batch processes of polymerisation\textsuperscript{135} and GC/FT-IR/MS chromatograms,\textsuperscript{136} although it has its origin in speech recognition.\textsuperscript{137} The signals are warped nonlinearly, point by point, to fit a reference profile and the cumulative distance (e.g. Euclidean, equation (35)) between points is optimised. Additional methods developed from DTW have been presented for chromatograms: one peak matching algorithm relying on peak detection and local warping (LW)\textsuperscript{138} and one parametric time warping method (PTW).\textsuperscript{139}

Piecewise linear correlation optimised warping (COW)\textsuperscript{140} has been applied to GC and LC data.\textsuperscript{141,142} This algorithm selects predetermined spectral segments of equal length and by linear interpolation stretches and shrinks them to fit a reference spectrum. The correlation coefficient is optimised. A method for peak alignment very similar to COW, but with sideways shifting instead of interpolation, is piecewise alignment (PWA).\textsuperscript{123} Another segmentwise peak alignment approach based on genetic algorithm optimisation has been applied to LC chromatograms on peptide maps.\textsuperscript{143}

Walczak and Wu recently presented a method for fuzzy warping (FW) of chromatograms relying on only the main peaks.\textsuperscript{144} They also compared the performance and processing time for several different methods of peak alignment referred to above: FW, COW, PAGA, PTW and DTW. PAGA (from paper III) performs really well in comparison to the others, although it is relatively time-consuming in some cases.

Profile alignment or peak alignment

Some of the methods mentioned rely on peak detection in some way (PARS, FW and LW) or include just one peak at a time. When a peak list is created, no noise is present and unit variance scaling may successfully be applied as in paper IV, for instance. This and the data reduction are some of the advantages of peak detection. The disadvantage is the uncertainty in defining the peaks.
The data set from the urine sample $^1$H-NMR analysis studied in this thesis has a lot of line-width variations due to, for example, shimming problems. An example is shown in Figure 17, which illustrates the difficulty of peak detection relying on a peak shape definition in these spectra. In order to avoid detection of peaks, methods relying on alignment of profiles are preferred, e.g. DTW, PTW, COW, PLF, PWA and PAGA.

Figure 17: Example of peak shifts and peak shape distortions due to varying background matrix and shimming between $^1$H-NMR spectra from rat urine samples. Black lines represents samples from control rats and grey lines represent samples from dosed rats (days 7-14). The upper figure shows unaligned spectra and the lower figure shows spectra aligned by PAGA.
Target for alignment

In first-order real data (one vector per sample), there are no data to guide or validate a peak alignment as in second-order data (one matrix per sample). Hence, a target must be chosen – it could be one spectrum, although various alternatives are possible.

When the same target is used for all the spectra in a data set, differences between groups of spectra will easily be detected in the same score plot from a multivariate analysis, assuming that peaks appearing from the same substance appear at the same position in all spectra. An unknown sample would be aligned to the previously chosen target and projected into the multivariate model space (loadings) for visual interpretation. This is the situation in papers III–IV and references,47,123,134,135,138-140,144,145 where one target spectrum is chosen and assumed to reflect all peaks of interest. Also a combination of targets may well be employed, such as the sum of two spectra (paper II) or a mean spectrum.138 For the PARS method,134 there is also a possibility of using a recursively updated target, where new (unaligned) peaks appearing in aligned spectra are added to the target (paper IV), or a dendogram alignment scheme where a root target is created.146

When interpreting data with defined class associations, this information could be used. All the spectra in one class may be aligned to one chosen target in its own class.143 The interpretation with all classes in one multivariate model will not be valid in this case. There are also examples where the whole data matrix (all samples in the study) is part of the alignment and no special target is chosen.129,130,132

Optimisation

An optimisation problem is a computational problem in which the objective is to find the best of all possible solutions, or more formally, find a solution in the feasible region which has the minimum or maximum value of the objective function. There are several solution methods to optimisation problems. I will bring up and compare some used in the methods for alignment mentioned above.
Dynamic programming is an algorithmic technique which decomposes a multivariate optimisation by evaluating all possible solutions. This is a “safe” method but could be very computer-intensive if the variables are numerous, as when comparing two NMR spectra with 65,000 variables each. Dynamic programming is utilised in COW, DTW or LW, for example.

Genetic algorithms (GAs), used for PAGA (paper III), search the solution space of a function by simulated evolution. A GA usually starts with a random population of candidate solutions. Each solution is subjected to a genetic search, which encompasses assignment to a fitness value according to an objective function, selection of strings for further replication, and recombination by crossover and/or mutation. This genetic search runs until a predefined optimisation criterion is met.

Beam search, which was used for alignment by Lee and Woodruff, is a heuristic search algorithm that includes an optimisation of best-first search, where a number of nearly optimal alternatives (“the beam”) are examined. Initially, a set of likely solutions is created on the edge of a predetermined search radius and subsequently evaluated. The algorithm then assigns one or more new candidate solutions by selecting the best steps from the current trial solution(s) where is the parametric input beam width. For each loop of the algorithm, the radius is reduced until a stop criterion is met. In paper IV and the paper by Lee and Woodruff, peak alignments by a genetic algorithm and by beam search are compared. According to these studies, they perform equally well, but beam search is about ten times faster.

In optimisation by tree-search algorithms, nodes (e.g. possible peak matches) are defined from the data structure. The nodes are then searched level by level until the globally optimal solution is found (breadth-first search as in PARS).
Objective function

Several objective functions are mentioned together with peak alignment techniques. In paper I, the fitness criterion was the sum of the absolute differences between the actual spectrum ($S$) and a target spectrum ($T$):

$$
\sum_{i=1}^{J} |s_i - t_i|
$$

(32)

where $s_i$ and $t_i$ represents each data point $i$ in spectrum and target respectively. The correlation coefficient ($CC$) was used as a measure for the goodness of alignment in, for example, COW, PWA, LW and PAGA (paper III). It is a numeric measure of the strength of linear relationship between variables and it is calculated by dividing the covariance (cov) by the product of their standard deviations ($\sigma$):

$$
CC(S, T) = \frac{\text{cov}(S, T)}{\sigma_S \cdot \sigma_T}.
$$

(33)

The standard deviation ($\sigma$) is estimated by the square root of the variance, the root mean square (RMS) of the deviations from the average:

$$
\sigma = \sqrt{\frac{1}{J-1} \sum_{i=1}^{J} (x_i - \bar{x})^2},
$$

(34)

where $J$ is the number of observations. In COW, PWA and LW, a Wallis filter was also used to minimise the effect of varying peak heights. This was not used, however, in paper III since we wanted the larger peaks to guide the alignment.

The root-mean-square error (RMSE) used as objective function in PLF and PTW, for example, is actually a measure of the mean Euclidean distances between profiles. The Euclidean distance is defined as:

$$
D(S, T) = \sqrt{\sum_{i=1}^{J} (s_i - t_i)^2}
$$

(35)

for $J$ variables.
NMR data compression

Generally, no data compression of a 1H-NMR spectrum is required to perform the data analysis. Since today’s powerful computers can handle (quite) large data volumes, the whole data matrix should not be a problem to cope with. However, special cases when, for example, neural networks are utilised for calibration, the data need to be reduced due to heavy computations, as shown in paper I. This reduction can be performed by more sophisticated methods than bucketing (mentioned above). Examples in the already referred literature are PARS where a sparse matrix replace the spectra by peak identification, or wavelet reduction and variable selection (paper I). In paper IV, data compressions by bucketing to different extents after peak alignment are compared. The conclusion was that no difference in class separation from a PLS-DA model is shown when a reduction from about 65,000 to 8,000 data points of spectra is performed.

Discussion

The optimum choice of method for peak alignment will differ between applications. However, most of the methods mentioned above could be applied to any data produced by an analytical chemist. On the other hand, it is not straightforward to put the optimum parameters into the algorithms. Either the alignment methods rely on peak detection, which implies some definition of a peak and may cause problems, e.g. if the peak shapes differ between samples, or the alignment requires some segment picking, in which optimum parameters will also differ between applications. One conclusion may be that the peak picking approach is favourable in samples with well-defined peaks, appearing similar between samples, and a profile alignment including segment picking is preferred when peaks or profiles are not that well defined. Point-by-point alignments such as DTW are, of course, also possible. However, for NMR data, consisting of around 60K data points, the available methods are very computer-intense.

Another parameter to choose in alignment tasks is the objective function, which should be chosen to fit the samples, either the largest intensities being allowed to guide the alignment (CC (paper III) or RMSE) or each peak being equally evaluated (Wallis filter).
As described in chapter 4, $^1$H-NMR spectrometry is a well-established technique for generating metabolic profiles. Recently, liquid chromatography/mass spectrometry (LC/MS) has also been used for the same purpose. Both techniques have their advantages in this field of research. For instance, $^1$H-NMR spectrometry requires minor sample pretreatment, while LC/MS in general possesses a higher sensitivity. $^1$H-NMR spectrometry may detect compounds that are not retained on the LC column or not ionisable in the mass spectrometer. LC/MS may detect compounds that are present in a concentration below the LOD for $^1$H-NMR spectrometry.

The rat urine samples described in chapter 4 have been analysed both by $^1$H-NMR (paper III) and by LC/MS. When comparing the data from the two techniques, it is most likely that they generate complementary information. A few previous metabolic studies have analysed both LC/MS and NMR spectrometry for analysing the same samples to verify and complement results of the other technique. The data sets have then been studied separately. In papers V and VI, we wanted to analyse the two data sets within the same chemometric model. This was carried out by data fusion and data correlation.

Original data

These data analyses started with the data schematically represented in Figure 18. The LC/MS data consisted of a 95 variable peak list for each sample, pretreated by curve resolution, peak alignment, log$_{10}$ transformation and mean centring. The $^1$H-NMR data were pretreated by peak alignment according to PAGA (but with the beam search optimisation algorithm), data reduction by segmentwise integration, normalisation to equal area and mean centring, resulting
in 8,192 variables for the data fusion and 2,048 variables for the data
correlation. The data pretreatments are further described in papers V and VI.

Figure 18: Original data.

Data fusion

In the field of chemometrics the number of possible operations
and combinations of method applications to data are numerous.
In paper V, the aim was to examine different possibilities of
performing data fusion of the measured LC/MS and $^1$H-NMR data.
Data fusion is defined as a method that combines information from
different sources to produce a single model or decision.152 There
are a few examples of data fusion in the literature153,154 showing that
this is not straightforward. Our hypothesis for this work was that
the classification of the samples would be improved by using data
fusion, making the diagnosis (i.e. pattern recognition) of the studied
phenomena more reliable.

Concatenation, full hierarchical modelling, batch modelling and
outer product analysis were the techniques examined for the fusion
of LC/MS and $^1$H-NMR data in papers V and VI. Various scalings of
the variables depending on the standard deviation and/or the mean
of each variable and two types of block scaling were tried, and the
possibilities of reaching an enhanced classification were explored.
Of course, various types of other data fusion methods, scalings
and multivariate analyses are possible; however, the optimum data
treatment must be tried out for each new case.
Data concatenation

Data concatenation (Figure 19) is denoted as data-level (low-level) fusion according to definitions in the literature.$^{152,155}$ It is straightforward, but most certainly requires appropriate block scaling$^{156-158}$ to prevent one block of data from being numerically dominant. In paper V, a number of different scalings of the concatenated data were evaluated prior to mean centring and multivariate analysis (MVA) – PCA or PLS.

![Figure 19: Data concatenation.](image)

Hierarchical modelling

Hierarchical modelling$^{157,159}$ was also tried in paper V. It is a way of modelling data at two levels: one lower level and one higher level. The lower-level modelling is performed on separate blocks of the data set (LC/MS and $^1$H-NMR in this case), while higher-level modelling is performed on the concatenated and scaled scores from the lower-level, Figure 20. Different combinations of PCA and PLS were tried. This type of data fusion has previously been termed decision-level (high-level) fusion in the literature.$^{152,155}$

![Figure 20: Hierarchical modelling.](image)
Batch modelling

Batch modelling \(^{85,160}\) is a special case of hierarchical modelling in which every batch (rat in paper V), is modelled with respect to one variable (time in paper V) in the lower-level modelling. Prior to the higher-level modelling, the scores from each rat and each technique at the lower level are concatenated and scaled (Figure 21). Different combinations of PCA and PLS were tried.

![Figure 21: Batch modelling.](image)

Outer product analysis

Outer product analysis (OPA) is another possible method of fusing two different data sets in order to determine the relations that exist between the two types of signals.\(^{161}\) The outer product matrix for two vectors is defined as all possible product combinations of the vector variables (Figure 22). In paper VI, this data fusion was performed samplewise, resulting in one matrix per sample.

![Figure 22: Outer product analysis.](image)
Data correlation

Correlation scaling

Paper VI shows the potential of using prior knowledge, in this case identified metabolites, in the data analysis. $^1$H-NMR variables correlated with drug metabolites variables in LC/MS$^{151}$ were identified by correlation scaling and PARAFAC analysis. The correlation scaling was performed by multiplying the NMR variables by the correlation coefficient (33), scaled to the range $[0, 1]$, between the time trajectory of each $^1$H-NMR variable and the LC/MS variables from the identified drug metabolites. These identified $^1$H-NMR signals (most likely arising from drug metabolites) were then eliminated from $^1$H-NMR data. The drug metabolite variables in LC/MS were also eliminated, and the two data sets were fused by OPA and again analysed by PARAFAC. The data fusion step was required to achieve satisfactory class separation. The PARAFAC analysis gave rise, of course, to new $^1$H-NMR variables, separating the classes of dosed and control rats. The kinetics of those variables differed from the kinetics of the drug metabolites and they were thus identified as potential endogenous markers for phospholipidosis. Some of the peaks could be identified from the literature (see paper VI). An example is shown in Figure 23.
Figure 23: A time profile of a $^1$H-NMR spectral segment. The peak at 7.75 ppm was selected in the correlation scaling analysis. It grows with time in the same way as a drug metabolite. The peaks at 6.64 and 7.56 ppm show a decrease (minimum at day 3) and then increase with time, these peaks being selected in the PARAFAC analysis after the elimination of data correlating with identified drug metabolites and being identified as hippurate.
PLS

PLS was also used to find correlating variables in paper VI. There the $^1$H-NMR data were chosen as $X$ and selected LC/MS variables identified as drug metabolites were chosen as $y$. The loadings from PLS (unfolded data) and N-PLS showed the variables separating the sample classes and correlating with the chosen LC/MS variables. These results agreed with the correlation scaling and PARAFAC analysis. However, problems occurred in the next step (with eliminated metabolite variables), when the separation of $^1$H-NMR data was not satisfactory, and no data fusion was possible since the LC/MS data were used as $y$ data.

Discussion

In paper V, data fusion by full hierarchical modelling and concatenation of data generally improved the classification, especially for the unsupervised models (PCA). This indicates that we may have some complementary data between the two data sets. The batch modelling, however, was not very successful in this study.

In paper VI, too, the separation between dosed and control classes was improved by data fusion, although this time OPA was utilised for the purpose. The large-scale use of data fusion in general was discovered in this work when some dominating peaks had been removed and the data fusion was actually needed to discriminate between the different groups of samples.

It is not possible to choose the best data fusion method on the basis of these studies. The four methods presented will be successful to different extents in different cases. The scaling of data is of great importance and should be optimised by a proper validation, as exemplified in paper V.

The data correlation methodology in Paper VI shows a lot of potential in the interpretation of metabolic data. More variables could be selected in both steps of the analysis and hence give more signals for interpretation. Moreover, further analysis of the LC/MS data still remains to be done.
Determination of low-level components by qNMR in organic and biological samples is very easy. The sample is simply dissolved in a deuterated solvent, perhaps buffered to reduce pH variations, put in an NMR sample tube and placed in the magnet. A lot of standard pulse programs have already been prepared, ready for use, in the modern NMR instruments.

However, to get good spectral resolution and quantitative information out of an NMR spectrum, the instrumental parameters have to be carefully chosen. Knowledge of the influence of the sample matrix, relaxation processes and the NMR instrumentation will be of great importance. The optimal parameter settings must be selected for each new type of sample. The work of this thesis includes only the initial step to enable qualified guesses to be made for the basic starting parameters in a qNMR experiment.

The evaluation of qNMR data has been the subject uppermost in my thoughts during this work. Classical data pretreatments (zero filling and line broadening) were tried, together with more unusual and novel ones. The conclusions from these studies were that the classical methods (with typical setups) had very little influence on the quantitative results (papers I–II), while the non-standard methods (convolution by a triangular function and peak alignment) significantly improved the evaluation of qNMR data (papers II–IV).

Extensions in the field of chemometrics were explored by fusion of data from the various analytical techniques $^1$H-NMR and LC/MS (papers V–VI). Data fusion was shown to be successful when more detailed studies in the interpretation of $^1$H-NMR data were performed by means of correlation with LC/MS data.
9 Future perspectives

Of course, several additional parameter setups, data pretreatments, and calibration models are possible in the field of qNMR. Experience from the work carried out on this thesis provides me with even more motivation to tackle new qNMR problems and try new methods. If I get the chance to proceed with this work, my immediate efforts would include further work on the profile/peak alignment problem and nonlinear modelling, which may handle peak shifts, incomplete nuclear relaxations and nonlinear relations between samples. Moreover, the project of data fusion has just been initiated and a lot of scaling and methods remain to be investigated.

In order to extract the maximum information out of a set of samples, the noise must be minimised and the resolution maximised. As the time devoted to experiments has taken up a very small part of the total time in the research projects presented, an increase of more transients in favour of S/N would be beneficial. More and shorter pulses would probably have given a lower LOD in papers I–II. For the metabonomic samples in papers III–VI, an improved spectral resolution, in particular, would have been desirable. This will require improved instrumental facilities such as the sample probe and improved shimming. This would facilitate and improve the final interpretation.

Metabonomic sample collection (which is really outside my subject) may also be an area for improvement. More frequently collected samples would, for example, give a more reliable estimate of the kinetics of a metabolite or biomarker (paper VI).

In the future, the metabolic profiling analysis will perhaps have a more important role to play in toxicology, enabling classification of the type of toxicity. This will probably require more extensive data.
analysis than the most straightforward types of analysis. In papers III–VI, the relevant goal is to find a biomarker time pattern for a specific type of toxicity, namely phospolipidosis. This task (and many others) still remains to be solved and presents a challenge. More than one analysis technique will most likely be needed to give a clear and significant diagnosis or classification. This will require appropriate methods of data fusion. Furthermore, metabolomics, including metabolic profiling, will most likely be of greater significance in the area of systems biology in the future, requiring data fusion techniques.
Acknowledgements

Jag vill hjärtligt tacka alla Ni som direkt eller indirekt hjälpt och stöttat mig under min tid som doktorand på Institutionen för Analytisk kemi (IAK). Det har varit roligt att gå till jobbet varje dag.

Sven, tack för din oändliga entusiasm när det gäller forskningen, det smittar av sig. Få har sån koll på vad som är rätt och hett, det har varit ett privilegium att få ha dig som handledare.

Bosse, tack för att du visat både vetenskapligt och personligt engagemang, trots att du inte har varit min huvudhandledare.


Och TACK alla ni samtida doktorander: Helena I (för ett fantastiskt samarbete i Bergen, på Smyril, på Färöarna, i källaren, i Ume och i Tänndalen – high five!), Malin (för allt allvarliga och lättsamma prat och ditt smittande skratt), Magnus Å (för alla bra diskussioner, mer och mindre vetenskapliga), Stina C.B. (för trevligt rumssällskap och handledning på kurslab), Leila (tack för att du hjälper mig att behålla mina vänner – och för att du dansar så bra!), Helena H (för att du är en lyssnare och en toppenträningsskompis), Petter (tack för allt tjört om JAVA, barn, riskokare, brödbak och locket på eller av), Stina M (för din humor och för att du är så bra på dansa – va fäen Määts!), Magnus E (för att du kan vara både arg och glad), Johanna (för ditt lugn och din attityd – det är en konst att leva i nuet), Sindra (för att du hjälpte mig överleva en grundkurs som assistent), Ove (för en helt underbar kajak-tur och för att du är en rättskaffens), Yvonne (som ju tydligen var doktor hela tiden, fast bara inte på pappret), Ragnar (tack för gott samarbete, och lunchsällskap i bögparken), Thorvald (för att du är så sträng mot studenterna men så barnkär), Ralf (för alla diskussioner som inte leder nånstans, och för alla de som leder nånstans), Kent (för bugg-kursen), Nana (för allt dagis-prat), Leo (för att du fortsätter att jobba för jämställdheten på IAK – och aligna NMR data), Lina (tack för din oblyghet och glada humör), Christoffer (den perfekta studenten), Axel (för det tyska popen), Helena B, Bodil, Cristina S och Sharam.
Tack alla ni andra som gör/gjorde det så trevligt på IAK: Jonas R (för att du så väl har balanserat upp mitt ”bråttom” med ditt lugn, du är ovärderlig), Anders Co (för att du är så män om att stämningen skall vara god på IAK), Anita (för trevligt samarbete på kurs), AnneMarie (för allt du alltid fixar), Carlo (för vin och kulinariska upplevelser), Håkan (för äventyrliga hemresor) Pol (för your whistling), Conny (för att du så vänligt lånar ut din scanner), Ulrika, Lena, Björn, Roger, Ulla, Ramon, Eva, Rolf och Boven.

Och tack Hillevi på kemibiblioteket för din gränslösa service!

Stort tack Ni på AstraZeneca som har hjälpt mig med allt från teoretiska problem till att hitta min dator och mina tofflor: Fredrik, Ina, Bengt, Johan, Ingrid, Joanna med flera. Tack särskilt till Susanne, Anders och Alex; er NMR-hjälp har varit ovärderlig!

Ett extra tack till Er som har korrekturläst och kommit med värdefulla åsikter om manus: Magnus, Susanne, Ralf, Bosse, Sven, Jonas R.

Jag vill också tacka för värdefull inspiration från Ume: Henrik, Johan, Suss, Mattias, Pera m.fl.

Tack alla ni sköna vänner som inte har med kemi att göra: Jenny, Micke, Camilla, Mathias, Gunilla, Alvin, Eva, Anna, Robert, Karin (tackas extra för snowboard och surfing), Danne, Jenny, Sandis, Christina, Magnus, Magda, Ludde, Annika, Peter, Martin, Marie, Ola, Klara.

Jag vill också tacka Ylva, Kjell, Folke (d.ä.), Monica, Tim och Frida för släktkalas och visat intresse.

Tack också alla trivsamma grannar, uthålliga barnvakter, trevliga dagisföräldrar och proffsig dagis- och skolpersonal.

Och min närmsta familj: Farmor och Morfar, tack för att ni finns/fanns. Mamma och Pappa, tack för eft uttröttliga och kravlösa stöd, jag är er evigt tacksam för det. Sara och Carl med familjer, tack för att ni gör något helt annat än jag och på så sätt ger mig perspektiv.

Men... ♫♪ Vad vore livet utan Östen, vad vore livet utan Folke, vad vore livet utan Mina – Ingenting!! – Ingenting!! – Ingenting!! ♪♫ Östen, tack för att du är som du är, jag älskar dig. Folke och Mina, mina kära solstrålar! Det är nog i alla fall ni som har hjälpit mig mest i detta arbetet, genom att ni ger mig den viktiga distansen till allt det här.
References


51. Wells, R.J., Hook, J.M., Al-Deen, T.S. and Hibbert, D.B., Quantitative nuclear magnetic resonance (QNMR) spectroscopy for assessing the purity of technical grade agrochemicals: 2,4-dichlorophenoxyacetic acid (2,4-D) and sodium 2,2-dichloropropionate (dalapon sodium), Journal of Agricultural and Food Chemistry 50 (2002) 3366-3374.


105. Pearson, K., On lines and planes of closest fit to systems of points in space, Philosophical Magazine and Journal B 2 (1901) 559-572.


References | 91


