DISSERTATIONS IN
HEALTH
SCIENCES

Mika Tiainen

Quantitative Quantum Mechanical Analysis of ^{1}H NMR Spectra

Applications and Strategies

Publications of the University of Eastern Finland
Dissertations in Health Sciences

UNIVERSITY OF
EASTERN FINLAND

MIKA TIAINEN

Quantitative Quantum Mechanical Analysis of ^{1}H NMR Spectra

Applications and Strategies

To be presented by permission of the Faculty of Health Sciences, University of Eastern Finland for public examination in Auditorium L2 in Canthia Building of the University of Eastern Finland, Kuopio, on Thursday, May 16th 2013, at 12 noon

Publications of the University of Eastern Finland
Dissertations in Health Sciences
Number 161

School of Pharmacy, Faculty of Health Sciences
University of Eastern Finland
Kuopio
2013

Kopijyvä Oy
Kuopio, 2013

Distributor:
University of Eastern Finland
Kuopio Campus Library
P.O.Box 1627
FI-70211 Kuopio, Finland
http://www.uef.fi/kirjasto

ISBN: 978-952-61-1076-9 (print)
ISBN: 978-952-61-1077-6 (PDF)
ISSN: 1798-5706 (print)
ISSN: 1798-5714 (PDF)
ISSN-L: 1798-5706

Author's address: School of Pharmacy
University of Eastern Finland
KUOPIO
FINLAND

Supervisors: Professor Reino Laatikainen, Ph.D.
School of Pharmacy
University of Eastern Finland
KUOPIO
FINLAND

Pasi Soininen, Ph.D.
School of Pharmacy
University of Eastern Finland
KUOPIO
FINLAND

Reviewers: Professor Erkki Kolehmainen, Ph.D
Department of Chemistry
University of Jyväskylä
JYVÄSKYLÄ
FINLAND

Professor Jacques Vervoort, Ph.D
Laboratory of Biochemistry
Wageningen University
WAGENINGEN
NETHERLANDS

Opponent: Docent Elina Sievänen, Ph.D.
Department of Chemistry
University of Jyväskylä
JYVÄSKYLÄ
FINLAND

Tiainen, Mika
Quantitative Quantum Mechanical Analysis of ^{1}H NMR Spectra: Applications and Strategies
University of Eastern Finland, Faculty of Health Sciences
Publications of the University of Eastern Finland. Dissertations in Health Sciences Number 161. 2013. 61 p.

ISBN: 978-952-61-1076-9 (print)
ISBN: 978-952-61-1077-6 (PDF)
ISSN: 1798-5706 (print)
ISSN: 1798-5714 (PDF)
ISSN-L: 1798-5706

ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is widely used for profiling of a variety of complex biological samples. The advantages of NMR are its quantitative and non-destructive nature, high reproducibility and that measurements can be done with minimal sample preparation and instrument calibration. However, accurate quantification of compounds from a ^{1}H NMR spectrum of biological sample, where the number and concentration range of compounds is large, is a demanding task due to spectral complexity and overlap. Further complications may arise from any instrumental or experimental artefacts. Additionally, the sample conditions, mainly pH and ionic strength, may cause variations to the chemical shifts and line widths.

A distinctive feature of high-resolution 1D NMR spectra is that even the most complex spectrum of a compound, composed of thousands of individual spectral lines, can be described by a few spectral parameters within experimental accuracy employing a quantum mechanical (QM) model. Thus, even in the case of the ^{1}H NMR spectrum of a complex mixture, there are strict QM rules between the lines of individual compounds. The spectral parameters can be extracted from the observed spectra by quantum mechanical spectral analyses (QMSA). In the present thesis, quantification methods, strategies and protocols for NMR spectra of different biological samples were developed. Also adaptive spectral library (ASL) principle including ^{1}H NMR pH indicators was developed and discussed. ASL can be described as a library of spectral parameters obtained through QMSA. The parameters in the library can be used to simulate the spectra of the compounds in any magnetic field, line shape, line widths and, also, taking into account different sample conditions like pH or solvent. Thus, these parameters can be used as a starting point of quantitative Quantum Mechanical Spectral analysis (qQMSA), which means the complete iterative analysis of the spectra based on the QM spectral model and offers an ideal tool for quantification of complex ^{1}H NMR spectra. qQMSA including models describing unknown components, background and prior knowledge from the sample enables modelling of even the smallest details of the spectrum and the maximal quantitative NMR information analysis. In addition to accurate concentrations of known metabolites, qQMSA offers chemical confidence, which means that individual components are identified with high confidence on the basis of their spectral parameters. Also, a protocol for quantification of amino acid ^{13}C isotopomers and determination of positional fractional ^{13}C enrichments for metabolic ^{13}C tracer experiments was developed. This application offers an extreme example of ASL. The protocols and tools developed in this study, ASL and qQMSA, enable accurate, robust and cost-effective way to quantify individual components from the NMR spectra of complex mixtures.

National Library of Medicine Classification: QU 25, QU 60, QV 25, QV 744
Medical Subject Headings: Chemistry Techniques, Analytical; Spectrum Analysis; Magnetic Resonance Spectroscopy; Nuclear Magnetic Resonance, Biomolecular; Metabolomics; Amino Acids/analysis; Glucose/analysis; Hydrogen-Ion Concentration

Tiainen, Mika
Kvantitatiivinen kvanttimekaaninen ^{1}H NMR-spektrien analyysi: Sovellukset ja strategiat
Itä-Suomen yliopisto, terveystieteiden tiedekunta
Publications of the University of Eastern Finland. Dissertations in Health Sciences Numero 161. 2013. 61 s.

ISBN: 978-952-61-1076-9 (print)
ISBN: 978-952-61-1077-6 (PDF)
ISSN: 1798-5706 (print)
ISSN: 1798-5714 (PDF)
ISSN-L: 1798-5706

TIIVISTELMÄ

Ydinmagneettista resonanssispektroskopiaa (NMR) käytetään yleisesti erilaisten biologisten näytteiden profilointiin. Sen etuja muihin yleisiin profilointimenetelmiin verrattuna ovat sen kvantitatiivisuus, hyvä toistettavuus, vähäinen näytteenvalmistuksen ja laitteen kalibroinnin tarve. Lisäksi mittaus ei tuhoa näytettä. Yleensä biologiset näytteet sisältävät suuren määrän yhdisteitä laajalta pitoisuusalueelta. Näiden yhdisteiden tarkkojen pitoisuuksien määrittäminen protoni-NMR-spektristä on kuitenkin vaativa tehtävä, sillä spektrit ovat monimutkaisia ja eri yhdisteiden signaalit ovat suureksi osaksi päällekkäin. Laitteen ja mittausmenetelmän aiheuttamat artefaktit aiheuttavat lisähankaluuksia. Lisäksi näytteen olosuhteet, esimerkiksi pH ja ionivahvuus, voivat aiheuttaa vaihtelua yhdisteiden kemiallisissa siirtymissä sekä juovanleveyksissä. NMR-spektroskopian hyvä puoli on, että jopa kaikkein monimutkaisin yhden yhdisteen tai monen yhdisteen seoksen NMR-spektri voidaan kuvata muutaman spektriparametrin avulla käyttämällä kvanttimekaanista mallia. Spektriparametrit voidaan määrittää havaitusta spektristä kvanttimekaanisen spektrianalyysin (QMSA) avulla.

Tässä väitöskirjatyössä on kehitetty työkaluja erilaisten biologisten näytteiden NMR-spektrien kvantitatiiviseen kvanttimekaaniseen spektrianalyysiin (qQMSA). Väitöskirjatyössä kehitettiin ja rakennettiin myös adaptiivinen spektrikirjasto (ASL), joka sisältää QMSA:llä määritetyt spektriparametrit mukaan lukien niiden riippuvuudet näyteolosuhteista. ASL:n sisältämät spektriparametrit mahdollistavat näytteen sisältämien yhdisteiden NMR-spektrien simuloinnin pH, juovan leveys ja muoto sekä magneettikentän voimakkuus huomioiden, joten niitä voidaan käyttää qQMSA:n lähtökohtana. qQMSA mahdollistaa spektrin pienimpienkin yksityiskohtien kuvaamisen ja maksimaalisen kvantitatiivisen NMR-informaation saamisen, koska käytetyn mallin avulla voidaan kuvata myös tuntemattomat yhdisteet ja spektritausta sekä hyödyntää näytteestä mahdollisesti saatavilla olevat ennakkotiedot. Tunnettujen yhdisteiden pitoisuuksien lisäksi qQMSA:lla saavutetaan kemiallinen varmuus, joka tarkoittaa, että yksittäiset yhdisteet voidaan tunnistaa ja kvantitoida luotettavasti spektriparametriensa avulla. Lisäksi tässä työssä kehitettiin protokolla aminohappojen ^{13}C-isotopomeerien pitoisuuksien ja paikallisten ^{13}C-rikastusasteiden määrittämiseen ^{13}C-leimauskokeissa, mikä on yksi esimerkki ASL:n sovelluksesta. Tässä työssä kehitetyt protokollat ja analyyttiset työkalut, ASL ja qQMSA, tarjoavat tarkan ja kustannustehokkaan tavan määrittää yksittäisten yhdisteiden pitoisuuksia monimutkaisten seosten NMR-spektreistä.

Luokitus: QU 25, QU 60, QV 25, QV 744
Yleinen suomalainen asiasanasto: analyysimenetelmät; spektrianalyysi; NMR-spektroskopia; aineenvaihduntatuotteet; aminohapot; glukoosi; pH

Foreword

The present study was started at the University of Kuopio, Department of Chemistry in 2004 and finished at the University of Eastern Finland, School of Pharmacy in 2013. During this quite a long process I have also worked in the Department of Biosciences (the former Department of Chemistry), in the institute of Clinical Medicine at the University of Oulu and in the PERCH Solution Ldt. The departments are acknowledged for providing the facilities for this work. The Finnish Funding Agency for Technology and Innovation and the Doctoral Program of Organic Chemistry and Biology are thanked for their financial support.

I would like to express my gratitude to my principal supervisor Professor Reino Laatikainen for giving me this opportunity to be a part of his group and for his guidance to the world of chemistry and NMR spectral analysis. I want to express my sincere thanks to Dr. Pasi Soininen, a supervisor and a friend, for his guidance, support, and friendship. We have gone through many things together, for example conference travels and relocating of the NMR magnets. Your "there are no problems" mentality drove me sometimes almost crazy, but it also made things to happen, one way or another.

I wish to express my gratitude to the reviewers Professor Erkki Kolehmainen and Professor Jacques Vervoort for their valuable and constructive comments that helped me to improve the thesis.

I wish to thank Docent Hannu Maaheimo for his contribution to the amino acid isotopomer studies. I would like to thank Dipl.Chem. Matthias Niemitz and Dr. Samuli-Petrus Korhonen for their contribution to the work and, also, for the years I spent in the firm. I also want to thank all of the other collaborators who have contributed to the present work and made it possible.

I wish to acknowledge the staff of the former Department of Chemistry for all the help and support I have received, as well as, the relaxed working atmosphere and all the moments that we have spent together out of hours. I would especially like to thank my good friend and roommate at the University, Tuulia, for her friendship and collaboration. I also want to thank our "Kirves group" (Pasi, Miikka, Tuomas, Petri, Jouko and Janne) for all the relaxing moments including conversations, good food and card games.

I am deeply thankful to my parents, Maire and Teppo, for their love and support during all the years. You have offered me a good and solid starting point for life. I am also grateful to my brother, Marko, and his family for all the moments that we have spent together.

Finally and most importantly, I owe my everything to my beloved wife Kirsi. Thank you for your endless love and support during these years. We have got a lovely daughter, Saana, who has taught me a great deal about life. The two of you are the light of my life.

Kuopio, March 2013

Mika Tiainen

List of the original publications

This dissertation is based on the following original publications:

I Tiainen M, Maaheimo H, Niemitz M, Soininen P and Laatikainen R. Spectral analysis of ^{1}H coupled ^{13}C spectra of the amino acids: Adaptive spectral library of amino acid ^{13}C isotopomers and positional fractional ^{13}C enrichments. *Magnetic Resonance in Chemistry 46: 125-137, 2008.*

II Tiainen M, Maaheimo H, Soininen P and Laatikainen R. ^{13}C Isotope effects on ^{1}H chemical shifts: NMR spectral analysis of ^{13}C-labelled D-glucose and some ^{13}C-labelled amino acids. *Magnetic Resonance in Chemistry 48: 117-122, 2010.*

III Tynkkynen T, Tiainen M, Soininen P and Laatikainen R. From proton nuclear magnetic resonance spectra to pH. Assessment of ^{1}H NMR pH indicator compound set for deuterium oxide solutions. *Analytica Chimica Acta 648: 105-112, 2009.*

IV Laatikainen R, Tiainen M, Korhonen S-P and Niemitz M. Computerized analysis of high-resolution solution state spectra. *Encyclopedia of Magnetic Resonance, 2011.*

V Tiainen M, Soininen P, Laatikainen R. Quantitative quantum mechanical spectral analysis (qQMSA) of ^{1}H NMR spectra of complex mixtures. Submitted.

Contents

Abbreviations

1D	one-dimensional
2D	two-dimensional
ACA	automated consistency analysis
ASL	adaptive spectral library
CASE	computer-aided structure elucidation
CCV	concurrent combined verification
COSY	correlation spectroscopy
CPMG	Carr–Purcell–Meiboom–Gill
CTLS	constrained total-line-shape
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
ERETIC	electronic reference to access *in vivo* concentrations
FC	Fermi contact
FID	free induction decay
GC-MS	gas chromatography-mass spectrometry
HMBC	heteronuclear multible bond correlation
HSQC	heteronuclear single quantum coherence
IR	infrared
IT	integral transform
JRES	*J*-resolved
LC-MS	liquid chromatography-mass spectrometry
MS	mass spectrometry
MWCO	molecular weight cut-off
NMR	nuclear magnetic resonance
PCA	principal component analysis
PULCON	pulse length based concentration determination
qHNMR	quantitative proton nuclear magnetic resonance
QM	quantum mechanical
QMSA	quantum mechanical spectral analysis
qNMR	quantitative nuclear magnetic resonance
qQMSA	quantitative quantum mechanical spectral analysis
SD	spin-dipole
SO	spin-orbit
TLS	total-line-shape
TMS	tetramethylsilane
TOCSY	total correlation spectroscopy
TSP	3-(trimethylsilyl)propionic acid-d_4 sodium salt
UV	ultraviolet

1 Introduction to NMR spectral analysis

1.1 NUCLEAR SPIN AND RESONANCE

The nuclear spin is a fundamental property of atoms arising from that the spinning nuclei possess angular momentum, which gives rise to an associated magnetic moment. When placed into an external magnetic field, such as the magnet of a nuclear magnetic resonance (NMR) spectrometer, the magnetic moments align themselves relative to the field in a discrete number of orientations because the energy states involved are quantised. For example, for proton and carbon-13 there are two possible spin states which align parallel (α-state) or antiparallel (β-state) to the external magnetic field. The α-state is lower in energy and the energy difference (ΔE) between the two spin states is

$$\Delta E = \frac{h\gamma B_0}{2\pi} \tag{1.1}$$

where h is Plank's constant, γ is the magnetogyric ratio of nuclide and B_0 is the external magnetic field.

Since there is an energy difference between the α- and β-states, there is also a difference between the occupancy of the states at the equilibrium. The relative population of a state is given by the Boltzmann distribution

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{kT}} \tag{1.2}$$

where $N_{\alpha,\beta}$ represent the number of nuclei in the spin orientation, k is the Boltzmann constant and T the temperature. Since the energy difference between the two states is small, the corresponding population differences are similarly small. However, the tiny excess of the nuclei in the more favourable α-state generates net magnetisation, which is fundamental to NMR spectroscopy. For example, in an 11.74 T (500 MHz) magnet, the population ratio of proton nuclei states is 0.999987. Thus, NMR is very insensitive relative to other spectroscopic techniques such as infrared (IR) and ultraviolet (UV) spectroscopy, where the ground and excited state energy differences are significantly greater. On the other hand, as a consequence of the low transition energies associated with nuclear resonance, the lifetimes of excited nuclear spins are very long and NMR resonances are very narrow. A distinctive feature of ^{1}H NMR spectrum is that it can be described accurately using quantum mechanical (QM) rules.

1.2 THE SPECTRAL PARAMETERS AND THEIR SENSITIVITY TO CONDITIONS

The main spectral parameters in NMR spectroscopy, discussed below, are the chemical shift (signal position in the spectrum), which tells the type of the nuclei, the coupling constant (splitting of the signal), which bears information about the connectivity of the nuclei, and the area of a signal, which is directly proportional to the number of nuclei in the corresponding group (Figure 1), or in the case of different molecules in a sample, proportional to the concentrations of the components. Also relaxation, line width and line shape are NMR parameters and are shortly discussed in the following chapters.

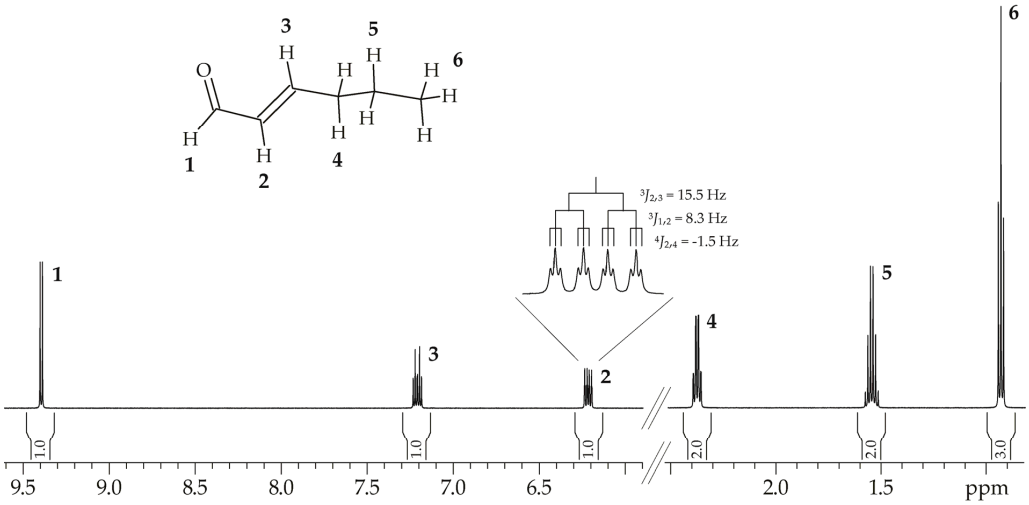

Figure 1. A 600 MHz ^{1}H NMR spectrum and the structure of 2-hexenal. Signal position in the spectrum is the chemical shift of the corresponding nuclei (e.g. the chemical shift of 2-hexenal methyl group (6) is ca. 0.9 ppm). In the insert, the splitting pattern and coupling constants of the signal representing proton 2 (e.g. $^3J_{2,3}$ means coupling over three bonds between protons 2 and 3) are shown. The integrals (the areas of the signals) below the spectrum are directly proportional to the number of nuclei in the corresponding group.

1.2.1 Chemical shift

That the nuclear magnetic moment has a resonance frequency shift from one electronic environment to another and that this shift can be measured easily and with high accuracy has made the chemical shift an excellent tool in distinguishing differences in electronic environments of a molecule. These differences can arise, for example, from chemical reaction, geometric isomerism or hydrogen bonding. (Jameson 1996) From the chemical shift of a nucleus, it is not only possible to determine corresponding functional group, but also more detailed structural information can be obtained, e.g. is the proton axial or equatorial or hydrogen-bonded or not. It can be said that, specially, the proton chemical shift is the most important single parameter in high-resolution NMR. (Abraham 1999) The chemical shift (δ) is given relative to that of a reference compound or standard, and it is defined as follows:

$$\delta = \frac{\nu_{substance} - \nu_{reference}}{\nu_0} \tag{1.3}$$

where ν_0 is the operating frequency of the spectrometer, $\nu_{substance}$ is the resonance frequency of the substance in question, $\nu_{reference}$ is the resonance frequency of a reference compound, and, as an unit for the δ scale, parts per million (ppm) is used. (Günther 1995) In proton NMR spectroscopy, chemical shifts are usually referenced to TMS (tetramethylsilane), TSP (3-(trimethylsilyl)propionic acid-d_4 sodium salt) or DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid), whose chemical shift is set to 0 ppm. However, as a result of the intermolecular interactions between reference and solute or solvent (e.g. TSP-protein interaction and TMS-aromatic stacking), these references are not perfect for automated quantitative NMR (qNMR) applications as the chemical shifts are not reported unambiguously. Thus, for example in the case of the serum, the spectra can be aligned according to an endogenous metabolite signal, for example alanine methyl signal.

Magnetic shielding

Although it is the chemical shift that is measured in practise, the actual magnetic property that can be defined and calculated quantum mechanically is the magnetic shielding of a nucleus, which is converted to a chemical shift when compared with experimental measurements. The theory and calculation of magnetic shielding have been discussed in numerous reviews. (de Dios 1996;Facelli & Orendt 2007;Fukui 1997) In the following description, the magnetic shielding of a proton is the main focus, although the same principles are valid also for all nuclei.

The local magnetic field (B_{obs}) at the nucleus is different than the applied magnetic field (B_0). This effect corresponds to a magnetic shielding (or deshielding) of the nucleus that reduces B_0 by an amount equal to σB_0 where σ is known as the shielding or screening constant of the particular nucleus:

$$B_{obs} = B_0(1-\sigma) \tag{1.4}$$

The shielding constant, σ, can be decomposed to internal, σ_{in}, and external, σ_{ex}, components:

$$\sigma = \sigma_{in} + \sigma_{ex} \tag{1.5}$$

A summary of internal and external components of shielding constant, including short descriptions, is given in Table 1.

The internal component of the shielding constant in equation (1.5) can be described with the following equation:

$$\sigma_{in} = \sigma(d) + \sigma(p) + \sum \sigma_{ik} + \sigma(r) + \sigma(E) \tag{1.6}$$

where $\sigma(d)$ is diamagnetic component, $\sigma(p)$ paramagnetic term, $\sum\sigma_{ik}$ contribution of neighbouring group currents, $\sigma(r)$ ring current term, and $\sigma(E)$ electric field term.

Diamagnetic component, $\sigma(d)$ in the equation (1.6), is a consequence of that the electrons oppose the external field. The external magnetic field induces electron circulations in the ground electronic state. In the case of an unperturbed spherical electron distribution the induced movement of charge leads to a pure diamagnetic effect. In molecules, or in other words, in practise, the situation is more complex since the electronic circulation within the entire molecule must be considered. (Günther 1995)

Paramagnetic term, $\sigma(p)$ in the equation (1.6), originates from that the external field induces a field that is parallel with the external one. Paramagnetic shielding is inversely proportional to the energy gap between the occupied and excited-state orbitals. Thus, for protons the direct paramagnetic contribution to the shielding constant is negligible, because of the large energy gap between the 1s and 2p orbitals. (Günther 1995) Paramagnetic

shielding depends on polarisability and, thus, it is notable, for example, for molecules with double bonds or aromatic rings. Also neighbour group term and ring current term, which are discussed next, depend on polarisability.

Contribution of neighbouring group currents (neighbour group term), $\sum\sigma_{ik}$ in the equation (1.6), is anisotropic. If the induced field does not depend on the orientation of the molecule in the external field, that is isotropic, the neighbour group term averages to zero. If this is not the case, for example in the case of a diatomic molecule AB, A possesses a magnetic anisotropy which can effect a paramagnetic or a diamagnetic shift of the resonance frequency of the nucleus B. Respectively, chemical shift anisotropy (Sitkoff & Case 1998) is defined as the chemical shift difference between the isotropic and anisotropic states.

Ring current term, $\sigma(r)$ in the equation (1.6), means that protons in the molecular plane and outside the ring (cyclic conjugated π system) are deshielded while protons in the region above or below the plane of the ring are strongly shielded. As a simplified model, an aromatic molecule can be visualised as a current loop where the π-electrons are free to move on a circle formed by the σ framework. When these molecules are subjected to the external field, a ring current is induced (Figure 2). Numerous ring current models have been developed to rationalise the peculiar magnetic properties of aromatic molecules. (Cuesta et al. 2009;Lazzeretti 2000)

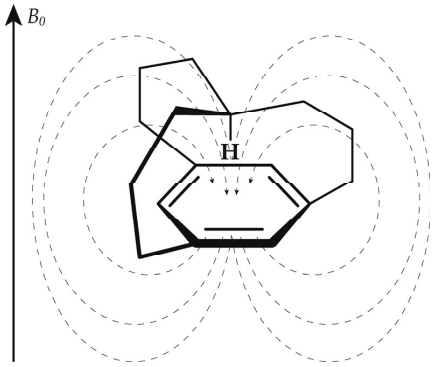

Figure 2. The secondary magnetic field (dashed lines) of a benzene ring resulting from a ring current. This secondary magnetic field can have significant effects to the chemical shifts of nuclei close to the ring. For example, the chemical shift of CH proton of in-[34,10][7]metacyclophane is -4.03 ppm (Pascal, Grossman, & Van Engen 1987) versus 1.74 ppm in isobutane.

Electric field term, $\sigma(E)$ in the equation (1.6), describes neighbouring group effect to charge density that effects magnetic shielding. The electric dipole moment may lead to a change of the charge density at particular protons because the charge cloud of the corresponding C-H bond can be distorted by electrostatic forces. This can happen in the molecules with highly polar groups.

The external components of the shielding constant in the equation (1.5) can be described with the following equation:

$$\sigma_{ex} = \sigma(b) + \sigma(w) + \sigma(e) + \sigma(c) + \sigma(a) \tag{1.7}$$

where $\sigma(b)$ is bulk susceptibility term, $\sigma(w)$ *van der Waals* term, $\sigma(e)$ reaction field term, $\sigma(c)$ effects of complex formation, and $\sigma(a)$ solvent effects.

Bulk susceptibility term, $\sigma(b)$ in the equation (1.7), describes the effects that originate from inconsistency of the sample and the magnetic susceptibility of the solvent. (Augustine & Zilm 1996) Related to this, the use of internal standard, instead of external, is highly

recommended because the differences in the bulk susceptibilities of the sample and the external standard can cause a minor bias to the observed chemical shift. (Live & Chan 1970) However, if the external standard is used, the bulk susceptibility correction should be taking into account. (Harris et al. 2001) For example, ^{31}P chemical shifts in aqueous samples are commonly referenced to phosphoric acid (δ = 0.00 ppm), but if the reference substance is in a coaxial insert, a correction of -0.73 ppm must be used. (Batley & Redmond 1982)

Table 1. Summary of internal and external components of shielding constant.

Component	**Description**
Internal components	
Diamagnetic, *σ(d)*	A consequence of that the electrons oppose the external field.
Paramagnetic, *σ(p)*	Originates from that the external field induces a field that is parallel with the external one.
Neighbouring group currents, $\Sigma\sigma_{ik}$	Anisotropic, almost all chemical bonds are magnetically anisotropic.
Ring current term, *σ(r)*	Protons in the molecular plane and outside the ring (cyclic conjugated *π* system) are deshielded while protons in the region above or below the plane of the ring are strongly shielded.
Electric field term, *σ(E)*	Describes the neighbouring group effect to charge density.
External components	
Bulk susceptibility, *σ(b)*	Describes the effects that originate from inconsistency of the sample and the magnetic susceptibility of the solvent.
van der Waals term, *σ(w)*	Arises from the strong steric interaction between a proton and a neighbouring group.
Reaction field term, *σ(e)*	Solvent-solute dipole-dipole effects.
Complex term, *σ(c)*	Effects of complex formation.
Solvent effects, *σ(a)*	Interactions between solute and solvent: hydrogen bonding, solvent molecules' anisotropy, polar effects, *van der Waals* interactions.

Van der Waals term, *σ(w)* in the equation (1.7), arises from the strong steric interaction between a proton and its neighbouring group. This deforms the electron cloud around the proton and the decreased spherical symmetry of the electron distribution causes a paramagnetic contribution to the shielding constant. (Abraham, Warne, & Griffiths 1997)

Reaction field term, *σ(e)* in the equation (1.7), describes the solvent-solute dipole-dipole effects (Kotowycz & Schaefer 1967), and complex term, *σ(c)* in the equation (1.7), effects of complex formation, which is related to complexation induced shifts (Hunter, Packer, & Zonta 2005). Above-mentioned effects are usually small, but they are origins of the solvent and concentration effects.

Solvent effects, *σ(a)* in the equation (1.7), originate from the interactions between solute and solvent. The interactions responsible for these effects are hydrogen bonding, the anisotropy of the solvent molecules, polar effects, and *van der Waals* interactions. (Abraham et al. 2006;Holzgrabe 2010) For example, aromatic solvents tend to produce high-field shifts in the solute. Related to this, solvent-reference combinations with specific interaction should be avoided. For example, chloroform associates with benzene in such a way that the chloroform proton is specifically shielded. A change of solvent can be used to change

chemical shifts by design. (Holzgrabe 2010) Solvent effects are particularly significant when intermolecular interactions in the solvent lead to the formation of weak complexes. Table 2 gives some examples of solute chemical shifts in different solvents.

Table 2. Examples of chemical shifts (ppm) for some small common organic compounds in different solvents (values are from the in-house database).

Solute	Solvent		
	DMSO	D_2O	$CDCl_3$
	δ (ppm)	δ (ppm)	δ (ppm)
Acetic acid	1.910	2.080	2.100
Acetone	2.090	2.220	2.162
Benzene	7.370	7.448	7.360
Chloroform	8.320	7.681	7.260
Dichloromethane	5.760	5.465	5.300
Ethanol	1.060, 3.440	1.190, 3.664	1.250, 3.720
n-pentane	0.860, 1.225, 1.278	-	0.883, 1.246, 1.303
Pyridine	8.580, 7.790, 7.390	8.600, 7.910, 7.480	8.615, 7.665, 7.275
Triethylamine	0.930, 2.430	0.990, 2.570	1.030, 2.530

Isotope Effects

When an isotopic label is introduced into a molecule, the neighbouring resonant nucleus experiences an observable chemical shift, and if the labelling percentage (enrichment) is not 100, the resonant nuclei in both the labelled and the unlabelled molecules are observed. There are two classes of isotope effects on nuclear shielding, primary and secondary. If the isotopic label itself is the resonant nucleus, the isotope effect is called a primary isotope effect. Secondary isotope effect is caused by the change in the isotope of the neighbouring atoms. Additionally, the isotope effects observed in an NMR spectrum can be divided into two categories according to that if they are directly caused by isotope effects on nuclear shielding (direct isotope effects) or indirectly by the fact that isotope substitution may cause a change in chemical equilibrium, which then causes a change in nuclear shielding. The isotope shift is by convention the chemical shift of the nucleus substituted by the heavier isotope minus that substituted by the lighter isotope. However, also the opposite sign convention is used. As a unit for the isotope effect, parts per billion (ppb) is used. (Hansen 1988;Jameson 2007)

The nuclear shielding can be considered as a function of the nuclear configuration of the molecule. The internuclear distances in a molecule are affected by the vibrational and rotational motions of the molecule, and since the vibration is in general anharmonic, the vibrating molecule is deformed from the equilibrium configuration. Additionally, the centrifugal forces caused by the overall rotation act on the atoms to shift their average positions away from the centre of gravity of the molecule. Thus, the observed nuclear shielding is a value characteristic of the thermal average of internuclear distances. The anharmonic vibration and the centrifugal distortion contribute to a larger mean bond extension in the lighter isotopomer than in the heavier one. The isotope dependence of the shielding is a consequence of the anharmonicity of molecules and the isotope effects result mainly from the fact that the heavier isotopomer has, on the average, shorter bond length. (Jameson 1981)

Some general rules about isotope effects on chemical shifts have been proposed: the size of the effects depends on (i) the mass ratio of the isotopes and (ii) the chemical shift range of the nucleus. In addition, (iii) isotope effects due to multiple isotope substitution are normally additive, (iv) the magnitude of the isotope effects decreases with increase in the distance to isotope, and (v) substitution with a heavier isotope usually shifts the NMR signal of the nearby nucleus towards lower frequencies (higher shielding). (Jameson 2007)

Temperature dependence

Temperature dependence of the chemical shift (shielding) of rigid molecules in the intramolecular level is normally dominated by bond stretching factors. When temperature is raised, a deshielding value is observed since the bonds are getting longer and the shielding derivative with respect to bond length changes is negative. However, in practise, temperature dependence of the chemical shift can be positive or negative and is caused by intermolecular interactions and conformational changes. (de Dios 1996;Jameson 1981)

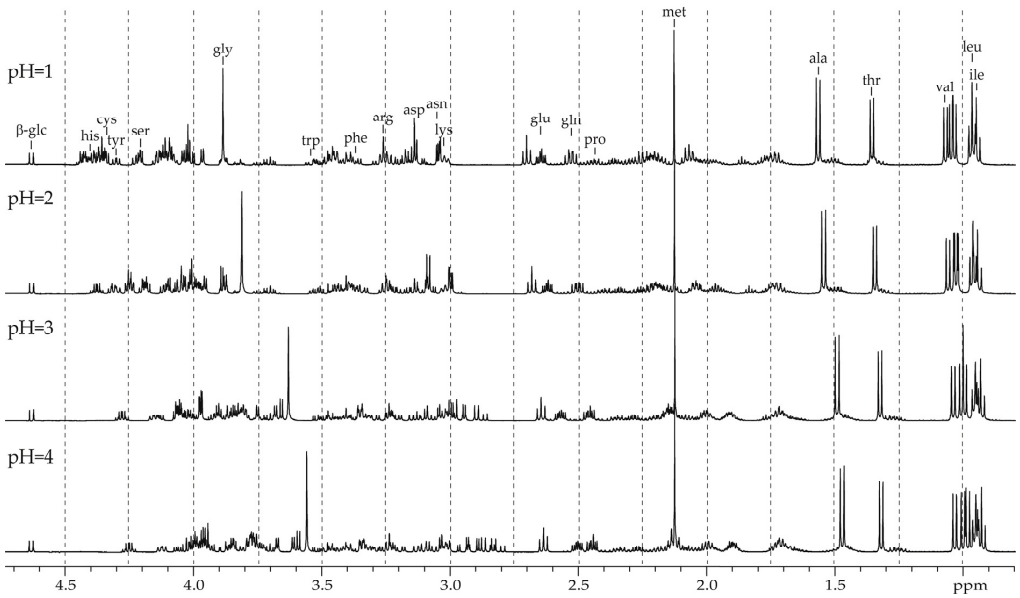

Figure 3. The 500 MHz ^{1}H NMR spectrum of a mixture of amino acids and D-glucose at four pHs. The chemical shifts of glucose (neutral compound) are not sensitive to pH as can be seen in the case of assigned signal (β-glc). The chemical shifts of amino acids α-protons are the most sensitive ones - typical shift being ca. 0.3 ppm (gly). Abbreviations: ala=alanine, arg=arginine, asn=asparagine, asp=aspartate, β-glc=β-glucose, cys=cysteine, gln=glutamine, glu=glutamate, gly=glycine, his=histidine, ile=isoleucine, leu=leucine, lys=lysine, met=methionine, phe=phenylalanine, pro=proline, ser=serine, thr=threonine, trp=tryptophan, tyr=tyrosine and val=valine.

pH dependence

The chemical shifts in molecules like amines and carboxylic acids show a strong dependence on pH due to the introduction of charges into a system (Figure 3). These charges create perturbations in the electron distribution around the nucleus causing a measurable change in the resonant frequency. Thus, it is possible to evaluate events such as ionisation and protonation state via changes in the chemical shift of nuclei near to the locus of the perturbation. (Reily et al. 2006) For fast proton exchange, an average spectrum is observed that results from the protonated and deprotonated species that are in equilibrium. The pH dependence of a chemical shift can expressed with the following equation:

$$\delta_{obs} = x_{HA}\delta_{HA} + x_{A^-}\delta_{A^-} = \frac{\delta_{A^-} + 10^{pK-pH}\delta_{HA}}{1+10^{pK-pH}} \tag{1.8}$$

where pK is the negative logarithm of the dissociation constant, δ_{HA} and δ_{A^-} are the chemical shifts, and x_{HA} and x_{A^-} the molar fractions of the protonated and the deprotonated forms, respectively. (Szakács & Hägele 2004)

The relationship between pH and the chemical shift can also be expressed with Henderson-Hasselbalch equation

$$pH = pK_a + \log\frac{\delta_{HA} - \delta_{obs}}{\delta_{obs} - \delta_A} \quad (1.9)$$

where pK_a is the negative logarithm of the acid dissociation constant, δ_{obs} is the observed chemical shift, δ_{HA} and δ_A are the chemical shifts of the protonated and the deprotonated forms, respectively. The Henderson-Hasselbalch equation or its modifications have been used in numerous works concerning pH indicators utilising chemical shift. (Rabenstein & Isab 1982;Szakács & Hägele 2004;Publication III)

Concentration dependence

Also concentration can cause variations to the chemical shifts due to the intermolecular interactions. For aromatic systems like quinolines, indols, and naphthalenes considerable shift changes (from -0.24 to 0.09 ppm/M) have been observed in different concentrations and solvents. (Mitra et al. 1998) The concentration dependence of the chemical shift can also be used to increase the information available in the spectrum. For example, with certain concentration of benzyl alcohol in acetone the resonance of the CH_2OH group appears as a singlet (instead of doublet and triplet), while in pure benzyl alcohol the expected AB_2 system is observed. (Günther 1995) In some cases, the concentration-dependent resonance can be used for quantification if the correlation between sample concentration and chemical shift is known. (Michaleas & Antoniadou-Vyza 2006)

In conclusion, chemical shift is a sum of several internal and external terms of same magnitude making it difficult to predict, but, on the other hand, chemical shift is easy to measure accurately and carries plenty of information about nucleus and the overall electronic environment surrounding the nucleus. This is why the chemical shift is the most important characteristic of a nucleus in terms of NMR. From the qNMR point of view, the sensitivity of the chemical shift to conditions forms the biggest challenge for the automated mixture analysis and this is why analysis methods that are capable to take chemical shift variations into account are needed.

1.2.2 Coupling constants

The energy state of a nucleus can be affected by the spin state of other nuclei. This interaction, called spin-spin coupling, is transmitted by the bonding electrons of a molecule and can be observed as splitting of signals. The spin-spin coupling constant between two nuclei depends on the distribution of electrons in a bond or bonds connecting these nuclei and, thus, it provides detailed information about the connectivity of the nuclei in a molecule. (Cremer & Grafenstein 2007) Also two other mechanisms of spin-spin interaction, direct spin-spin and through-space couplings/interactions are shortly introduced at the end of the following chapter.

Mechanisms

Spin-spin coupling is transmitted by four different mechanisms from a perturbing nucleus, which by its magnetic spin moment perturbs the surrounding electron density, to the responding nucleus, whose magnetic moment receives the perturbation of the electron density and responds to it. (1) The Fermi contact (FC) mechanism relies on the probability of finding an electron in nuclei. It is expected that σ electrons will have a very significant role since these are the only electrons that do not have nodes at the nuclear sites. However, the FC coupling can be transmitted long distances through the π-electronic system because of exchange interactions between the σ- and π-electronic systems. The spin-orbit (SO) mechanisms are associated with orbital currents generated by the spin moment of the

perturbing nucleus; the electron currents are accompanied by a magnetic field, which is experienced by the responding nucleus. (2) In the diamagnetic SO case, the circular currents depend on the molecular ground state, whereas (3) in the paramagnetic SO case, the orbital currents depend on the existence of appropriate excited states. (4) The spin dipole (SD) mechanism originates from the spin polarisation caused by the external magnetic field. The SO mechanisms and the SD mechanism, which are notable for heavier nuclei with d-electrons, are related to polarisability and cause solvent dependence. In general, the FC term is the most important contribution to the scalar coupling, the exception being couplings involving fluorine atoms. (Alkorta & Elguero 2003;Gräfenstein, Tuttle, & Cremer 2004)

Coupling interaction

In the case of two directly bonded nuclei A and X, the coupling interaction proceeds as follows. The magnetic moment of nucleus A causes a weak magnetic polarisation of the bonding electrons so that the neighbouring electron's spin is lined up in opposition to the nuclear spin of A. Then by the Pauli exclusion principle the two electrons in the A-X bond must be antiparallel and that is why the other electron's spin is lined up in opposition to the previous one. Finally, the latter electron interacts with the X nucleus to again produce an antiparallel orientation of the spins. As a result of this coupling interaction, the A and X nuclear spins are in opposite orientations and this is a positive coupling. By definition the coupling constant is positive when the low-energy state has an antiparallel arrangement of nuclear moments and when the low-energy state has a parallel arrangement coupling constant is negative. (Abraham 1971)

Simple splitting rules

Some simple rules for splitting can be represented. The multiplicity of an NMR signal caused by n neighbouring nuclei is given by $2nI + 1$. Thus, for nuclei with spin quantum number I=½, like protons, coupling to n nuclei splits the signal into an n+1 multiplet with intensity ratios following the Pascal's triangle. Coupling to additional spins will lead to further splittings of each component of the multiplet e.g. coupling to two different nuclei with significantly different coupling constants will lead to a doublet of doublets. If spin-spin coupling involves nucleus that has a spin quantum number I greater than ½, the multiplicity and the intensity distribution of the splitting pattern differ from those described above. For example, proton's coupling to a deuteron (I=1) splits the proton signal into a triplet with equal intensities, because the spin 1 has three equally probable spin states (m_I = +1, 0 and -1). The line separations expressed in Hz correspond to the coupling constants between the nuclei under consideration. In general, the magnitude of the coupling between protons decreases as the number of bonds between the coupled nuclei increases. Finally, the splitting patterns are independent of the signs of the coupling constants. (Günther 1995)

Above described simple splitting rules fail in a couple of cases. Coupling between nuclei that are magnetically equivalent, e.g. the protons in a methyl group, has no effect on the outlook of the spectrum. In addition, second-order spectrum (discussed more detailed later on) does not follow simple splitting rules. Instead, increased multiplicity and altered intensity distribution are observed. On the other hand, the rule that any coupling, which is observed in the signal of one nucleus, must also be found in that of the coupled nucleus is valid even in complex spectra in which the line separation observed does not equal the coupling constant. (Abraham 1971)

Structural dependence

Spin-spin coupling constant is sensitive to the electronic structure, geometry, and conformation of a molecule. One-bond coupling constant, ^1J, reflects the nature of the chemical bond; two-bond coupling constant, ^2J, known as a geminal coupling constant, depend on the bond angle and, thus it is sensitive to the bond angle strain. Also, three-bond coupling constant, ^3J, known as a vicinal coupling constant, depend on the dihedral angle of a three-bond fragment in a characteristic way (see below) (Wu, Gräfenstein, & Cremer 2003) Finally, long-range coupling constants, ^nJ ($n \geq 4$), are sensitive to the stereochemistry of the molecule. (Schaefer 2007) In the following text only the angular dependence of couplings is introduced, even though other factors affect the coupling constants too, including bond lengths, substituent electronegativity, and orientation. (Altona 2007;Esteban et al. 2001;Günther 1995;Minch 1994;Tormena et al. 2004)

The angular dependence of couplings can be divided into three different types. (1) Hybridization effect describes the dependence of couplings on the bond angles between the bonds containing any of coupled nuclei and other bonds attached to the same atom. (2) Dependence of couplings on the dihedral angles defined by bonds along the coupling pathway. (3) How coupling is affected by the orientation of a moiety proximate in space to the coupling pathway. For ^1J and ^2J, the angular dependence of type (2) does not apply, whereas for ^nJ n ≥ 3 (2) is the main factor. In general, the angular dependence of long-range couplings is different for saturated and for partially saturated systems because in unsaturated systems the π-electrons have often important role in transmission of long-range coupling. (Barfield 1971;Contreras & Peralta 2000)

The best known example of the angular dependence of coupling constant is the dependence of vicinal proton-proton coupling constant on the dihedral angle, ϕ, of a three bond fragment (Figure 4), which is exploited in the empirical functions. For example, in the original Karplus equation the dependence is described by the relation

$$^3J(H,H) = A + B\cos\phi + C\cos 2\phi \tag{1.10}$$

where A, B, and C are constants and ϕ is the dihedral angle. (Karplus 1959;Karplus 1963) Many modifications to this original equation have been introduced (Altona 2007;Haasnoot, de Leeuw, & Altona 1980) aiming to improve the accuracy of the calculated coupling constants. A series of important regularities is explained by this relation. For example, in olefinic systems the coupling of trans protons is always greater than that between cis protons and, same way in 1,2-disubstituted ethane, J_{trans} is greater than J_{gauche}. Therefore, in the chair conformation of cyclohexane the coupling between two axial protons (J_{aa}) is bigger than that between two equatorial protons (J_{ee}) or between an equatorial and an axial one (J_{ea}) ($J_{aa} > J_{ea} \approx J_{ee}$). Preceding regularity is an important criterion in the conformational analysis of cyclohexane derivatives and carbohydrates. (Günther 1995)

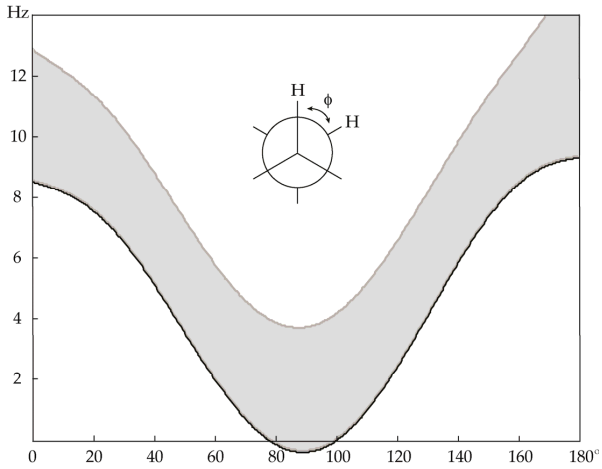

Figure 4. The Karplus curve for the dependence of vicinal proton-proton coupling on the dihedral angle (φ): the dark line represents theoretical curve and the shaded area the range of empirical results.

Solvent and temperature dependence

In almost all cases solvent effects amount to only a few per cent of the total value of the coupling constant. Geminal proton-proton coupling constants often exhibit substantial variations. Solvent dependence of vicinal coupling constants is caused by solvent-induced changes in conformational populations. (Barfield & Johnston 1973) The temperature dependence analysis of the vicinal coupling constants can be used to characterise the conformational behaviour of compounds. For example, the vicinal coupling constant ($^3J_{2,3'}$) of *n*-butane measured in chloroform at different temperatures (from 240 to 320 K) varies from 9.212 Hz to 8.557 Hz and can be used in the analyses of the conformational behaviour of *n*-butane. (Tynkkynen et al. 2012)

Direct spin-spin interaction

The direct magnetic interaction of nuclear moments through space is called a dipole-dipole or simply a dipolar coupling. As a result of the random thermal translational and rotational motions of the molecules in liquid, no line splitting originating from dipolar coupling is observed. On the other hand, in a solid (Middleton 2011) or liquid crystal (Fung 2002) sample where motions of the molecules are restricted, line splitting due to dipolar coupling will occur. (Günther 1995)

Through-space interaction

Through-space coupling is a variation of the spin-spin coupling transmitted by electrons and it can be detected when, as the result of steric compression, an extensive non-bonding or *van der Waals* interaction of orbitals occurs. This leads to transmission of magnetic information through a "short circuit" where no formal bonds are present. This mechanism has more extensive significance for spin-spin coupling between a proton and a fluorine nucleus or between two fluorine nuclei than two proton nuclei. (Günther 1995;Tuttle, Gräfenstein, & Cremer 2004)

Opposite to chemical shifts, spin-spin coupling constants are not as sensitive to the conditions and molecular tertiary structure, where intramolecular interactions may lead to large effects on chemical shifts. Thus accurate coupling constants can be highly diagnostic and used to identify a certain type of fragment even in mixtures. (Hanhineva et al. 2009)

1.2.3 Relaxation times

The lifetimes of excited nuclear spins are remarkably long when compared to the excited electronic states of optical spectroscopy. These extended lifetimes, which are a consequence of the low transition energies associated with nuclear resonance, are crucial to the success of NMR spectroscopy. As consequences, NMR resonances are narrower than rotational, vibrational or electronic transitions and, additionally, it also provides time to manipulate the spin systems after initial excitation. Latter is vital for multi-pulse NMR experiments. (Claridge 1999)

As a result of pulse excitation of nuclear spins the net magnetisation vector is moved away from the thermal equilibrium, which means a change in the spin populations. The recovery of the magnetisation, which corresponds to the equilibrium populations being re-established, is called longitudinal relaxation (Figure 5). The energy lost by the spins, related to recovery of the magnetisation, is transferred in the form of heat into the sample itself. After 90° pulse, the recovery of magnetisation follows the expression

$$M_z = M_0\left(1 - e^{-\frac{t}{T_1}}\right) \tag{1.11}$$

where M_0 is the magnetisation at thermal equilibrium, t is time, and T_1 is the first-order time constant for this process. T_1 is usually referred to as the longitudinal relaxation time, although it is a time constant rather than a measure of the time required for recovery. (Claridge 1999;Traficante 2007)

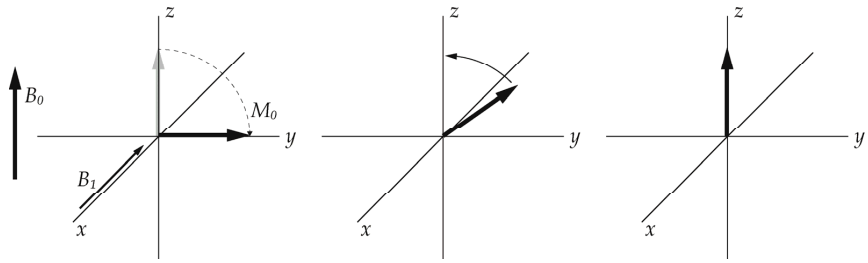

Figure 5. As a result of a 90° pulse (B_1) the net magnetisation vector is moved away from the thermal equilibrium to the (x, y) plane. The recovery of the magnetisation is called longitudinal relaxation.

The net magnetisation vector is generally pictured as a single vector, although it is a vector sum of many smaller ones. When the net magnetisation vector reaches the (x, y) plane after a 90° pulse, the inhomogeneous external magnetic field will act on these individual vector components, which are said to possess phase coherence following the pulse (Figure 6). Even though homogeneous field is desired, the field remains inhomogeneous even after careful shimming and the individual components, located in different positions throughout the sample, will experience different field strength. This results in a fanning-out of the individual magnetisation vectors, which cause net magnetisation to decay in the transverse plane (Figure 6). This form of relaxation is called transverse relaxation and corresponding time constant T_2. (Claridge 1999;Günther 1995;Traficante 2007) Although the inhomogeneity of the B_0 field is the main cause of this decay of magnetisation for high-resolution proton and carbon spectra of liquids, this decay will still occur, albeit more slowly, even in the case of perfectly homogeneous field, owing to an exchange of energy between two spins (random phase changes). (Traficante 2007)

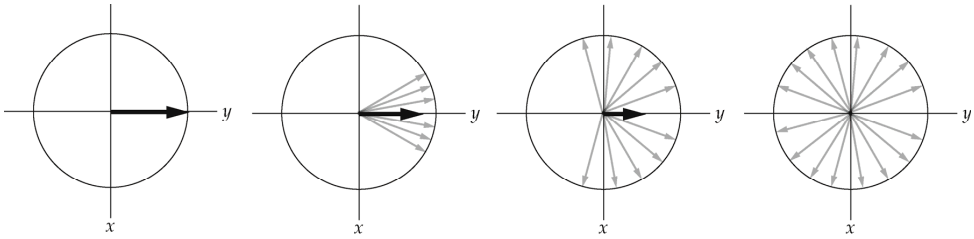

Figure 6. As a result of local field differences within the sample, spins precess with slightly differing frequencies eventually leading lost of phase coherence and zero net transverse magnetisation.

Magnetic field differences arise from static magnetic field inhomogeneity throughout the sample and from the local magnetic fields arising from intramolecular and intermolecular interactions in the sample. The former is an instrumental imperfection and the latter represent 'genuine' or 'natural' transverse relaxation processes. The corresponding relaxation time constant is T_2^*

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2(\Delta B_0)} \tag{1.12}$$

where T_2 refers to the contribution from genuine relaxation processes and $T_2(\Delta B_0)$ to that from field inhomogeneity. (Claridge 1999)

Since the time required for spontaneous emission in NMR is long, it has no effect on the spin populations and so stimulated emission must be operative for relaxation to occur. The fundamental requirement for inducing nuclear spin transitions is a magnetic field oscillating at the Larmor frequency of the spins. These fields can originate from a variety of sources coming from local molecular motions. (Claridge 1999) There are six sources for these fields: magnetic dipole-dipole interaction, spin-rotation interaction, chemical shift anisotropy, quadrupolar interaction, scalar interaction, and electron-nuclear or paramagnetic interaction. Magnetic dipole-dipole interaction results from a 'through-space' interaction of magnetic moments of other nuclei. Spin-rotation interaction originates from the distribution of electrons in the rotating molecule, or portions of the molecule (segmental motion). Chemical shift anisotropy is due to the variability of the nuclear shielding as a function of the orientation of the molecule with respect to the B_0 lines of force. Quadrupolar interaction, directly relevant only for those nuclei having a nuclear spin quantum number, I, greater than ½, is a consequence of the changing direction of the electric field gradient at the nucleus. Scalar interaction originates from fluctuations of the coupling constant or fluctuations of the spin state of the coupled nucleus. Electron-nuclear or paramagnetic interaction results from the presence of unpaired electrons. (McConnell 1987;Traficante 2007) For spin-½ nuclei in solution state the two most significant interactions contributing to the relaxation are dipole-dipole and spin-rotation or chemical shift anisotropy interaction depending on the type (size) of the molecule (small organic molecule versus bigger biomolecule). (Nicholas et al. 2010;Traficante 2007)

Knowledge of relaxation processes is important for many reasons. The proper operation of a spectrometer, even in the case of a single pulse experiment, requires the precise settings of pulse widths and time delays to optimise the resolution and sensitivity. In addition, these have effect to, for example, integral accuracy and nuclear Overhauser enhancements. (Claridge 1999;Traficante 2007) It is crucial for qNMR to use relaxation delay (the delay before the excitation) that is long enough since using too short relaxation delay relative to T_1, leads to a substantial decrease in signal's intensity. After a 90° pulse, it is recommended to wait at least 5xT_1 of the slowest relaxing nuclei. At this point, the magnetisation has recovered by 99.33%. Measurements of relaxation times are useful for calculating diffusion

constants, overall molecular motion and segmental motion. They can also be used for making line assignments, determining internuclear distances, and establishing molecular structure. (Traficante 2007) However, the relationship between relaxation rates and structural features are not as well defined as those of the chemical shift and coupling constants. This is mainly caused by the numerous extraneous effects that influence experimental results. (Claridge 1999) The transverse relaxation time (T_2) is inversely related to the sample viscosity, whereas the longitudal relaxation time (T_1) is smallest when the rotational fluctuations occur at rate comparable to the Larmor frequency of the nuclei in question. The relaxation times of substrate nuclei can also be strongly decreased by paramagnetic impurities or relaxation reagents, like chromium tris(acetylacetonate). These reagents are typically used in ^{13}C NMR spectroscopy. (Levy & Komoroski 1974)

1.2.4 Line width and shape

The widths of NMR resonances are inversely proportional to T_2^*. A short T_2^* corresponds to a fast decay of the transverse magnetisation, which means a greater frequency difference between the individual magnetisation vectors and hence a greater dispersion in the frequency dimension, on the other words, broader line. For exponential relaxation the line shape is Lorentzian with a half-height line width, $\Delta\nu_{½}$ of

$$\Delta\nu_{½} = \frac{1}{\pi T_2^*} \tag{1.13}$$

For most spin-½ nuclei in small, rapidly tumbling molecules in low-viscosity solutions, it is inhomogeneity of the field that dominates the observed line width. The effect of inhomogeneity factor is more critical for nuclei with higher Larmor frequency, e.g. more critical for proton than for carbon. (Claridge 1999;Harcken et al. 2010)

The line shape of NMR signals is sensitive to chemical exchange processes if they affect the NMR parameters of the corresponding nucleus. This phenomenon is exploited in dynamic NMR studies, for example, to study fast reversible reactions. The classical example of dynamic NMR is the ^1H NMR spectrum of *N*,*N*-dimethylformamide in which one or two methyl signals are observed depending on the used sample temperature. (Günther 1995)

By modifying the free induction decay (FID), one can influence the line width and shape. The most common modification that is done to the FID is noise reduction by exponential multiplication. As the result of this exponential window function, lines are broadened, but a Lorentzian line shape of NMR signal is remained. Thus, exponential multiplication is a compromise between resolution and signal-to-noise ratio (S/N). Multiplication with Gaussian function is used to change the shape of the lines so that the line shape becomes a mixture of Lorentzian and Gaussian. The Gaussian line shape is narrower than the Lorentzian, especially near the root of the line. As a result of Gaussian multiplication, the resolution of the spectrum is enhanced but this happens at the expense of the S/N. Long-range coupling constants have contribution to the effective line width and, as a result of many small coupling constants, lines can appear very broad and no splitting at all is observed. In some cases, resolution enhancement can be used to reveal this coupling information. The line width variations are one of the main reasons why quantitative Quantum Mechanical Spectral Analysis (qQMSA) (Manuscript V), which means the complete iterative analysis of the spectra based on the quantum mechanical spectral model, outruns conventional quantification methods.

1.2.5 The area of the signal

The area of a signal is proportional to number of nuclei in corresponding group or, in the case of different molecules in a sample, proportional to concentrations of the molecules. Thus, the area of the signal is per se the most important parameter when speaking on qNMR, although, in practise, it is not solely usable, since the bare area does not carry any information about the corresponding signal's chemical environment or connectivity. This means that the area of the signal does not offer identification for corresponding signal.

In theory, each proton within a compound gives equal signal area. However, there are reasons why this is not fulfilled in practise: different relaxation times, water suppression, macromolecular interactions (with T_2-editing), exchangeable protons and imperfect pulse calibration can cause up to tens of percentages bias to the areas. These effects have a special importance in the accurate quantification of low concentration components overlapping with high concentration components. (Manuscript V)

1.3 QUANTITATIVE ^{1}H NMR

Quantitative ^{1}H NMR spectroscopy (qHNMR), or more generally qNMR, refers to the use of NMR spectroscopy to determine the concentration of one or more components in the sample based on the earlier mentioned fact that the area of an NMR signal is directly proportional to number of nuclei or concentration (see chapter 1.2.5). qNMR is an invaluable tool used in the analysis of mixtures in many areas such as pharmaceutical industry (Holzgrabe et al. 2005), natural products (Pauli, Jaki, & Lankin 2007), metabolic profiling (Beckonert et al. 2007), food and beverages (Consonni et al. 2008;Nord, Vaag, & Duus 2004), combinatorial chemistry (Rizzo & Pinciroli 2005) and on-flow monitoring of reaction processes (Bernstein, Štefinovic, & Sleigh 2007).

If NMR data is to be used for quantitative purposes, the data acquisition and processing must follow quantitative conditions discussed in many reviews. (Barding, Salditos, & Larive 2012;Holzgrabe 2010;Pauli, Jaki, & Lankin 2004) In addition to the appropriate parameters, high S/N is needed for accurate quantitative work. For absolute quantification, a S/N larger than 100:1 is preferred, whereas for relative quantification, a S/N larger than 1:10 is acceptable. (Barding, Salditos, & Larive 2012) Also, a flat spectral baseline without phase distortions is essential for the accurate determination of signal areas. Modern spectrometers have a very linear baseline, and thus the baseline describes mostly broad signals originating from macromolecules, not instrumental artefacts. Additionally, the ^{13}C satellites of the more abundant compounds can interfere with the quantification of lower abundance compounds from ^{1}H spectrum. This can be handled by using ^{13}C decoupling. (Soininen et al. 2005) Solvent suppression and spectral editing techniques, for example the removal of protein background by T_2-editing, are other matters that need to be taken into account in qHNMR. (Barding, Salditos, & Larive 2012)

One-dimensional (1D) qHNMR techniques, such as 1D ^{1}H NMR spectroscopy (Pauli, Jaki, & Lankin 2004), 1D *J*-resolved (JRES) spectroscopy (Fonville et al. 2010), Carr–Purcell–Meiboom–Gill (CPMG) edited ^{1}H spectroscopy (Nicholson et al. 1995), and 1D heteronuclear single quantum coherence (HSQC) spectroscopy (Koskela et al. 2007), have been demonstrated to work well in quantitative studies. Nowadays, there is also a significant number of quantitative two-dimensional (2D) NMR techniques that have been proved to provide reliable results when compared to the 1D qHNMR spectral data. (Pauli et al. 2012) These include 2D JRES experiment (Ludwig & Viant 2010), homonuclear 2D correlation spectroscopy (COSY) experiment and its variations (Martineau et al. 2011), and 2D HSQC experiment (Heikkinen et al. 2003;Koskela et al. 2010;Peterson & Loening 2007). Although these experiments are customised for qNMR, the use of calibration is often a necessity. (Pauli et al. 2012)

Spectral area can be determined using several different approaches. Classical integration is performed by defining an integral region around the signal of interest and then measuring the relative step heights from the step curve. Integral regions should be chosen so that the overlapping of the integrated signals is minimised, and, at the same time, integral regions should be wide enough due to the Lorentzian nature of NMR peaks. In practise, this means that the parts of the peaks that are left outside of the integration range and also the tails of the neighbouring peaks (even if visually resolved), affect the integration results. (Griffiths & Irving 1998) However, in the case of complex biological samples, reasonable definition of the integral regions is often impossible. (Barding, Salditos, & Larive 2012;Griffiths & Irving 1998) Also software for batch integration and analysis of large NMR datasets has been introduced. (Mäkelä et al. 2011) Other methods, used for area determination, are binning and deconvolution. Also qQMSA (Manuscript V) is a deconvolution method which can be seen as an ultimate example of previously developed constrained total-line-shape (CTLS) approach (Soininen et al. 2005). These are discussed in more detail later on (see chapter 1.5.3)

qNMR experiments require both reference and calibration standards, which are used to reference chemical shifts and to the absolute quantification of the NMR signals, respectively. The absolute concentration of a component can be determined by comparing the signal area of component and the signal area of calibration standard which concentration is known. For some experiments, relative differences in the concentrations of compounds in two sample groups (e.g. control versus treated) may be more useful than the absolute concentrations and, as a result, determination of the absolute concentrations and the use of calibration standard are not needed. While TMS and DSS are the IUPAC-approved (International Union of Pure and Applied Chemistry) internal NMR standards for referencing, and often used also as calibration standards, there are numerous compounds available as calibration and reference standards described in recent reviews. (Pauli et al. 2012;Rundlöf et al. 2010) Choice of a standard is case-specific and depends on several factors such as solubility, solvent and overlapping with other signals. For example, TSP is not a suitable calibration standard for biological samples due to its interactions with proteins. (de Graaf & Behar 2003) External reference capillaries, which demand calibration against known concentration, are used mainly in biological studies and in analyses where analyte contamination must be avoided.

ERETIC (Electronic REference To access *In vivo* Concentrations) has been introduced as an alternative approach to an internal standard as a way of determining absolute concentrations. (Akoka, Barantin, & Trierweiler 1999) It is based on a calibrated reference signal, which is not a real NMR line, but an NMR-like, electronically produced signal. The PULCON (PUlse Length based CONcentration determination) method was developed to measure protein concentration in the NMR sample, but it was shown to be suitable also for small molecules. (Wider & Dreier 2006) PULCON correlates the absolute intensities in two NMR spectra by the measurement of a precise 360° radio frequency pulse, and the concentration can be determined using an external reference sample.

1.4 THEORY OF NMR SPECTRAL ANALYSIS

The aim of the NMR spectral analysis is to extract the spectral parameters (chemical shift, coupling constant, line width and area of the signal) from the observed spectra for the needs of structural and quantitative applications of NMR. As mentioned earlier, the area of the signal is the most important parameter for qNMR, whereas the other parameters offer chemical confidence. Although good estimates of spectral parameters can often be obtained just by visually inspecting the observed spectrum as, based on the simple rules mentioned earlier, the chemical shifts are the midpoints of the patterns and the coupling constants are the observed splittings, this application is limited to first order spectra and the more complicated spin systems can only be analysed by QM methods.

1.4.1 Nomenclature for spin systems

The notion spin system is used for a group of nuclei that interact magnetically with each other (i.e. are coupled). The relative chemical shifts of different nuclei in a spin system are indicated by the position in the alphabet of the labelling letter. Thus, nuclei with similar chemical shifts are given letters A, B, C, etc. and, if there are also interacting nuclei with different chemical shifts, then these are given letters X, Y, Z. Nuclei with identical chemical shifts (chemical equivalence) are denoted by a subscript. Using above mentioned nomenclature, the protons of ethyl group form an A_3B_2 system while for a CH_3CF_2-group the designation A_3X_2 is used to indicate the large difference between the chemical shifts of the protons and the fluorine nuclei. Two nuclei are magnetically equivalent if they have identical chemical shifts and they couple equally to the every other nucleus in the spin system. Chemically equivalent, but magnetically non-equivalent nuclei are distinguished by using primed letters. Thus, the two protons and the two fluorine nuclei in 1,1-difluoroethylene and the two pairs of protons in 1,2-dichlorobenzene are an AA'XX' system and an AA'BB' system, respectively.

1.4.2 The quantum mechanics of NMR in brief

Quantum mechanics is needed in order to describe, understand and devise NMR phenomena, experiments, and especially the principles of the quantum mechanical spectral analysis (QMSA) properly. In the following a short introduction to the quantum mechanics of NMR is given.

The Hamiltonian operator, H, measures the energy (E_n) of the various states (Ψ_n) of the system by the famous Schrödinger equation

$$H\Psi_n = E\Psi_n \tag{1.14}$$

The Hamiltonian is important because its *eigenvalues* and *eigenfunctions* are the energy levels of the system, and it is transitions between these energy levels which are detected in NMR. The Hamiltonian does not simply determine the energy levels in NMR, but it also affects how the spin system evolves in time, which is the keystone of multiple-pulse NMR. The idea of matrix representation of the theory (density matrix theory) forms a convenient framework for calculating the outcome of NMR experiments and spectral analysis. (Abraham 1971;Keeler 2005)

Coupled nuclear spins form a system which can be treated within experimental accuracy using QM theory. In the case of AX system, the spin states of the two-spin system can be completely described by wave functions $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, and $\beta\beta$ that are products of wave functions describing the individual nuclei. If the A and X nuclei are coupled, the indirect interaction (via electrons) of their z-component of spin magnetic moments leads to the regular doublets and the spectrum can be analysed also manually as a first-order spectrum. However, if the nuclei are strongly coupled (the chemical shift difference between the two nuclei is not more than 10 times larger than coupling), the xy-components of spin magnetic

moments start precess with each other which leads to the mixing of the $\alpha\beta$ and $\beta\alpha$ states. In quantum mechanics this can be described by forming new wave functions that are linear combinations of the original ones: the coefficients c_{ij} in $c_{11}\alpha\beta + c_{12}\beta\alpha$ and $c_{21}\alpha\beta + c_{22}\beta\alpha$ are selected so that the two functions are orthogonal. This mixing of the states leads to second-order effects, which appear in spectra as changes of intensities and splittings (Figure 7). In symmetrical systems, like para-disubstituted benzenes with AA'BB' system with chemically equivalent nuclei, the interaction leads to completely mixed wave functions $[\alpha\beta + \beta\alpha]/2^{½}$ and $[\alpha\beta - \beta\alpha]/2^{½}$ and to NMR multiplets which can be converted into coupling information only by using complex rules or by computerised analysis. (Abraham 1971;Publication IV)

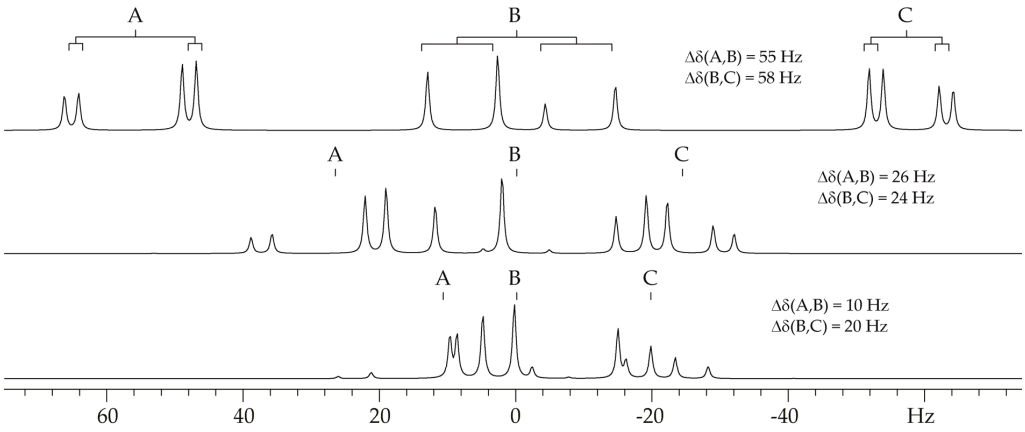

Figure 7. Three theoretical spectra of an ABC spin-system calculated with the same coupling constants but different chemical shifts. The coupling constants (J_{AB}=17.38 Hz, J_{AC}=1.79 Hz, and J_{BC}=10.46 Hz) and the relative chemical shifts in the lowest spectrum correspond to the parameters of acrylic acid. First-order coupling trees are shown in the uppermost spectrum that is not anymore that strongly second-order spectrum although leaning of the signals is still apparent. The chemical shifts are indicated with capital letters.

1.4.3 Computerised spectral analysis

The chemical shifts of multiplets, excluding strongly second-order ones, can usually be estimated manually with a precision better than 0.01 ppm using the centres of gravity of the multiplets. This is sufficient for most of the structural applications of NMR and in such applications very accurate chemical shifts are not the reason to use computerised analysis. Instead, in applications of qNMR, high accuracy shifts (even better than 0.0001 ppm) can be useful, specially, in the case of intense spectral overlapping. Unlike chemical shifts, coupling constants are relatively insensitive to the sample conditions and molecular tertiary structure. Thus, accurate coupling constants can be used to identify, for example, a certain type of aromatic fragment even in mixtures (Hanhineva et al. 2009).

One reason for computerised analysis is that, when the spectrum is complex or second-order type or the spin-system owns symmetry, it may even be impossible to determine the coupling constants directly from the spectrum without computerised analysis. In such cases, computerised spectral analysis helps to explain spectral details and to reveal coupling information which is not obtained or understood on the basis of manual analysis. Whenever high similarity between the observed and quantum mechanically calculated spectra is obtained, it is a strong argument for the consistency of the spectral assignments. In addition, even when it is possible, the manual analysis of complex spectra is a time-demanding and subjective task.

Modern NMR technology allows automatic measurement of NMR spectra. Computerised analysis offers a fast way to obtain accurate, unbiased values of parameters to be reported, archived, and used in structure verification. Quantum mechanically derived spectral parameters are the most efficient way to storage spectral data (see chapter 4.1). NMR based automatic structure verification, based on consistency of observed and predicted spectral parameters, is one of the novel applications of computerised spectral analysis and will be discussed more detailed later on.

The principle of computerised spectral analysis is always the same. As the first step, some trial parameters are obtained from the spectrum by manual analysis or, nowadays, if a good guess of the structure exists, by prediction of the spectral parameters. The trial parameters are used to simulate the trial spectrum based on quantum mechanics, and then in the iterative phase the trial parameters are adjusted iteratively in order to minimise the residual sum of the squares (*SQ*) between the observed and calculated spectral observables *O(i)*

$$SQ = \sum \left(O_{obs}(i) - O_{cal}(i) \right)^2 \tag{1.15}$$

The problem is more or less non-linear, depending on the type of the observables and the second-order nature of the spectrum.

The QMSA protocols can be divided into two main classes. In the frequency based methods, the observables *O(i)* are transition frequencies or spectral maxima (peak-tops) which may contain contributions from many transitions. In the Total-Line-Shape (TLS) methods the observables *O(i)* are spectral intensities or numbers derived from them. (Publication IV)

Frequency-based methods

In the earliest frequency based methods LAOCOON and NMRIT/NMREN, the observables *O(i)* were the positions (frequencies) of NMR signals (Castellano & Bothner-By 1964;Swalen & Reilly 1962). In PERCHit iterator, a peak-top fitting algorithm was developed in order to obtain accurate non-biased couplings for the cases where NMR signals are composed of many degenerate transitions. (Laatikainen 1986) In practise, the newer TLS methods are superior as to the convenience and usability, with a couple of exceptions.

The first type of exception is the spectral analysis of strongly resolution-enhanced spectra. One example of such peak-top-fitting analysis is the analysis of ^{1}H spectrum of naphthalene. (Laatikainen 1988) Its AA'A"A"'BB'B"B"' system yields ca. 800 observable lines but only ca. 300 peak-tops are visible even after heavy resolution enhancement which, unfortunately, distorts the line intensities.

The second type of exception is where some essential spectral information on coupling constants is carried by very weak signals that are not easy to separate from baseline artefacts and ^{13}C satellites. In such cases, the peak-top frequency information can be used as an extra piece of information in a TLS analysis. As a common example, analysis of the couplings of $-CH_2-CH_2-$ fragments requires that the very weak satellite lines are included into the TLS analysis. (Publication IV)

Total-Line-Shape (TLS) iterators

The idea of TLS analysis was introduced to QMSA by Heinzer (Heinzer 1977) and the principle has been developed in a number of programs. (Diehl, Sýkora, & Vogt 1975;Golotvin & Chertkov 1997;Hägele, Engelhardt, & Boenigk 1987;Laatikainen et al. 1993;Stephenson & Binsch 1980;Weber & Thiele 1998;Zubkov, Golotvin, & Chertkov 2002) The principle of integral transforms (IT) was introduced by Diehl (Diehl, Sýkora, & Vogt 1975) and later further developed by Stephenson and Binsch in the program DAVINS. (Stephenson & Binsch 1980) The basic idea of the IT-approach in PERCHit (Laatikainen et

al. 1996b) is similar to the programs DAVINS, DAISY (Hägele, Engelhardt, & Boenigk 1987) and WIN-DAISY (Weber & Thiele 1998).

In TLS fitting the objective is to minimise the difference between the observed and calculated spectral intensity information: $O(i)$ in equation (1.15) can be the intensity $I_i(\nu)$ at a given spectral point or an integral (bin) derived from the spectrum. In general, the intensity of an NMR spectrum $I(\nu)$ is the sum of the spectra S_n of the chemical components, baseline $B(\nu)$ and noise:

$$I(\nu) = \sum x_n S_n(\nu) + B(\nu) + noise(\nu) \tag{1.16}$$

where x_n is the population of the component n. Each spectrum $S_n(\nu)$ can be written as a function of spectral parameters:

$$S_n(\nu) = F_n(\nu, W, J, R, \Delta, lineshape) \tag{1.17}$$

where W, J and Δ are the vectors for chemical shifts, coupling constants and line widths, which may be different for each species. The response factors R accounts for the small differences caused by different relaxation in the signal integrals of different species and they should ideally be close to 1.0. The line shape parameters can be assumed to be the same for all species in the same spectrum and can be expressed as a sum of Lorentzian, Gaussian and dispersion terms. (Laatikainen et al. 1996a) The baseline $B(\nu)$ can be described with a Fourier expansion that can also be optimised during the iteration. While fitting small signals located close to a strong signal, one can include also a few polynomial terms that can be used to describe the contribution of the strong signal to the effective baseline. (Soininen et al. 2005)

The TLS fitting is a very strongly nonlinear problem, and demands calculation of numerous partial derivatives $\partial I_n(\nu)/\partial P$ with respect to different parameters P. If the calculated and observed spectra do not overlap at all, a situation known as the basic problem of TLS method, all partial derivatives are zero at the points where the observed spectrum has intensity and this is why the derivatives do not offer any information for the iteration. However, if the lines are broadened artificially, the derivatives become non-zero and the iteration starts converging. Figure 8 illustrates this problem and represents integral transforms based solution to it. In PERCHit triangular (Bartlett) window function (Bartlett 1950) is used in forming the ITs, which lead effectively to broadening and also packing of the spectral information, and for which the calculations are fast. An essential tool in the protocol is Principal Component Analysis (PCA), which is used to automatically find the spectral parameters that can be optimised with at a given broadening, and to identify cases in which the spectrum is not sensitive to some spectral parameter. At the beginning of the iteration, all the spectral details arising from coupling constants are removed using large broadening and only the chemical shifts and the sums of couplings that have a significant effect on the appearance of the calculated spectrum will be adjusted. After the shifts and sums of couplings are optimised, the broadening is decreased so that large couplings (or in symmetric systems those combinations of couplings that have a clear effect on the spectral appearance) become adjustable. The process is continued until the broadening is of the same magnitude as the line width. After this the very final iteration can be done without any broadening and including both the line widths and line shape in the analysis.

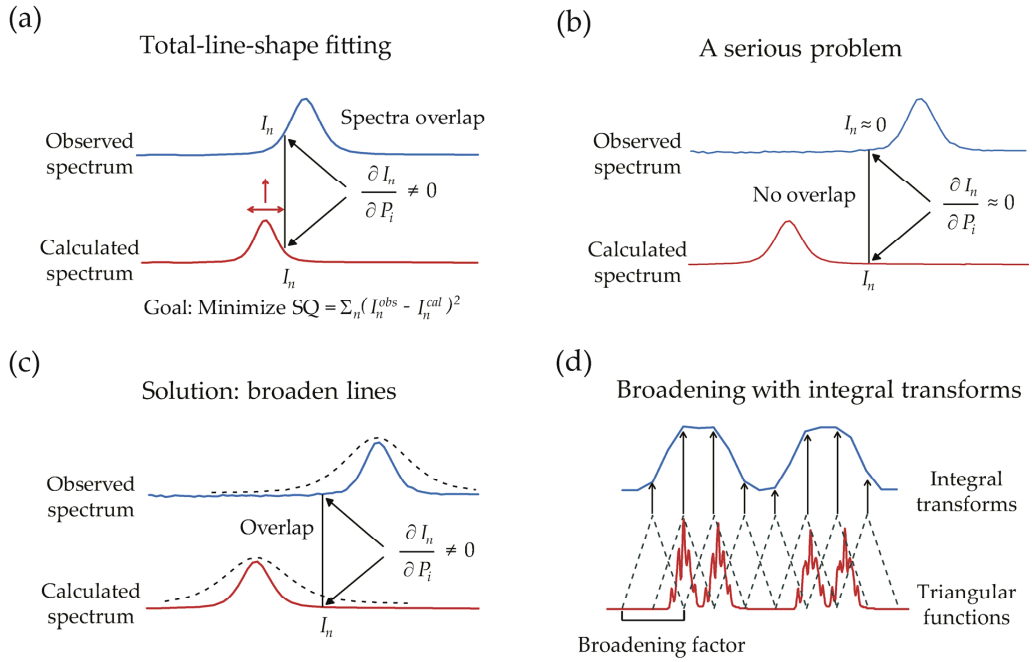

Figure 8. Integral transforms based solution to the basic problem of total line shape fitting. (a) Observed and calculated lines overlap and the derivative ∂I/∂P (I=intensity, P=frequency of the signal) is non-zero, and, thus, can be used to estimate in which direction the calculated signal must be moved in order to minimise difference. (b) The signals do not overlap, and because ∂I/∂P=0 the direction of calculated signal cannot be defined. (c) However, if the lines are artificially broadened (dashed lines), the derivative with respect to P becomes non-zero. (d) Artificial broadening can be obtained by multiplying the spectrum with triangular window functions (dashed line triangles) and then integrating the product functions.

The convergence of the IT iterations depends on the complexity of the system. In the case of ABC system there are three multiplets, at positions corresponding to the three chemical shifts and with multiplet widths that define the three couplings. Thus, the spectrum provides six well-defined "features" to be used in the determination of six spectral parameters and, therefore, the IT method allows automatic analysis of the ABC system. In practise, the analysis of ABCD system (>8 features/10 spectral parameters) and even larger systems usually succeeds without any prior knowledge about the spectral parameters if the spectra are not so strongly second-order that some essential information is carried only by the weak lines (typical of pro-chiral CH_2 proton signals) which are not recognised by the TLS criteria. Unfortunately, when the number of spins and non-zero couplings increase, the convergence of the iteration to the correct solution is far from sure, especially, if the trial parameters do not bear a fair similarity to the correct ones.

1.4.4 Large systems

As mentioned above, the Schrödinger equation for the nuclear spin system can be written in the matrix form. This QM formalism leads to Hamiltonian matrices (Abraham 1971;Corio & Smith 2007), which can be diagonalised in order to obtain the *eigenvalues* and *eigenvectors* for the calculation of transition frequencies, intensities and derivatives in the iterative calculations. Unfortunately, the sizes of the matrices, the number of transitions and evidently the demand of computer resources grow quickly when the spin system grows. Fortunately, the spin system can be almost always simplified by using X-approximation,

symmetry, and dividing the system into subsystems so that QM calculations are realisable for almost any molecular system.

X-Approximation (Abraham 1971;Corio & Smith 2007) can be used to split the Hamiltonian matrices into smaller ones. If the "X-factor" = $|(H_{ii} - H_{jj})/H_{ij}|$ (where H_{ij}, H_{ii}, and H_{jj} are Hamiltonian matrix elements) is large enough, the spin states i and j are not mixed significantly and the off-diagonal element H_{ij} can be ignored. In the case of ABX system, those off-diagonal elements of the Hamiltonian matrix that occur between states with different magnetic quantum numbers m(X) of the X nucleus are negligibly small compared with the diagonal elements and can be ignored. Setting the "X-factor" bigger than 200 yields almost unbiased spectral intensities and with smaller values some bias may arise but a fair spectral analysis is usually possible if the factor is kept over 10.

In practise, the largest spin-½-system that can be calculated accurately without X-approximation and within reasonable efforts is the spin-system of 12 protons, which are all coupled with each other, with the largest Hamiltonian of 924 x 924 matrix and each 12 species giving 4 096 transitions, which overlap so strongly that no single lines can detected. In many cases, such as peptides and carbohydrates, the spin-system can be divided into sub-systems, which can be treated independently. Thus, for example, the 89 protons of the cyclic peptide Cyclosporin A yield 55 spin-particles in 11 sub-systems and totally ca. 8 150 non-degenerate transitions. Large spin networks such as steroids (Figure 9) can be treated by dividing the spin network into partly overlapping sub-systems. Each sub-system is composed of "primary" species (for which the transitions are calculated through this spin sub-system) and "secondary" species (transitions not calculated) which are included in the calculation because of their couplings to the "primary" species. Any "tertiary" species that are strongly coupled and mixed with the "secondary" species may couple virtually to the "primary" species and thus need to be included into the sub-system. Virtual coupling can be observed if one nucleus of a set of strongly coupled nuclei is additionally coupled to a third nucleus with a very different resonance frequency. Thus, for example, the multiplicity of the X portion of an ABX system can be higher than a simple first order approach suggests. The above described protocol is used in PERCHit (Laatikainen et al. 1996b) and it is analogous to that described recently by Castillo *et al*. (Castillo, Patiny, & Wist 2011)

Symmetry also simplifies the spin systems effectively: CH_3 and homotopic CH_2 protons can be treated as one spin particle so that, for example, the 12 spin states of the five spin-system of CH_3CH_2 give only 14 nondegenerate lines, which reduce to 6 lines if the chemical shifts differ enough, instead of the 80 lines that theoretically would arise from an ABCDE system. Twofold symmetry with both magnetic and chemical equivalence, used originally in the program NUMARIT, (Quirt & Martin 1971) is usually adequate for efficient analysis of common spin systems. A good point in twofold symmetry is that the couplings can be presented in unsymmetrical tables (setting $J_{ij} \neq J_{ji}$). If the system has a higher symmetry, for example, like that of naphthalene (AA'A"A'"BB'B"B"') and butane ($A_3A_3'B_2B_2'$), some parameters must be kept equal during the iteration. For example, in the case of butane the chemical shifts of the B_2 and B_2' must be kept equal because otherwise they can diverge during the iteration. (Tynkkynen et al. 2012)

Although computerised spectral analysis is already at a level allowing automation, understanding of a few special points, such as the consequences of the magnetic and chemical equivalence, is needed in critical work. One should also be able to recognise cases where spectral parameters cannot be determined from the spectrum at all because the spectrum simply does not depend on them. Multiple solutions are typical for strongly second-order spectra and symmetric systems, and it is important that the analyst is aware of this potential problem. Additional challenges are formed by signs of couplings and long-range coupling constants which may have significant effects of outlook of the second-order spectrum.

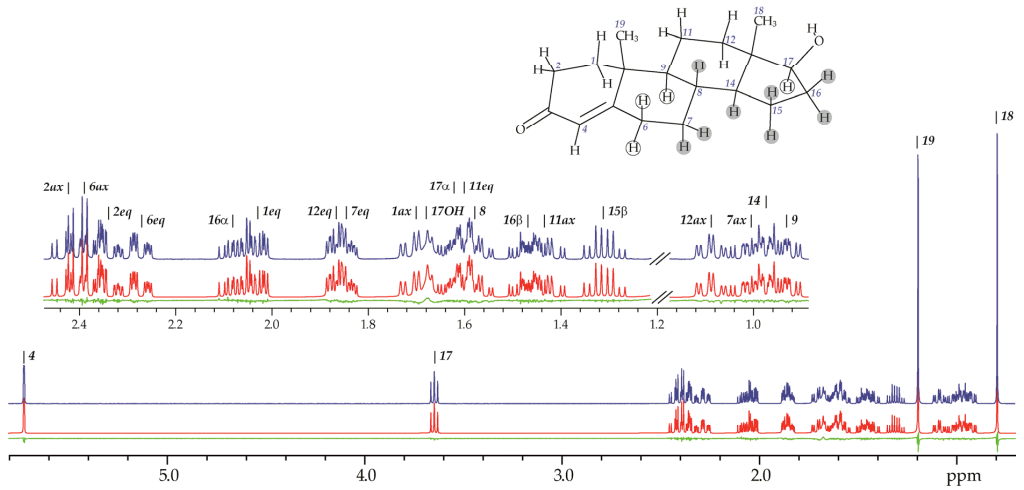

Figure 9. The structure of testosterone and its 500 MHz ^{1}H NMR spectrum in chloroform (blue), the simulated spectrum (red), and the difference between the observed and the simulated spectra (green). The spin network of testosterone can be divided into 7 partly overlapping sub-systems. For example, the transitions of protons 7ax, 7eq, 8, 14ax, 15α, 15β, 16α, and 16β are calculated from the 12 spin system formed by protons 7ax, 7eq, 8, 14ax, 15α, 15β, 16α, 16β, 6eq, 6ax, 9, and 17 where the 6eq, 6ax, 9, and 17 protons (open circles in the structure) are "secondary" species which are included in the calculations because of their couplings to the "primary" species (shaded circles in the structure).

1.4.5 Automated structure elucidation and verification

NMR based automatic structure elucidation and verification are novel applications of computerised spectral analysis. Structure elucidation is the *de novo* determination of a structure, whereas structure verification means the confirmation of a postulated structure. Structure elucidation is used, for example, to the identification of only just isolated natural products of unknown structure, to the analysis of a product or by-product of a chemical reaction, and to the characterisation of chemical degradation products and metabolites of drugs. Structure verification is applied to cases, for example, in combinatorial chemistry and parallel synthesis, where one or more structures are hypothesised and experimental spectra are used to confirm one hypothesis and/or reject others. (Elyashberg, Williams, & Martin 2008) In the following text, short introductions to structure elucidation and verification are given; verification being discussed in more detail.

Structure elucidation

In structure elucidation, information from several types of spectra is commonly used. The molecular formula of the substance is generally obtained from a mass spectrum and structural hypotheses are deduced from other spectral data such as NMR, IR, and UV spectra. NMR spectra are considered as the primary source of structural information since from them the environment of magnetically active nuclei can be revealed through the chemical shifts and spin-spin couplings. (Elyashberg et al. 2009)

The structure elucidation can be subdivided into data acquisition, data interpretation, and structure generation and evaluation. (Dunkel 2007) Data acquisition part of this process can often be automated and, for example, modern NMR technology allows fully automatic measurement of NMR spectra. The rest of the process forms a whole which can be referred to as Computer-Aided Structure Elucidation (CASE). The running of CASE can be simplified to a two-step process: 1) generation of all possible isomers from a given molecular formula; 2) testing of different structural constraints one by one until there is

only one possible isomer left. Actually this approach can be used only for small molecules (not more than 15 skeletal atoms), because in the case of larger molecules, this will lead to a situation known as a "combinatorial explosion". This problem is solved by using molecular fragments that account for a significant number of the skeletal atoms and, thus, reduce the dimension of the problem. (Blinov et al. 2003;Elyashberg et al. 2009)

The Structure Elucidator software based on NMR spectral data was used to elucidate successfully more than 300 complex compounds from which more than 100 were large molecules with between 30 to 106 skeletal atoms. In this program, the molecular formula or molecular mass of analysed compound, HSQC, heteronuclear multible bond correlation (HMBC), and COSY spectra, as well as 1D ^{13}C and ^{1}H spectra were used as initial data. (Elyashberg et al. 2009)

Structure verification

Structure verification differs from structure elucidation in that there is a proposed and therefore kind of known structure. Thus, the bottleneck of structure elucidation, structure generation, can be avoided and, additionally, the proposed structure can be used to predict the properties of molecule, for example, chemical shifts. Consequently, structure verification addresses questions such as "How well does the measured spectrum correspond to the given structure?" or "Is the spectrum consistent with the given structure?".

One dimensional proton spectroscopy is the most widely used method in small molecule structure confirmation. Structure verification methods based on the comparison of experimental, obtained from ^{1}H NMR spectra, and predicted chemical shifts, integrals, and multiplicities for proposed structures have been published. (Golotvin et al. 2006;Griffiths & Bright 2002) However, the success of these methods depends heavily on the accuracy of the chemical shift prediction and, additionally, the presence of labile proton peaks limits the performance of these methods. Improvement in verification has been gained by using ^{1}H-^{13}C HSQC spectrum or its variants together with the 1D proton spectrum. (Elyashberg et al. 2009;Golotvin et al. 2007;Griffiths, Beeley, & Horton 2008) Also, concurrent combined verification (CCV) in which combined verification (HSQC spectrum together with 1D proton spectrum) is performed simultaneously on a presumed correct structure and alternative incorrect structures has been introduced and evaluated. It was found that the CCV reduces the pass rates of analyses but, on the other hand, it identifies the incorrect structures (i.e. false positives) better than the standard combined verification. (Golotvin et al. 2012) The limited throughput of 2D NMR relative to 1D NMR is the reason why 1D NMR together with other analytical techniques such as mass spectrometry (MS) has remained as the primary tool in pharmaceutical industry for validating combinatorial libraries. Use of Hadamard encoding, which eliminates the extensive and uniform sampling in the indirect dimension, has been proposed as a partial solution to this problem. (Ruan et al. 2009)

Also QM analysis based structure verification has been proposed. Automated consistency analysis (ACA) performs an iterative QM analysis in full automation. In contrast to the previous approaches, structure verification based on QMSA can extract accurate chemical shifts and coupling constants even for overlapping and higher order multiplets providing more reliable results. The 96 different pyrrole derivates were analysed using ACA and, for 86% of the tested samples with impurities less than 20%, complete analysis could be successfully automated yielding fully assigned spectra. (Thiele et al. 2011)

1.5 NMR IN METABOLOMICS

High-throughput metabolic profiling focuses on the characterisation, identification, and quantification of metabolites present in biological samples such as serum, urine, cerebrospinal fluid, faeces or tissue extracts. (Beckonert et al. 2007;Jukarainen et al. 2008;Nicholson et al. 2011;Psychogios et al. 2011;Soininen et al. 2009;Suhre et al. 2011;Wu et al. 2010) Metabolomics describes the physiological endpoint, which originates from the interplay of all regulatory and enzymatic processes in the system under research at a given time, providing snapshots of the molecular machinery in action. Thus, metabolomics is complementary to other "omics" disciplines.

The most commonly employed analytical tools, NMR and MS, have been used extensively in metabolomics. MS combined with chromatography, e.g. gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS), provides high sensitivity for target compound analysis. On the other hand, NMR spectroscopy has a unique potential since it is non-destructive and highly reproducible and it does not require potentially labour-intensive and costly physical separation of the components. (Cui et al. 2008;Zhang et al. 2008) Furthermore, NMR is quantitative whereas MS is only semi-quantitative by nature.

1.5.1 Database types used in metabolomics

A database is a collection of data which is organised so that it can easily be accessed, managed, and updated. As other 'omics' disciplines, also metabolomics is highly dependent on the availability and quality of electronic databases. Additionally, because metabolomics combines molecular biology with chemistry and physiology, there is a need for a wide variety of electronic resources providing various data from chemical structures to linked metabolic pathways. (Go 2010;Kastenmüller et al. 2011;Mendes 2002;Wishart et al. 2009)

In the recent years, advances in NMR spectroscopy and MS have allowed quantification of hundreds of metabolites in biological samples in a high-throughput manner. With this vast amount of metabolomic information harvested using high-throughput techniques, new types of metabolomic databases are emerging. These databases are not only designed to store, manage, and analyse metabolomic data but are also meant to serve as gateways to the broad information space of metabolism. Data housed in these databases, which are freely or commercially available, cover the wide-spectrum of research being done in the metabolomics field providing information on the chemical structures, physiochemical and pharmacological properties, spectral profiles (spectra), experimental workflows, and biological functions of metabolites. (Go 2010;Kastenmüller et al. 2011)

At present, there are at least five different types of databases used in metabolomics (Table 3). These include: (i) metabolic pathway databases, (ii) compound-specific databases, (iii) spectral databases, (iv) disease (physiology) databases, and (v) comprehensive, organism-specific metabolomic databases. (Wishart et al. 2009) This or any other categorisation is not that tight because, naturally, there is some overlap in content of these database types. Reactome database (Joshi-Tope et al. 2005) and KEGG database (Kanehisa et al. 2006) are examples of the metabolic pathway databases that contain metabolic pathways with general metabolite information. Compound specific databases such as PubChem (Wheeler et al. 2006), Lipid Maps (Fahy et al. 2007) and DrugBank (Wishart et al. 2006) provide detailed nomenclature, structural or physicochemical data on restricted classes of compounds. The Madison Metabolomics Consortium Database (MMCD) (Cui et al. 2008), the BioMagResBank (BMRB) (Ulrich et al. 2008) and the Golm Metabolome Database (GMD) (Kopka et al. 2005) are examples of spectral databases that contain reference NMR and MS spectra used in metabolomics. These databases include also software for identification of compounds by spectral matching. Disease databases such as Online Mendelian Inheritance in Man (OMIM) (Hamosh et al. 2005) and The Online

Metabolic and Molecular Bases of Inherited Disease (OMMBID) (http://www.ommbid.com) contain descriptions of the causes, clinical symptoms, diagnostic indicators or genetic mutations related to diseases. The Human Metabolome Database (HMDB) (Wishart et al. 2007) and the database SYSTems biology of pseudOMONAS (SYSTONOMAS) (Choi et al. 2007) are examples of comprehensive, organism-specific metabolomic databases, which attempt to combine all of the information from most of the former types of databases.

Table 3. Representative examples of databases used in metabolomics.

Database	Content
(i) Metabolic pathway database	
KEGG	Metabolic and regulatory pathways, protein-protein interactions, genes, drugs, glycans, small molecules, reactions, functional hierarchies.
Reactome	Metabolic pathways, reactions, proteins, nucleic acids, small molecules, macromolecular complexes, links to other databases.
(ii) Compound-spesific database	
DrugBank	Drugs, drug targets, chemical, pharmacological and pharmaceutical data, 3D and chemical structure, links to other databases.
Lipid Maps	Lipids, chemical structure and formula, physical properties, lipid-related proteins and genes, links to other databases, experimental data and protocols.
Pubchem	Small organic compounds, chemical structures, bioactivity data, links to *Entrez* databases.
(iii) Spectral Database	
BMRB	NMR, spectra, quantitative data, proteins, peptides, nucleic acids and other biomolecules, experimental data.
GMD	GC-MS, spectra, retention times, metabolite profiles, experimental methods and protocols
MMCD	NMR, LC-MS, spectra, small molecules, chemical structure, physical and chemical properties, links to other databases.
(iv) Disease database	
OMIM	Human genes and genetic phenotypes, relationship between phenotype and genotype.
OMMBID	Genetics, metabolism, diagnosis and treatment of metabolic disorders, pathways, chemical structures, physiological data.
(v) Comprehensive, organism-specific database	
HMDB	NMR, GC-MS and MS/MS data, small molecule metabolites found in the human body, chemical, clinical and molecular biology/biochemistry data, experimental methods, links to other databases.
SYSTONOMAS	Pseudomanas species, transcriptomic, proteomic and metabolomic data, reconstructions of metabolic networks.

1.5.2 NMR spectra databases

In metabolomics, most compounds are identified by comparing own, experimental spectrum against spectral library (known compound spectra), just as most genes and proteins are identified via sequence comparisons in genomics and proteomics. (Wishart et al. 2009) Especially automatic identification and quantification of metabolites in NMR spectra are dependent upon spectral libraries. Meaningful interpretation of metabolomics experiments results is possible only in a specific experimental context that needs to be known. Thus, data unification and standardisation are critical for facilitating the reporting and archiving of NMR spectra in public databases, as well enabling efficient querying of data. Standards should provide an unambiguous description of the metabolite measurements and associated metadata. (Fiehn et al. 2007;Rubtsov et al. 2007;Sumner et al. 2007)

Spectra database can be stored as a complete self-contained data set or as an online repository that can be accessed and searched remotely. Examples of NMR databases with short describtions of content are given in Table 4. NMRPredict (Modgraph), ACD (ACD/labs), BBIOREFCODE (Bruker Biospin) and Chenomx (Chenomx) are examples of offline NMR databases whose use require a license. BioMagResBank (Ulrich et al. 2008), Birmingham Metabolite Library (BML-NMR) (Ludwig et al. 2012), Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do), a database of Refined solution NMR structures (DRESS) (Nabuurs et al. 2004), MMCD (Cui et al. 2008), HMDB (Wishart et al. 2007), Spectral Database for Organic Compounds (SDBS) (National Institute of Advanced Industrial Science and Technology, Japan) and NMRShiftDB (http://nmrshiftdb.nmr.uni-koeln.de/) are examples of online databases that are freely available. The form in which the spectral data is stored ranges widely from line lists that can be graphically displayed to raw FID data.

NMR parameters (chemical shift, coupling constant, line width and area of the signal) obtained by computerised spectral analysis offer an efficient way to store spectral information. For example, even the finest details of spectra containing tens of thousands spectral points, measured at one magnetic field strength, can be compressed into a few numbers that can then be used to simulate the spectrum at any magnetic field strength and with any line shape. Even the sample pH can be taken into account by measuring the spectra for the libraries at different pH values and then using interpolation to the wanted pH value. This is the principle of adaptive spectral library (ASL). (Publication I) There are already commercial applications based on quantum mechanically prepared libraries such as the BBIOREFCODE, which contains spectra and structures of metabolites, offered by Bruker Biospin.

Table 4. Examples of NMR databases.

Database	Content
(i) Offline NMR databases	
ACD	Training databases of the ACD/NMR predictors, nuclei (chemical shifts): ^{1}H (1.7 million), ^{13}C (2.5 million), ^{15}N (ca. 22 000), ^{19}F (ca. 35 000), ^{31}P (ca. 34 000).
BBIOREFCODE	Bruker's reference compound database, up to 600 metabolites measured at pH's from 3 to 8 at intervals of 0.5 pH units.
Chenomx	^{1}H database of ca. 300 small molecules, chemical shifts modelled as a function of pH.
NMRPredict	Ca. 465 000 records, ca. 6.3 million chemical shifts in total, nuclei: ^{1}H, ^{13}C, ^{19}F, ^{31}P, ^{11}B, ^{15}N, ^{17}O, ^{29}Si.
(ii) Online NMR databases	
BioMagResBank	Repository for data from NMR spectroscopy on proteins, peptides, nucleic acids, and other biomolecules, chemical shifts: protein ca. 5.2 million, DNA ca 27 000, RNA ca. 43 000.
HMBD	905 compounds with experimental ^{1}H NMR spectra and 899 with experimental ^{13}C spectra, also HSQC data.
NMRShiftDB	Database for organic structures and their NMR spectra, peer-reviewed submission of datasets, ca. 49 000 observed spectra.
SDBS	Integrated spectral database for organic compounds, ca. 15 000 ^{1}H NMR and ca. 14 000 ^{13}C NMR spectra.

1.5.3 From NMR spectra to composition

Determining the chemical composition of biological samples is a main focus of systems biology and metabolic profiling. Comprehensive studies of these complex mixtures require reliable, efficient, and automatable methods, which are used to identify and quantify the underlying metabolites and natural products. Because of its rich structural information content and quantitative nature, NMR spectroscopy has a unique potential for this task. (Zhang et al. 2008)

The sensitivity of NMR parameters to the chemical environment and physical properties of sample create challenges; slight differences in the sample conditions such as pH, ionic strength, and temperature or protein content between samples will cause differences in the spectra of metabolites. Differences in the properties of biological samples, for example ionic strength in urine samples, can cause peaks to shift as much as 0.1 ppm even in the case of pH buffered samples measured at controlled temperatures. In addition, metabolites are differently sensitive to the above-mentioned effects, and in many cases, the NMR resonances of a certain compound are affected independently. Furthermore, spectral overlapping in certain regions of the NMR spectrum complicates analysis of these complex mixtures.

A single compound can be quantified from the ^{1}H NMR spectrum of a complex mixture whenever even a single line of the compound can be seen, but even though the separation of the signals can be improved by expanding the spectrum to more than one dimension, the 2D NMR spectra are typically lower in either sensitivity or throughput. The use of selective total correlation spectroscopy (TOCSY) experiments (Ludwig et al. 2009;Sandusky, Appiah-

Amponsah, & Raftery 2011) is an alternative, but these experiments typically detect a few targeted metabolites. When the number of samples is large and their concentrations low, quantification based on a 1D proton spectrum is desirable.

There are two fundamentally different approaches to analyse NMR metabolomic data. In chemometric approaches (non-targeted metabolomics) pattern-recognition methods are used to analyse whole spectra, and only the metabolites that seem to be responsible for selected patterns are identified. (Ebbels & Cavill 2009;Trygg, Holmes, & Lundstedt 2006) Conversely, quantitative metabolomics (targeted metabolomics) approaches based on fitting or deconvolution methods are used to match the resonances to the known metabolites, and so the concentrations of known metabolites are directly quantified instead of some spectral areas. (Weljie et al. 2006;Wishart 2008) Both of these approaches have their own advantages, disadvantages and, of course, supporters. The major advantage of chemometric methods is that spectra can be analysed without the prior knowledge of metabolites present and that is why these methods can be used even for wholly novel compounds or sample materials. However, underlying any statistical treatment of NMR spectra in metabolomics is the basic notion that metabolites are the actual variables of interest, since they represent the underlying physical model that generated the observed data, and that is why quantitative data is much more valuable and multi-usable. Above all, the quantitative data does not require any spectroscopic knowledge in further analysis and it can be handled by any other specialist than spectroscopist, e.g. clinician or statistician, same way as the results of conventional laboratory tests.

Non-targeted metabolomics

The traditional method to overcome above mentioned difficulties caused by the sensitivity of NMR parameters to the sample conditions is a form of data reduction referred to as spectral binning. (Anthony et al. 1994) In this method, the spectrum is subdivided into a number of regions, bins, and the total areas of these bins are then used in further analysis. The assumption is that the small variations in compounds' spectral parameters between samples, especially chemical shifts, can be handled by using regions of spectra, bins, instead of individual data points. Binning itself is a simple and very fast method, and in addition, can also be automated. (Holmes et al. 1998) However, this method has its own drawbacks such as loss of information and the occurrence of artefacts caused by above mentioned peak shifts. Loss of information refers to lack of structural information in the resulting bins. For example, if there is a significant change in some low concentration compound concentration, this can be missed if there is some high concentration metabolite in the same bin (bins can contain peak/peaks from one or many compounds). Peak shift artefacts originate from that while uniform binning mitigates the effects from small variations of the peaks positions, shifts occurring near the boundaries of bins can cause dramatic quantitative changes in the adjacent bins due to the non-overlapping boundaries.

Problems arising from uniform binning have been tried to overcome by using more sophisticated binning methods such as non-equidistant binning, adaptive binning, adaptive intelligent binning and dynamic adaptive binning. Non-equidistant binning tries to take the width and the variability of peak shifts into account; an average spectrum of all the spectra in the study is calculated and the borders of the bins are defined by the five-point minima of the average spectrum. (Dieterle et al. 2006) Adaptive binning method uses the undecimated wavelet transform to smooth a reference spectrum and then the observed peaks and minima of the smoothed reference spectrum are used to dynamically bin the spectra. (Davis et al. 2007) Another dynamic binning method is the adaptive intelligent binning, which recursively identifies bin edges in the existing bins by dividing each existing bin into two new bins which are then evaluated according to specific criterion. (De Meyer et al. 2008) In the dynamic adaptive binning, bin boundaries are dynamically determined via dynamic programming by optimizing an objective function that measures the quality of the bin configuration. (Anderson et al. 2011) These methods were shown to be superior to the

traditional uniform binning technique, but on the other hand, data reduction can be done by more sophisticated methods, which are used in targeted metabolomics, than binning.

Targeted metabolomics

Since the area determination of overlapping NMR signals by routine integration or bucketing is problematic and produce inaccurate results, methods using line-fitting or, more commonly, deconvolution have been created. Deconvolution is based on peak fitting onto the observed spectrum by using a least-squares-based method. The initial values of peak parameters (frequency, width, height, and line shape) for line-fitting analysis can be obtained by using the prior knowledge, a spectral parameter library, calculations (prediction), or by performing a peak picking.

Methods using a database to acquire initial values needed for fitting are all based on the same principle: database contains model spectra of pure individual components recorded with certain parameters in certain conditions, and the observed spectrum of the mixture is reconstructed as a linear sum from the model spectra. Bruker AMIX (Bruker Biospin), LCModel (http://s-provencher.com/pages/lcmodel.shtml), and Chenomx NMR Suite (Chenomx) are examples of programs developed for this purpose. However, these programs suffer from the inflexibility of the model spectra, which means that no variation is allowed in the chemical shifts and coupling constants of a model compound spectrum. This can lead to errors in signal positions, because the frequencies of NMR signals depend on conditions, and thus, generate quantification errors. Other applications that fit model spectra to an observed spectrum, are, for example, weighted least-squares deconvolution method (Gipson et al. 2006) and a method based on linear least-squares fitting using singular value decomposition (Xu et al. 2006). The DemixC method (Zhang & Brüschweiler 2007) uses TOCSY combined with covariance NMR to deconvolute the mixture spectra into its components, thus allowing the quantification of overlapping signals. Automated targeted spectral profiling is an example of library-based method of using mathematically modelled reference spectra for quantification of metabolite concentrations in NMR mixture analysis. In this approach, metabolites are modelled using their peak centre and *J*-coupling information. (Mercier et al. 2011;Weljie et al. 2006) A fully automated annotation and quantification procedure is also proposed. (Alm et al. 2012) In this procedure, metabolites need to be annotated manually in one spectrum which is then used to build the reference database for similar samples (datasets with a similar matrix). Alignment that is needed to handle chemical shift variations between samples is based on the generalized fuzzy Hough transform and peak selection and quantification on PCA. However, none of these methods fully utilises the strict QM rules that exist between the spectral signals and, as a result, they cannot handle, for example, all the effects that variations in the sample conditions cause to the spectral parameters. Specially, in the case of strongly second-order spectrum, even small variations can have significant effects.

In the ASL based quantification (Figure 10), spectral parameters from QMSA of individual metabolites are used as a starting point of the QM mixture analysis. The qQMSA, which means the complete iterative analysis of the spectra based on the QM spectral model, offers an ideal tool for quantification of complex ^{1}H NMR spectra. qQMSA including models describing unknown components, background and prior knowledge from the sample enables modelling of even the smallest details of the spectrum and the maximal quantitative NMR information analysis. In addition to quantitative information, qQMSA offers chemical confidence, which means that individual components can be identified and quantified with a high confidence on the basis of their spectral parameters. (Manuscript V)

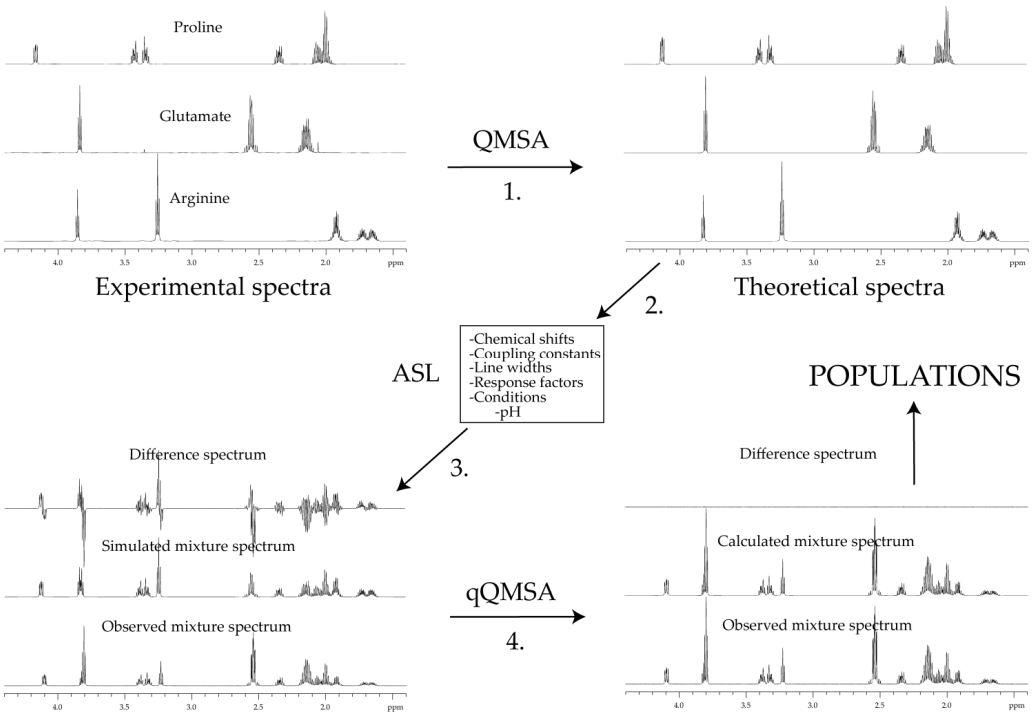

Figure 10. An illustration of the principle of ASL and qQMSA with a mixture of three amino acids. 1. Spectra of the pure compounds are measured and analysed (conditions, e.g. different pHs are also taken into account). 2. The obtained spectral parameters form the ASL. The spectra of the individual compounds simulated from these spectral parameters are free of instrumental artefacts and impurity signals. 3. The starting parameters of the QM mixture analysis are obtained from ASL (compounds and conditions). 4. Spectrum of the mixture is analysed using qQMSA, which gives populations (concentrations) of the individual compounds. In real case the number of the components may be even > 100 and the relative populations vary from 1–100%.

2 Aims of the study

General objective of the present study was to develop and test quantification methods, strategies and protocols for NMR spectra of different biological samples. The objective can be further divided to projects 1–3:

1. To develop adaptive spectral library principle including ^{1}H NMR pH indicators. This forms the basis for quantitative Quantum Mechanical Spectral Analysis. (Publications I and III)

2. To develop a protocol for quantification of amino acid ^{13}C isotopomers and determination of positional fractional ^{13}C enrichments for metabolic ^{13}C tracer experiments including rules for prediction of the ^{13}C isotope effects on ^{1}H chemical shifts. (Publications I and II)

3. To develop a quantitative Quantum Mechanical Spectral Analysis based quantification method and protocols for ^{1}H NMR spectra of complex mixtures. Human serum was used as the model system. (Publication IV and manuscript V)

3 Materials and methods

3.1 SAMPLES

3.1.1 Adaptive spectral library

Sample concentrations for ^{1}H NMR spectra were ca. 10 mg/ml and for ^{13}C spectra from 20 to 40 mg/ml of amino acid in D_2O. In addition, a mixture of amino acids with concentrations of individual amino acids between 1.2 and 3.1 mM was prepared. The pH^* of the samples was adjusted to 1.0 (pH^* refers to the glass electrode pH reading obtained without deuterium isotope correction). TSP was used as the internal reference. The experimental conditions were selected based on needs of the intended application of amino acid isotopomer analysis of protein hydrolysate.

3.1.2 Isotope shifts

The mixtures of ^{13}C-labelled and corresponding non-labelled compounds were prepared with different amounts (e.g. 4 and 6 mg) of non-labelled and labelled compound in question making the spectral analyses easier. The amino acids were dissolved into 0.1 M DCl (pH^* ca. 1) and D-glucose into D_2O. TSP was used as the internal chemical shift reference. The experimental conditions were selected based on needs of intended application of amino acid isotopomer analysis of protein hydrolysate.

3.1.3 pH indicators

To find out the pH-dependencies of the chemical shifts of the indicator molecule candidates, D_2O solutions including the indicator compounds were prepared. In these mixtures, the concentration of each indicator compound was ca. 2.8 mM. Trace amounts of TSP and DSS were used as internal shift references. The effects of ionic strength (0.15–0.45 M) on the chemical shifts were explored by preparing D_2O solutions containing also sodium, calcium, or magnesium chloride. The solutions were held in a water bath during the pH adjustment keeping the temperature constantly at 298.15 K. The pH^* values were determined using a pH meter (WTW Inolab pH 720) in equipped with a glass electrode (Hamilton Single Pore Glass). In preliminary studies, the NMR samples were taken at intervals of ca. 0.5 pH units, and for potential indicator compounds the pH adjustment was repeated collecting the samples at intervals of ca. 0.25 pH units.

To asses this approach with complex mixtures, indicator compounds (0.3–1.0 mM) together with TSP and DSS were added to two mixtures; one containing 1–4 mM amino acids, lactate and glucose and the other containing in addition to the above-mentioned also 0.1% human serum albumin. NMR samples were taken at intervals of ca. 0.5 pH units.

3.1.4 Artificial serum

Metabolite mixtures with physiological metabolite concentrations of human serum (Table 5) were prepared by using sodium phosphate buffer (75 mM Na_2HPO_4 in 80%/20% H_2O/D_2O, pH 7.4; including also 0.08% sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4 and 0.04% sodium azide) as a solvent. NMR samples with and without human serum background (pairs of parallel samples with the same metabolite concentrations) were prepared by adding 300 µl of the above-mentioned metabolite mixtures to 300 µl of dialysed serum or dialysis buffer (10 mM Na_2HPO_4, 150 mM NaCl and 2.7 mM KCl), respectively.

Table 5. The concentration (μM) ranges of metabolites in the metabolite mixtures.

Metabolite	Concentration range	Metabolite	Concentration range
3-Hydroxybutyrate	60–271	Glycine	169–286
Acetoacetate	24–118	Histidine	39–83
Alanine	312–491	Isoleucine	36–86
Citrate	56–130	Lactate	555–1680
Creatine	20–80	Leucine	64–126
Creatinine	42–93	Phenylalanine	46–100
Glucose	3 880–6 580	Pyruvate	35–99
Glutamine	361–586	Tyrosine	38–76
Glycerol	26–104	Valine	155–262

Dialyses cassettes (PIERCE, Slide-A-Lyzer Dialyses Cassette, 10 K molecular weight cut-off (MWCO)) were washed in 300 ml dialyses buffer (5 min magnetic stirring) to remove glycerol from the membrane. Five 3 ml aliquots of human serum from five different individuals were dialysed in three phases with magnetic stirring: 1) 1.5 l dialysis buffer, 1 h at room temperature; 2) 1.5 l dialysis buffer, 2 h at 4°C; 3) 1.5 l dialysis buffer, overnight (20 h) at 4°C. Protein concentrations of sera were measured before and after the dialyses using biuret method. After the dialyses, the sera were concentrated by ultrafiltration in order to obtain original protein concentrations (Figure 11). The centrifugal filters (VWR Centrifugal Filter, 10 K MWCO) were washed several times with water to remove glycerol from the filter membrane. Sera were filtered at 10 000 g and 4°C.

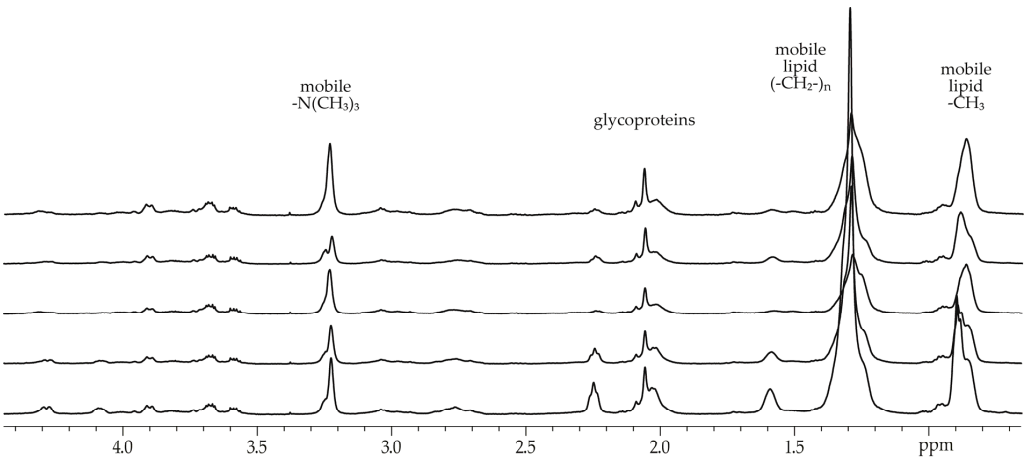

Figure 11. 500 MHz T_2-edited (T_2-filter time 80 ms) ^{1}H NMR spectra of five different serum backgrounds used in the artificial sera samples.

3.2 NMR MEASUREMENTS

3.2.1 Adaptive spectral library

^{1}H and ^{13}C NMR of samples were recorded on a Bruker Avance DRX 500 (Bruker, Karlsruhe, Germany) spectrometer operating at 500.13 and 125.77 MHz, respectively. Typical ^{1}H NMR spectra were acquired with a spectral width of 5 kHz, 64k data points and 16 scans. Typical ^{13}C spectra were measured with gated decoupling using a spectral width of 21 kHz with 256k data points and 8 192 scans. The ^{13}C-decoupled ^{1}H spectrum of acid-

hydrolysed yeast with 60% ^{13}C labelling degree was recorded on a Varian Inova 600 spectrometer operating at 599.91 MHz. Adiabatic WURST200 sequence was used for broad band ^{13}C-decoupling. The spectral width was 6 kHz, and 32k data points were recorded with 256 scans.

3.2.2 Isotope shifts

^{1}H and ^{13}C NMR spectra of samples were recorded on a Bruker Avance DRX 500 spectrometer operating at 500.13 and 125.77 MHz, respectively. Typical ^{1}H NMR spectra were acquired with a spectral width of 3.4 kHz, 64k data points and 16 scans. ^{13}C spectra of fully labelled D-glucose and leucine, needed for complete spectral analysis, were measured with gated decoupling using a spectral width of 15 kHz for glucose and 25 kHz for leucine with 128k data points and 512 scans. The sample temperature was 303.5 K for D-glucose and 300 K for the amino acids.

^{1}H spectra of the fully labelled D-glucose and mixture of the non-labelled and fully labelled D-glucose were recorded on an 800 MHz Varian Inova spectrometer. ^{1}H NMR spectra were acquired with a spectral width of 8 kHz, 64k data points and 32 scans for fully labelled and 16 for mixture. The sample temperature was 293K.

3.2.3 pH indicators

All the measurements were performed on Bruker Avance 500 DRX spectrometer operating at 500.13 MHz equipped with a 2.5 mm inverse broad band probe. Shimming and tuning of the samples were performed automatically. Standard ^{1}H NMR spectra were acquired with a spectral width of 6 kHz, 128k data points, 32 scans and the total recycling time 11 s. The sample temperature was 298.1 K.

3.2.4 Metabolite mixtures

All the measurements were performed on Bruker AVANCE III 500 spectrometer operating at 500.36 MHz equipped with a 5 mm selective inverse probe. Shimming and tuning of the samples were performed automatically. The baseopt digital filter, which is needed to achieve a flat baseline and zero degree first order phase correction, was used. The spectra were measured at a physiological temperature of 310 K.

^{1}H NMR spectra of the mixtures without serum background were recorded with 160k data points after 4 dummy scans using 8 transients acquired with an automatically calibrated 90° pulse and applying a Bruker noesypresat pulse sequence with a mixing time of 10 ms and irradiation field of 25 Hz to suppress the water peak. The acquisition time was 5.5 s, the relaxation delay (d1) 3.0 s and the additional relaxation delay (d2) before the suppression 51.5 s.

In order to remove majority of protein background and detect the low-molecular-weight metabolites from the spectra of the samples with serum background, T_2-editing was used. The T_2-edited spectra were collected with 64k data points using 24 transients acquired after 4 dummy scans with a Bruker 1D CPMG pulse sequence with water peak suppression and a 78 ms T_2-filter with a fixed echo delay of 403 μs to minimise diffusion and J-modulation effects. The acquisition time was 3.3 s and the relaxation delay 3 s.

T_2 times of the metabolites were estimated by measuring T_2-edited spectra of metabolite mixture with different T_2-filters (60–4000 ms). The spectra were recorded with 64k data points using 128 transients acquired after 4 dummy scans. The acquisition time was 3.3 s and the relaxation delay 60 s.

3.3 SPECTRAL ANALYSIS

3.3.1 Adaptive spectral library and isotope shifts

Preparation and analysis of the spectra were done using PERCH NMR Software. The spectral parameters were solved first by applying the IT method, after which the parameters were refined by the TLS fitting mode of PERCHit iterator. The programs MIX and UNMIX, needed in preparing the synthetic mixture spectra of isotopomers and in their decomposition analysis, were written for this purpose. The proposed protocol for ^{13}C isotopomer population analysis from ^{1}H NMR spectra was tested with simulated 600 MHz spectra of 232 isotopomers selected using metabolic rules (Maaheimo et al. 2001) for the 16 amino acids that survive protein-acid hydrolysis and isolation. The amino acid composition of bovine serum albumin was used in the simulation. The ^{13}C isotope effects on ^{1}H chemical shifts of benzene were obtained by reinvestigating the spectrum measured in our laboratory earlier (Laatikainen et al. 1995).

3.3.2 pH indicators

All the spectra were processed using PERCH NMR Software. The measured FIDs were multiplied by an exponential window function (LB=0.1 Hz) to increase the signal-to-noise ratio. In the cases, where the TSP and DSS signals overlapped strongly, the FIDs were multiplied by $Asin^2x+Be^y$ window function prior to Fourier transformation to enhance the resolution of these two signals. The chemical shifts of the indicator compound signals were determined by using the TLS tool of PERCH NMR Software.

3.3.3 qQMSA

The program developed in this work, qQMTLS (quantitative Quantum Mechanical Total-Line-Shape), was written in FORTRAN and it combines our previous CTLS approach (Laatikainen et al. 1996a;Soininen et al. 2005) and the iterative QM spectral analysis of the PERCHit iterator (Laatikainen et al. 1996b). Although the graphical interface of qQMTLS allows all the spectral preparations needed for the QMSA the file administration is based on support of PERCH NMR Software.

In qQMTLS, the total line shape $I(\nu)$ of an NMR frequency domain spectrum (intensity I vs. spectral frequency ν) is expressed as the sum of the QM systems $Q_n(\nu)$, the signals $S_m(\nu)$ which are not described using QM model, and the baseline $B(\nu)$ which may contain the protein background and instrumental artefacts and noise:

$$I(\nu) = \sum X_n Q_n(\nu) + \sum x_m S_m(\nu) + B(\nu) + noise(\nu) \tag{3.1}$$

The concentrations X_n and x_m are the parameters whose values are wanted. The QM spectra $Q_n(\nu)$, can be described by the equation

$$Q_n(\nu) = F_n(\nu, w, J, R, \Delta, line\ shape) \tag{3.2}$$

where the function F_n is a non-explicit function defined by the spin system and the vectors w, J, R and Δ contain the chemical shifts, coupling constants, response factors and line widths needed to describe the spin system, respectively. All these parameters reflect the physical and chemical conditions of the sample and can be therefore worth inspection after spectral analysis. The line shape is expressed as a sum of Lorentzian, Gaussian and dispersion functions. (Laatikainen et al. 1996a)

The non-quantum mechanical signals $S_m(\nu)$ can be used to describe signals which arise from an unknown or chemically non-stoichiometric component. The general formula of $S_m(\nu)$ is

$$S_m(\nu) = G_m(\nu, w, J, \Delta, line\ shape) \tag{3.3}$$

where G is a multiplet function (singlet, doublet, triplet, etc.) which can be regular (1:1 doublet, 1:2:1 triplet, etc.) with splitting described by J or irregular one for which the relative positions, intensities and line widths are freely optimisable.

Also the baseline $B(\nu)$ can be optimised during iteration. In qQMTLS, optimisable multiterm baseline function (Tynkkynen et al. 2012)

$$B(\nu) = \sum_n b_n b(\nu) \tag{3.4}$$

is employed. The $b(\nu)$ terms are cos^2-functions

$$b(\nu) = \cos^2(\nu - \nu_n) / d\nu \tag{3.5}$$

where $b(\nu) = 0$ when $\nu > \nu_n + d\nu$ or $\nu < \nu_n - d\nu$ and they are set to evenly distributed positions ($\nu_n+1 = \nu_n + d\nu$) so that their sum $B(\nu) = 1.0$ at every point when all $b_n = 1$. The advantage of the localised function is that any part of the spectrum can be fitted independently of the rest of the spectrum, preserving the exact form of the baseline function in memory.

4 Results and discussion

4.1 ADAPTIVE SPECTRAL LIBRARY

Amino acid ^{13}C isotopomer ratios carry invaluable information on intracellular metabolic flux distribution that can be used as additional constraints in metabolic flux analysis. In fact, the strategies based on the isotopic labelling offer the only gate to direct experimental quantification of intracellular metabolic fluxes. (Szyperski 1998) Previously, there has not been a fast method available to determine local ^{13}C enrichments of amino acids. In this work, a 1D NMR spectrum based approach to determine the local enrichments and isotopomer populations was established and the complete spectral analyses of ^{1}H coupled ^{13}C NMR spectra of all proteogenic amino acids were reported. Also, the ASL principle was introduced and discussed.

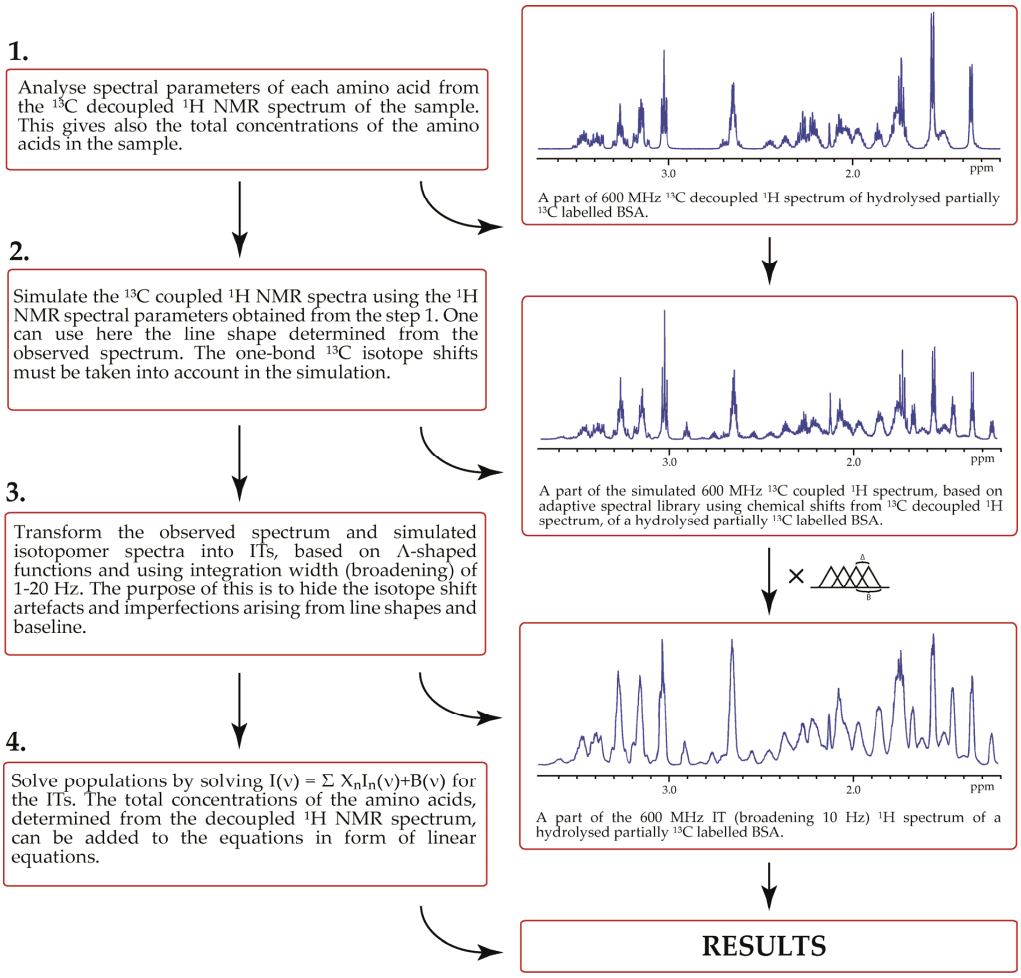

Figure 12. Proposed protocol for ^{13}C isotopomer population analysis.

ASL can be described as a library of spectral parameters obtained through QMSA. The parameters in the library can be used to simulate the spectra of the compounds in any magnetic field, line shape, line widths and, also, taking into account different sample conditions like pH or solvent. We also proposed a protocol for ^{13}C isotopomer population analysis from ^{1}H NMR spectra based on the simulated isotopomer spectra (Figure 12). The proposed protocol was tested with simulated cases and the results suggested that invaluable information about the positional fractional ^{13}C enrichments could be extracted from analysis of 1D ^{13}C-coupled ^{1}H NMR spectrum, especially, when combined with data obtained from biological experiment and MS. During the work, it became clear that robust automated quantification requires also good estimates of ^{13}C isotope effects on ^{1}H chemical shifts. (Publication II)

4.2 ^{13}C ISOTOPE EFFECTS ON ^{1}H CHEMICAL SHIFTS

Isotope effects on chemical shifts in NMR spectroscopy have been commonly used for structural and bonding studies, signal assignment and testing theories of chemical shifts (Buceta et al. 2008;Hansen 1988;Vidossich et al. 2006). While ^{2}H isotope effects on ^{13}C chemical shifts are obviously the most interesting and important ones, the ^{13}C isotope effects on ^{1}H chemical shifts have not been similarly in focus — there are only a few previous studies about ^{13}C induced isotope shift on proton. (Chertkov & Sergeyev 1983;Espinosa & Parella 2008;Everett 1984;Hoffman, Treitel, & Rabinovitz 2001;Laatikainen et al. 1995) The main reason for this is that these effects are small making them more difficult to exploit and normally they can be even ignored. However, in determination of the enrichment ratios of amino acid ^{13}C isotopomers by NMR spectroscopy (Publication I), these effects become significant and may even be useful.

Our results outlined some general rules for ^{13}C isotope effects on ^{1}H chemical shifts for amino acids and glucose. Based on our experiments and literature, it could be said that the one-bond effects in non-cyclic aliphatic systems are typically around -2 ppb. Instead, in cyclic systems, like in glucose, these shifts may be far larger, up to -4.4 ppb. The effects through two bonds varied in our cases between -0.3 and -1.0 ppb. Also, the long-range isotope effects through three to five bonds were observed, but for them no clear trends could be seen on the basis of our data. In the multi-labelled glucose the isotope effects appeared strongly non-additive, but for amino acids the effects were additive and by using additivity of the effects, the isotope effects for the non-cyclic amino acids can be predicted with sufficient accuracy.

4.3 ASSESMENT OF ^{1}H NMR PH INDICATORS

The proton NMR chemical shifts of typical metabolites are sensitive to the sample conditions of which the most important property is pH. This complicates quantification of metabolites from biological samples. However, if both the pH value of the sample and the pH-dependencies of the component chemical shifts are known, the component model spectra can be predicted sufficiently for model based quantification protocols. The use of ^{1}H NMR pH indicator compounds is proposed to be a repeatable and easily applicable method for pH determination of NMR sample. For ^{1}H pH indicators, several candidates had already been suggested. (Baryshnikova, Williams, & Sykes 2008;Pan et al. 1988;Rabenstein & Isab 1982;Rothman et al. 1997;Szakács, Hägele, & Tyka 2004;Vermathen, Capizzano, & Maudsley 2000) Our emphasis was placed generally on applicability to biological samples and, in addition, systematic study of the ionic strength effects was reported for the first time.

In this work, the accurate models for chemical shift pH-dependencies were given for selected 10 compounds that could be used as NMR pH indicators. The proposed set of indicators (dichloroacetic acid, glycine, 1,2,4-triazole, creatine, chloroacetic acid, formic acid, acetic acid, creatinine, histidine and TSP versus DSS) allow pH determination from pH^* 0 to 7.2 (Figure 13). Since the models were created for D_2O solutions, the numerical pH values (marked here by pH^*) refer to the glass electrode pH readings obtained without deuterium isotope correction. In opposition to the potentiometric method depending on the instrumentation and calibration, NMR method gives the same pH_{nmr}^* value for the same sample measured with any spectrometer if the temperature is same.

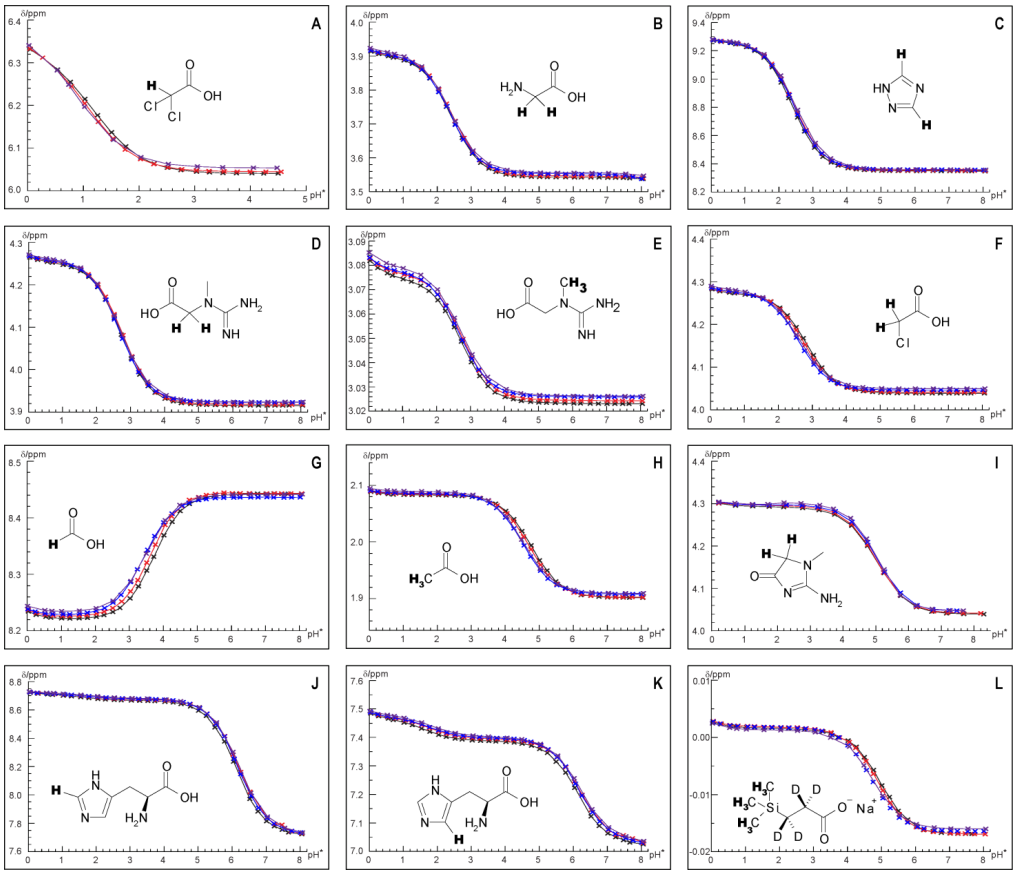

Figure 13. The chemical shifts of the pH indicator compounds as a function of pH^* (black, I = 0) referenced to DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid). The effect of ionic strength on the chemical shifts was explored with 0.15 M sodium chloride (red, I = 0.15 M), 0.15 M magnesium chloride (blue, I = 0.45 M) and 0.15 M calcium chloride (purple, I = 0.45 M). A dichloroacetic acid, B glycine, C 1,2,4-triazole, D creatine CH_2, E creatine CH_3, F chloroacetic acid, G formic acid, H acetic acid, I creatinine CH_2, J histidine H2, K histidine H4, L TSP versus DSS.

4.4 QUANTUM MECHANICAL QUANTIFICATION

The information content of ^1H NMR spectra of a biological sample is high, but transformation of this complex information into accurate concentrations of individual

compounds is not necessarily straightforward due to the spectral complexity and overlap. Despite this, ^{1}H NMR spectroscopy is widely used for profiling of a variety of complex biological samples such as urine, cerebrospinal fluid, serum, faeces, and tissue extracts. (Beckonert et al. 2007;Kettunen et al. 2012)

A distinctive feature of high-resolution 1D NMR spectra is that even the most complex spectrum of a compound, composed of thousands of individual spectral lines, can be described by a few spectral parameters within experimental accuracy, employing a quantum mechanical (QM) model (Abraham 1971). Hence, even in the case of the ^{1}H NMR spectrum of a complex mixture, there are strict QM rules between the signals of individual compounds. However, none of the earlier quantitative analysis methods like the spectral binning (Anthony et al. 1994), peak alignment algorithms and curve-fitting methods (Crockford et al. 2005;Davis et al. 2007;Stoyanova et al. 2004;Torgrip et al. 2006), singular value decomposition (Xu et al. 2006) and targeted profiling (Mercier et al. 2011;Weljie et al. 2006) fully utilise these rules.

In this work, our aim was to build up such a spectral analytical tool that uses and interprets the spectral information in maximal way taking also advantage from the prior knowledge available from the sample. In this approach, the signals of well-defined, stoichiometric components are described by QM model while unknown signals and the background are described using fittable lines and an optimisable multiterm baseline function (Tynkkynen et al. 2012), respectively. The program qQMTLS (quantitative Quantum Mechanical Total-Line-Shape) combines our previous CTLS approach and the iterative QMSA. In addition to the concentrations of known metabolites, this approach offers chemical confidence, which means that individual components are identified and quantified with high confidence on the basis of their spectral parameters such as coupling constants and chemical shifts. The assessment of the approach called herein qQMSA was performed with metabolite mixtures without and with human serum background. The intent of the first assessment was to test the performance of the program, different kinds of fitting protocols and parameters with known metabolite mixtures without background. The intent of the second assessment was to test the performance of qQMSA with serum spectra measured in high-throughput manner, with emphasis on the effects of T_2-editing. Figure 14 gives an example of qQMSA of the T_2-edited serum spectrum.

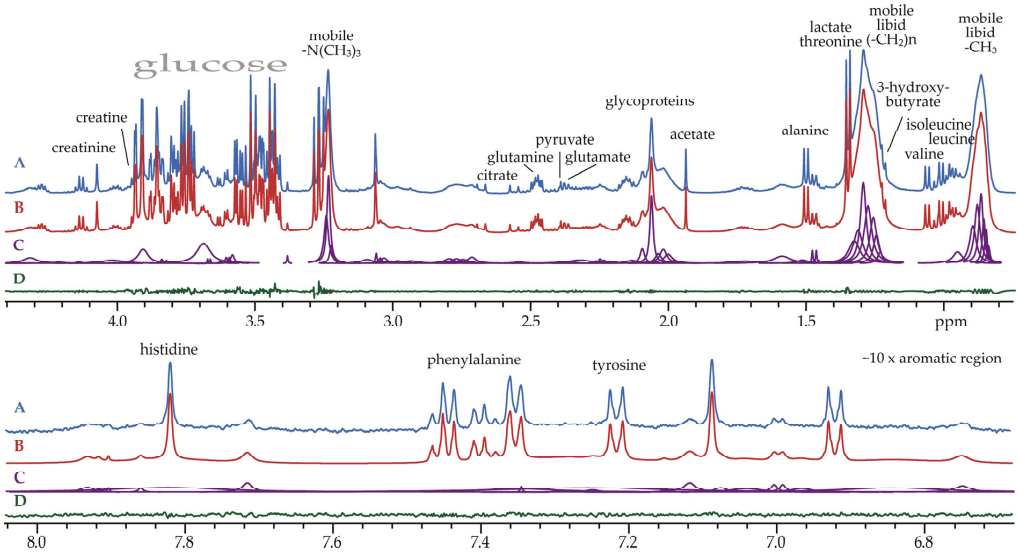

Figure 14. The qQMSA of the T_2-edited serum spectrum (A = observed spectrum, B = calculated spectrum, C = fitted lines, and D = difference between the calculated and the observed spectrum).

Table 6. The response factors of glucose obtained from the analysis of spectra measured with different settings.

	qH[a] D_2O	H[b] D_2O	qpresat[c] D_2O	presat[d] D_2O	qpresat[c] H_2O+ D_2O	presat[d] H_2O+ D_2O
α-H1	0.962	0.875	0.960	0.880	0.950	0.924
α-H2	0.974	0.993	0.965	0.993	0.904	0.909
α-H3	1.000	0.910	1.000	0.920	0.969	1.000
α-H4	0.978	0.953	0.990	0.990	1.000	0.978
α-H5	0.965	0.997	0.975	1.000	0.850	0.885
α-H6A	0.977	0.997	0.953	0.994	0.884	0.868
α-H6B	0.975	1.000	0.955	0.981	0.811	0.840
β-H1	nd	nd	nd	nd	nd	nd
β-H2	0.988	0.869	0.949	0.840	1.000	0.993
β-H3	0.996	0.955	0.978	0.945	0.986	1.000
β-H4	0.986	0.959	0.951	0.926	0.952	0.954
β-H5	0.989	0.993	1.000	1.000	0.974	0.989
β-H6A	1.000	1.000	0.913	0.914	0.870	0.881
β-H6B	0.982	0.987	0.904	0.908	0.845	0.863

[a] Basic proton spectrum (zg): 128k data points (td), 4 dummy scans (ds), 8 transients (ns), acquisition time (aq) 7.7 s and relaxation delay (d1) 52.3 s.

[b] Basic proton spectrum (zg): td = 128k, ds = 4, ns = 32, aq = 7.7 s and d1 = 2.3 s.

[c] Noesypresat pulse sequence (noesygppr1d): td = 128k, ds = 4, ns = 8, aq = 7.7 s, d1 = 3.0 s and additional relaxation delay before suppression (d2) 49.3 s.

[d] Noesypresat pulse sequence (noesygppr1d): td = 128k, ds = 4, ns = 8, aq = 7.7 s, d1 = 3.0 s and d2 = 0 s.

For mixtures without complex background, the components can be easily quantified with relative average error in concentrations less than 5% with appropriate fitting protocol. It is also important to take the response factors into account in the analysis, especially if there are overlapping major and minor components, because the response factors may differ a few per cent from 1.0 even in the simplest experiments (Table 6). In the best fitting protocol average errors of all the metabolite concentrations were smaller than 10% and the average of all the metabolites average errors was 3.8% (Figure 15). In this protocol line widths and shapes were optimised, overweighting and locking of populations were used, some response factors and coupling constants (glucose and lactate) were optimised, and a template for line widths was used.

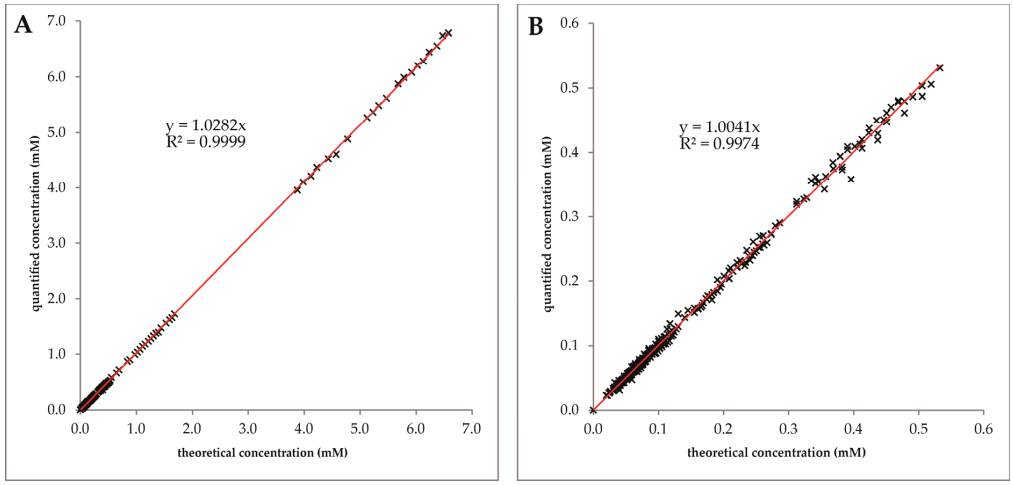

Figure 15. Correlations between the theoretical and quantified metabolite concentrations in metabolite mixtures without serum background. A includes all the metabolites. For more detailed view, glucose and lactate are excluded from B.

In order to estimate the bias that T_2-editing causes to metabolite concentrations, T_2 times of metabolites were estimated from T_2-edited spectra of a metabolite mixture measured with different T_2-filters lengths (Table 7). The experiments show that even the cautious T_2-editing yields small systematic bias in the absolute concentrations of the metabolites. However, the concentrations can be converted to absolute ones if the magnitudes of the effects are known.

Table 7. The determined T_2 times and the calculated recovery percentages for two different T_2-filters.

		recovery-% with given T_2-filter	
Metabolite	**T_2/s**	**80 ms**	**400 ms**
3-Hydroxybutyrate	1.2	95	79
Acetate	4.3	99	94
Alanine	1.0	95	76
Citrate	0.4	88	53
Creatine	1.2	96	80
Creatinine	2.3	98	89
Glucose	0.8	93	70
Glutamine	0.8	93	71
Histidine	1.3	96	80
Lactate	1.5	96	83
Leucine	1.0	94	75
Phenylalanine	2.1	97	87
Pyruvate	3.8	99	93
Threonine	1.0	95	76
Tyrosine	1.3	96	81
Valine	1.0	95	77

The effects of the T_2-editing, high-throughput manner performed measurements and serum background (protein interactions) complicate the qQMSA and the use of appropriate fitting protocols is invaluable in the quantification of low concentration compounds. In the

case of T_2-edited spectra measured in high-throughput manner, response factors must be used in the analysis. For example, the response factor of lactate CH proton can be as low as 0.70 when compared to that of the lactate CH_3 protons. On the basis of our observations, we ended up to the practise that the largest response factor of a compound is set to 1.0. In the best fitting protocol line widths were optimised, the certain integrals were locked (3-hydroxybutyrate, acetoacetate, aromatic amino acids, isoleucine, leucine, pyruvate and valine) after separated fitting, a template for line widths was used and the response factors obtained from the T_2-edited spectra of individual metabolites were applied (Figure 16). Our results suggest that moderately T_2-edited serum spectra obey well the QM rules and that qQMSA allows reliable quantification of the most common metabolites in human serum when variations in the response factors are taken into account.

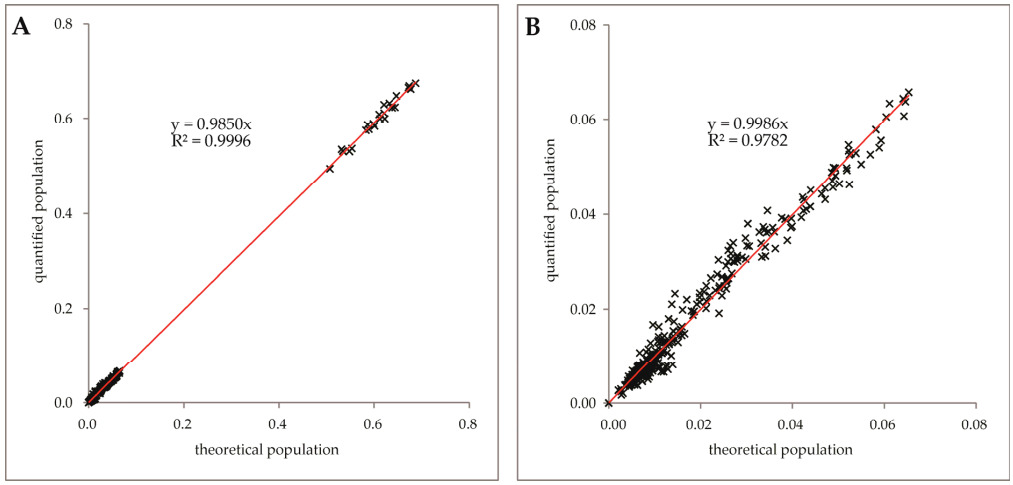

Figure 16. Correlations between the theoretical and quantified metabolite populations in the metabolite mixtures with serum background. A includes all the metabolites. For more detailed view, glucose and lactate are excluded from B.

5 Summary and conclusions

In the present study, quantification methods, strategies and protocols for NMR spectra of different biological samples were developed and assessed.

A distinctive feature of ^{1}H NMR spectrum is that it can be described accurately using quantum mechanical (QM) rules, and, as a result of this, the spectral parameters can be extracted from the observed spectra by quantum mechanical spectral analysis (QMSA). As an application of QMSA, the adaptive spectral library (ASL) principle was introduced in this project. ASL can be described as a library of spectral parameters obtained through QMSA. The parameters in the library can be used to simulate the spectra of the compounds in any magnetic field, line shape, line widths and, also, taking into account different sample conditions like pH or solvent. (Publication I)

Our aim was to build up such a spectral analytical tool that uses and interprets the spectral information in maximal way taking also advantage from the prior knowledge available from the sample. In our approach, called herein quantitative Quantum Mechanical Spectral Analysis (qQMSA), the signals of well-defined, stoichiometric, components can be described by QM model and the unknown signals and the possible spectral background can be described using fittable lines and optimisable multiterm function. The program qQMTLS (quantitative Quantum Mechanical Total-Line-Shape) combines our previous constrained total-line-shape (CTLS) approach and the iterative QMSA. In addition to the accurate concentrations of known metabolites, this approach offers chemical confidence, which means that individual components are identified with high confidence on the basis of their spectral parameters. The development and assessment of the approach was performed with metabolite mixtures without and with human serum background. In the case of the metabolite mixtures without complex background the qQMSA gives nearly unbiased estimates of the components in large concentration range. Additionally, our results suggest that moderately T_2-edited serum spectra obey well the QM rules and that qQMSA allows reliable quantification of the most common metabolites in human serum when variations in the response factors are taken into account. (Publication IV and manuscript V)

As an application of ASL principle (Publication I), a 1D NMR spectrum based approach to determine the positional fractional ^{13}C enrichments and ^{13}C isotopomers populations was established and the complete spectral analyses of ^{1}H coupled ^{13}C NMR spectra of all the proteogenic amino acids were reported. We also proposed a protocol for ^{13}C isotopomer population analysis from ^{1}H NMR spectra based on the simulated isotopomer spectra. The proposed protocol was tested with simulated cases and the results suggested that invaluable information about the positional fractional ^{13}C enrichments could be extracted from analysis of 1D ^{13}C-coupled ^{1}H NMR spectrum, especially, when combined with data obtained from biological experiment and mass spectroscopy. During the work, it became clear that robust automated quantification requires also good estimates of ^{13}C isotope effects on ^{1}H chemical shifts. These effects are small making them difficult to exploit and normally they can be even ignored. However, in determination of the enrichment ratios of amino acid ^{13}C isotopomers by NMR spectroscopy (Publication I), these effects become significant. Our results outlined some general rules for ^{13}C isotope effects on ^{1}H chemical shifts for amino acids and glucose. In the multi-labelled glucose the isotope effects appeared strongly non-additive, but for amino acids the effects were additive, and, by using additivity of the effects, the isotope effects for the non-cyclic amino acids can be predicted with sufficient accuracy (Publication II)

As closely related to the Publications I and II, the accurate models for chemical shift pH-dependencies were given for selected 10 compounds. In addition, systematic study of the

ionic strength effects was reported for the first time. The proposed set of indicators allow pH determination from pH^* 0 to 7.2. In opposition to the potentiometric method depending on the instrumentation and calibration, NMR method gives the same pH_{nmr}^* value for the same sample measured with any spectrometer if the temperature is same (Publication III)

This was the first systematic study of QMSA approach for quantification of NMR spectra of complex mixtures. It is expected that the protocols and tools, developed in this study, ASL and qQMSA, will enable accurate, robust and cost-effective way to quantify individual components from the NMR spectra of complex mixtures. ASL principle is the most efficient and adaptive way to storage spectral data and qQMSA can be apply to almost any mixture.

6 References

Abraham, J., Warne, A., & Griffiths, L. 1997, "Proton chemical shifts in NMR. Part 9.1 Steric and electric field effects in chlorine substituent chemical shifts (SCS)", *Journal of the Chemical Society, Perkin Transactions 2,* no. 5, pp. 881-886.

Abraham, R. J. 1971, *Analysis of high resolution NMR spectra,* Elsevier Publishing Co, New York.

Abraham, R. J. 1999, "A model for the calculation of proton chemical shifts in non-conjugated organic compounds", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 35, no. 2, pp. 85-152.

Abraham, R. J., Byrne, J. J., Griffiths, L., & Perez, M. 2006, "^{1}H chemical shifts in NMR: Part 23, the effect of dimethyl sulphoxide versus chloroform solvent on ^{1}H chemical shifts", *Magnetic Resonance in Chemistry*, vol. 44, no. 5, pp. 491-509.

Akoka, S., Barantin, L., & Trierweiler, M. 1999, "Concentration measurement by proton NMR using the ERETIC method", *Analytical Chemistry*, vol. 71, no. 13, pp. 2554-2557.

Alkorta, I. & Elguero, J. 2003, "Review on DFT and *ab initio* calculations of scalar coupling constants", *International Journal of Molecular Sciences*, vol. 4, no. 3, pp. 64-92.

Alm, E., Slagbrand, T., Åberg, K., Wahlström, E., Gustafsson, I., & Lindberg, J. 2012, "Automated annotation and quantification of metabolites in ^{1}H NMR data of biological origin", *Analytical and Bioanalytical Chemistry*, vol. 403, no. 2, pp. 443-455.

Altona, C. 2007, "Vicinal coupling constants and conformation of biomolecules," in *Encyclopedia of Magnetic Resonance,* John Wiley & Sons, Ltd.

Anderson, P., Mahle, D., Doom, T., Reo, N., DelRaso, N., & Raymer, M. 2011, "Dynamic adaptive binning: An improved quantification technique for NMR spectroscopic data", *Metabolomics*, vol. 7, no. 2, pp. 179-190.

Anthony, M. L., Sweatman, B. C., Beddell, C. R., Lindon, J. C., & Nicholson, J. K. 1994, "Pattern recognition classification of the site of nephrotoxicity based on metabolic data derived from proton nuclear magnetic resonance spectra of urine", *Molecular Pharmacology*, vol. 46, no. 1, pp. 199-211.

Augustine, M. P. & Zilm, K. W. 1996, "Observation of bulk susceptibility effects in high-resolution nuclear magnetic resonance", *Journal of Magnetic Resonance Series A*, vol. 123, no. 2, pp. 145-156.

Barding, G. A., Salditos, R., & Larive, C. K. 2012, "Quantitative NMR for bioanalysis and metabolomics", *Analytical and Bioanalytical Chemistry*, vol. 404, no. 4, pp. 1165-1179.

Barfield, M. 1971, "Conformation and substituent dependence of long-range hydrogen-hydrogen coupling over four bonds", *Journal of the American Chemical Society*, vol. 93, no. 5, pp. 1066-1071.

Barfield, M. & Johnston, M. D. 1973, "Solvent dependence of nuclear spin-spin coupling constants", *Chemical Reviews*, vol. 73, no. 1, pp. 53-73.

Bartlett, M. S. 1950, "Periodogram analysis and continuous spectra", *Biometrika*, vol. 37, no. 1-2, pp. 1-16.

Baryshnikova, O. K., Williams, T. C., & Sykes, B. D. 2008, "Internal pH indicators for biomolecular NMR", *Journal of Biomolecular NMR*, vol. 41, no. 1, pp. 5-7.

Batley, M. & Redmond, J. W. 1982, "^{31}P NMR reference standards for aqueous samples", *Journal of Magnetic Resonance (1969)*, vol. 49, no. 1, pp. 172-174.

Beckonert, O., Keun, H. C., Ebbels, T. M. D., Bundy, J., Holmes, E., Lindon, J. C., & Nicholson, J. K. 2007, "Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts", *Nature Protocols*, vol. 2, no. 11, pp. 2692-2703.

Bernstein, M. A., Štefinovic, M., & Sleigh, C. J. 2007, "Optimising reaction performance in the pharmaceutical industry by monitoring with NMR", *Magnetic Resonance in Chemistry*, vol. 45, no. 7, pp. 564-571.

Blinov, K. A., Carlson, D., Elyashberg, M. E., Martin, G. E., Martirosian, E. R., Molodtsov, S., & Williams, A. J. 2003, "Computer-assisted structure elucidation of natural products with limited 2D NMR data: Application of the StrucEluc system", *Magnetic Resonance in Chemistry*, vol. 41, no. 5, pp. 359-372.

Buceta, N. N., la Vedova, C. O., Romanelli, G. P., Autino, J. C., & Jos, J. L. 2008, "Deuterium isotopic effect on ^{13}C NMR chemical shifts of 1-(2-hydroxyphenyl)-3-aryl-1,3-propanediones: Hydrogen bond and substituent effects", *Journal of Molecular Structure*, vol. 878, no. 1-3, pp. 50-59.

Castellano, S. & Bothner-By, A. A. 1964, "Analysis of NMR spectra by least squares", *The Journal of Chemical Physics*, vol. 41, no. 12, pp. 3863-3869.

Castillo, A. M., Patiny, L., & Wist, J. 2011, "Fast and accurate algorithm for the simulation of NMR spectra of large spin systems", *Journal of Magnetic Resonance*, vol. 209, no. 2, pp. 123-130.

Chertkov, V. A. & Sergeyev, N. M. 1983, "^{13}C isotope effects on the parameters of proton magnetic resonance spectra of benzene", *Journal of Magnetic Resonance (1969)*, vol. 52, no. 3, pp. 400-406.

Choi, C., Münch, R., Leupold, S., Klein, J., Siegel, I., Thielen, B., Benkert, B., Kucklick, M., Schobert, M., Barthelmes, J., Ebeling, C., Haddad, I., Scheer, M., Grote, A., Hiller, K., Bunk, B., Schreiber, K., Retter, I., Schomburg, D., & Jahn, D. 2007, "SYSTOMONAS - an integrated database for systems biology analysis of Pseudomonas", *Nucleic Acids Research*, vol. 35, no. suppl 1, p. D533-D537.

Claridge, T. D. W. 1999, *High-Resolution NMR Techniques in Organic Chemistry*, 1 edn, Pergamon Press.

Consonni, R., Cagliani, L. R., Benevelli, F., Spraul, M., Humpfer, E., & Stocchero, M. 2008, "NMR and Chemometric methods: A powerful combination for characterization of Balsamic and Traditional Balsamic Vinegar of Modena", *Analytica Chimica Acta*, vol. 611, no. 1, pp. 31-40.

Contreras, R. H. & Peralta, J. E. 2000, "Angular dependence of spin-spin coupling constants", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 37, no. 4, pp. 321-425.

Corio, P. L. & Smith, S. L. 2007, "Analysis of high-resolution solution state spectra," in *Encyclopedia of Magnetic Resonance*, John Wiley & Sons, Ltd.

Cremer, D. & Grafenstein, J. 2007, "Calculation and analysis of NMR spin-spin coupling constants", *Physical Chemistry Chemical Physics*, vol. 9, no. 22, pp. 2791-2816.

Crockford, D. J., Keun, H. C., Smith, L. M., Holmes, E., & Nicholson, J. K. 2005, "Curve-fitting method for direct quantitation of compounds in complex biological mixtures using ^{1}H NMR: Application in metabonomic toxicology studies", *Analytical Chemistry*, vol. 77, no. 14, pp. 4556-4562.

Cuesta, I. G., De Merás, A. S., Pelloni, S., & Lazzeretti, P. 2009, "Understanding the ring current effects on magnetic shielding of hydrogen and carbon nuclei in naphthalene and anthracene", *Journal of Computational Chemistry*, vol. 30, no. 4, pp. 551-564.

Cui, Q., Lewis, I. A., Hegeman, A. D., Anderson, M. E., Li, J., Schulte, C. F., Westler, W. M., Eghbalnia, H. R., Sussman, M. R., & Markley, J. L. 2008, "Metabolite identification via the Madison Metabolomics Consortium Database", *Nature Biotechnology*, vol. 26, no. 2, pp. 162-164.

Davis, R. A., Charlton, A. J., Godward, J., Jones, S. A., Harrison, M., & Wilson, J. C. 2007, "Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform", *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 1, pp. 144-154.

de Dios, A. C. 1996, "*Ab initio* calculations of the NMR chemical shift", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 29, no. 3-4, pp. 229-278.

de Graaf, R. A. & Behar, K. L. 2003, "Quantitative ^{1}H NMR spectroscopy of blood plasma metabolites", *Analytical Chemistry*, vol. 75, no. 9, pp. 2100-2104.

De Meyer, T., Sinnaeve, D., Van Gasse, B., Tsiporkova, E., Rietzschel, E. R., De Buyzere, M. L., Gillebert, T. C., Bekaert, S., Martins, J. C., & Van Criekinge, W. 2008, "NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm", *Analytical Chemistry*, vol. 80, no. 10, pp. 3783-3790.

Diehl, P., Sýkora, S., & Vogt, J. 1975, "Automatic analysis of NMR spectra: An alternative approach", *Journal of Magnetic Resonance (1969)*, vol. 19, no. 1, pp. 67-82.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. 2006, "Metabolite projection analysis for fast identification of metabolites in metabonomics. Application in an amiodarone study", *Analytical Chemistry*, vol. 78, no. 11, pp. 3551-3561.

Dunkel, R. 2007, "Computer-assisted structure elucidation," in *Encyclopedia of Magnetic Resonance*, John Wiley & Sons, Ltd.

Ebbels, T. M. D. & Cavill, R. 2009, "Bioinformatic methods in NMR-based metabolic profiling", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 55, no. 4, pp. 361-374.

Elyashberg, M. E., Williams, A., & Martin, G. E. 2008, "Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 53, no. 1-2, pp. 1-104.

Elyashberg, M., Blinov, K., Molodtsov, S., Smurnyy, Y., Williams, A., & Churanova, T. 2009, "Computer-assisted methods for molecular structure elucidation: realizing a spectroscopist's dream", *Journal of Cheminformatics*, vol. 1, no. 1, p. 3.

Espinosa, J. F. & Parella, T. 2008, "First experimental determination of two-bond ^{13}C isotopic effects on ^{1}H NMR chemical shifts", *Tetrahedron Letters*, vol. 49, no. 16, pp. 2562-2565.

Esteban, A. L., Galache, M. P., Mora, F., Díez, E., Casanueva, J., San Fabián, J., Barone, V., Peralta, J. E., & Contreras, R. H. 2001, "Vicinal NMR proton-proton coupling constants. An NBO analysis", *The Journal of Physical Chemistry A*, vol. 105, no. 21, pp. 5298-5303.

Everett, J. R. 1984, "An experimental study of carbon-13 isotope effects on proton nuclear magnetic resonance chemical shifts", *Journal of the Chemical Society, Perkin Transactions 2*, no. 7, pp. 1151-1153.

Facelli, J. C. & Orendt, A. M. 2007, "Magnetic shielding and chemical shifts: Basics," in *Encyclopedia of Magnetic Resonance*, John Wiley & Sons, Ltd.

Fahy, E., Sud, M., Cotter, D., & Subramaniam, S. 2007, "LIPID MAPS online tools for lipid research", *Nucleic Acids Research*, vol. 35, no. suppl 2, p. W606-W612.

Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J. C., & Sansone, S. A. 2007, "The metabolomics standards initiative (MSI)", *Metabolomics*, vol. 3, no. 3, pp. 175-178.

Fonville, J. M., Maher, A. D., Coen, M., Holmes, E., Lindon, J. C., & Nicholson, J. K. 2010, "Evaluation of full-resolution *J*-resolved ^{1}H NMR projections of biofluids for metabonomics information retrieval and biomarker identification", *Analytical Chemistry*, vol. 82, no. 5, pp. 1811-1821.

Fukui, H. 1997, "Theory and calculation of nuclear shielding constants", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 31, no. 4, pp. 317-342.

Fung, B. M. 2002, "^{13}C NMR studies of liquid crystals", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 41, no. 3-4, pp. 171-186.

Gipson, G. T., Tatsuoka, K. S., Sweatman, B. C., & Connor, S. C. 2006, "Weighted least-squares deconvolution method for discovery of group differences between complex biofluid ^{1}H NMR spectra", *Journal of Magnetic Resonance*, vol. 183, no. 2, pp. 269-277.

Go, E. P. 2010, "Database resources in metabolomics: An overview", *Journal of Neuroimmune Pharmacology*, vol. 5, no. 1, pp. 18-30.

Golotvin, S. S. & Chertkov, V. A. 1997, "Pattern recognition of the multiplet structure of NMR spectra", *Russian Chemical Bulletin*, vol. 46, no. 3, pp. 423-430.

Golotvin, S. S., Pol, R., Sasaki, R. R., Nikitina, A., & Keyes, P. 2012, "Concurrent combined verification: Reducing false positives in automated NMR structure verification through the evaluation of multiple challenge control structures", *Magnetic Resonance in Chemistry*, vol. 50, no. 6, pp. 429-435.

Golotvin, S. S., Vodopianov, E., Lefebvre, B. A., Williams, A. J., & Spitzer, T. D. 2006, "Automated structure verification based on ^{1}H NMR prediction", *Magnetic Resonance in Chemistry*, vol. 44, no. 5, pp. 524-538.

Golotvin, S. S., Vodopianov, E., Pol, R., Lefebvre, B. A., Williams, A. J., Rutkowske, R. D., & Spitzer, T. D. 2007, "Automated structure verification based on a combination of 1D ^{1}H NMR and 2D ^{1}H-^{13}C HSQC spectra", *Magnetic Resonance in Chemistry*, vol. 45, no. 10, pp. 803-813.

Gräfenstein, J., Tuttle, T., & Cremer, D. 2004, "Decomposition of nuclear magnetic resonance spin–spin coupling constants into active and passive orbital contributions", *The Journal of Chemical Physics*, vol. 120, no. 21, pp. 9952-9968.

Griffiths, L. & Irving, A. M. 1998, "Assay by nuclear magnetic resonance spectroscopy: Quantification limits", *Analyst*, vol. 123, no. 5, pp. 1061-1068.

Griffiths, L. & Bright, J. D. 2002, "Towards the automatic analysis of ^{1}H NMR spectra: Part 3. Confirmation of postulated chemical structure", *Magnetic Resonance in Chemistry*, vol. 40, no. 10, pp. 623-634.

Griffiths, L., Beeley, H. H., & Horton, R. 2008, "Towards the automatic analysis of NMR spectra: Part 7. Assignment of ^{1}H by employing both ^{1}H and ^{1}H/^{13}C correlation spectra", *Magnetic Resonance in Chemistry*, vol. 46, no. 9, pp. 818-827.

Günther, H. 1995, *NMR spectroscopy: Basic principles, concepts, and applications in chemistry*, 2 edn, John Wiley & Sons Ltd, Chichester.

Haasnoot, C. A. G., de Leeuw, F. A. A. M., & Altona, C. 1980, "The relationship between proton-proton NMR coupling constants and substituent electronegativities–I : An empirical generalization of the karplus equation", *Tetrahedron*, vol. 36, no. 19, pp. 2783-2792.

Hägele, G., Engelhardt, M., & Boenigk, W. 1987, *Simulation und Automatisierte Analyse von Kernresonanzspektren* Wiley-VCH, Weinheim.

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. 2005, "Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders", *Nucleic Acids Research*, vol. 33, no. suppl 1, p. D514-D517.

Hanhineva, K., Soininen, P., Anttonen, M. J., Kokko, H., Rogachev, I., Aharoni, A., Laatikainen, R., & Kärenlampi, S. 2009, "NMR and UPLC-qTOF-MS/MS characterisation of novel phenylethanol derivatives of phenylpropanoid glucosides from the leaves of strawberry (Fragaria x ananassa cv. Jonsok)", *Phytochemical Analysis*, vol. 20, no. 5, pp. 353-364.

Hansen, P. E. 1988, "Isotope effects in nuclear shielding", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 20, no. 3, pp. 207-255.

Harcken, H., Ketzler, R., Albrecht, M., Burghoff, M., Hartwig, S., & Trahms, L. 2010, "The natural line width of low field nuclear magnetic resonance spectra", *Journal of Magnetic Resonance*, vol. 206, no. 1, pp. 168-170.

Harris, R. K., Becker, E. D., De Menezes, S. M. C., Goodfellow, R., & Granger, P. 2001, "NMR nomenclature. Nuclear spin properties and conventions for chemical shifts - (IUPAC recommendations 2001)", *Pure and Applied Chemistry*, vol. 73, no. 11, pp. 1795-1818.

Heikkinen, S., Toikka, M. M., Karhunen, P. T., & Kilpeläinen, I. A. 2003, "Quantitative 2D HSQC (Q-HSQC) via suppression of *J*-dependence of polarization transfer in NMR spectroscopy: Application to wood lignin", *Journal of the American Chemical Society*, vol. 125, no. 14, pp. 4362-4367.

Heinzer, J. 1977, "Iterative least-squares NMR lineshape fitting with use of symmetry and magnetic equivalence factorization", *Journal of Magnetic Resonance (1969)*, vol. 26, no. 2, pp. 301-316.

Hoffman, R. E., Treitel, N., & Rabinovitz, M. 2001, "^{1}H NMR isotope shifts arising from the substitution of ^{12}C by ^{13}C: application to polycyclic aromatic hydrocarbon anions", *Magnetic Resonance in Chemistry*, vol. 39, no. 8, pp. 489-494.

Holmes, E., Nicholson, J. K., Nicholls, A. W., Lindon, J. C., Connor, S. C., Polley, S., & Connelly, J. 1998, "The identification of novel biomarkers of renal toxicity using automatic data reduction techniques and PCA of proton NMR spectra of urine", *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1-2, pp. 245-255.

Holzgrabe, U. 2010, "Quantitative NMR spectroscopy in pharmaceutical applications", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 57, no. 2, pp. 229-240.

Holzgrabe, U., Deubner, R., Schollmayer, C., & Waibel, B. 2005, "Quantitative NMR spectroscopy - Applications in drug analysis", *Journal of Pharmaceutical and Biomedical Analysis*, vol. 38, no. 5, pp. 806-812.

Hunter, C. A., Packer, M. J., & Zonta, C. 2005, "From structure to chemical shift and vice-versa", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 47, no. 1-2, pp. 27-39.

Jameson, C. J. 1981, "Effects of intermolecular interactions and intramolecular dynamics on nuclear resonance", *Bulletin of Magnetic Resonance*, vol. 3, no. 1, pp. 3-29.

Jameson, C. J. 1996, "Understanding NMR chemical shifts", *Annual Review of Physical Chemistry*, vol. 47, pp. 135-169.

Jameson, C. J. 2007, "Isotope effects on chemical shifts and coupling constants," in *Encyclopedia of Magnetic Resonance*, John Wiley & Sons, Ltd.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G. R., Wu, G. R., Matthews, L., Lewis, S., Birney, E., & Stein, L. 2005, "Reactome: A knowledgebase of biological pathways", *Nucleic Acids Research*, vol. 33, no. suppl 1, pp. D428-D432.

Jukarainen, N. M., Korhonen, S. P., Laakso, M. P., Korolainen, M. A., Niemitz, M., Soininen, P. P., Tuppurainen, K., Vepsäläinen, J., Pirttilä, T., & Laatikainen, R. 2008, "Quantification of ^{1}H NMR spectra of human cerebrospinal fluid: a protocol based on constrained total-line-shape analysis", *Metabolomics*, vol. 4, no. 2, pp. 150-160.

Kanehisa, M., Goto, S., Hattori, M., oki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., & Hirakawa, M. 2006, "From genomics to chemical genomics: New developments in KEGG", *Nucleic Acids Research*, vol. 34, no. suppl 1, pp. D354-D357.

Karplus, M. 1959, "Contact electron-spin coupling of nuclear magnetic moments", *The Journal of Chemical Physics*, vol. 30, no. 1, pp. 11-15.

Karplus, M. 1963, "Vicinal proton coupling in nuclear magnetic resonance", *Journal of the American Chemical Society*, vol. 85, no. 18, pp. 2870-2871.

Kastenmüller, G., Römisch-Margl, W., Wägele, B., Altmaier, E., & Suhre, K. 2011, "metaP-server: A web-based metabolomics data analysis tool", *Journal of Biomedicine and Biotechnology*, vol. 2011, no. Article ID 839862.

Keeler, J. 2005, *Understanding NMR spectroscopy*, 1 edn, John Wiley & Sons, Ltd.

Kettunen, J., Tukiainen, T., Sarin, A. P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L. P., Kangas, A. J., Soininen, P., Würtz, P., Silander, K., Dick, D. M., Rose, R. J., Savolainen, M. J., Viikari, J., Kähönen, M., Lehtimäki, T., Pietiläinen, K. H., Inouye, M., McCarthy, M. I., Jula, A., Eriksson, J., Raitakari, O. T., Salomaa, V., Kaprio, J., Järvelin, M. R., Peltonen, L., Perola, M., Freimer, N. B., Ala-Korpela, M., Palotie, A., & Ripatti, S. 2012, "Genome-wide association study identifies multiple loci influencing human serum metabolite levels", *Nature Genetics*, vol. 44, no. 3, pp. 269-276.

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dörmann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A. R., & Steinhauser, D. 2005, "GMD@CSB.DB: the Golm Metabolome Database", *Bioinformatics*, vol. 21, no. 8, pp. 1635-1638.

Koskela, H., Rapinoja, M. L., Kuitunen, M. L., & Vanninen, P. 2007, "Determination of trace amounts of chemical warfare agent degradation products in decontamination solutions with NMR spectroscopy", *Analytical Chemistry*, vol. 79, no. 23, pp. 9098-9106.

Koskela, H., Heikkilä, O., Kilpeläinen, I., & Heikkinen, S. 2010, "Quantitative two-dimensional HSQC experiment for high magnetic field NMR spectrometers", *Journal of Magnetic Resonance*, vol. 202, no. 1, pp. 24-33.

Kotowycz, G. & Schaefer, T. 1967, "Concerning the role of the reaction field in the proton shielding of molecules in solution", *Canadian Journal of Chemistry*, vol. 45, pp. 1093-1096.

Laatikainen, R. 1986, "An application of $^{n}J(CH_2,H)$ and $^{5}J(OH,H)$ long-range couplings in the conformational analysis of 2-hydroxybenzyl methyl ether in various solvents. Computational strategies for spectral analysis and determination of association thermodynamic parameters", *Magnetic Resonance in Chemistry*, vol. 24, no. 7, pp. 588-594.

Laatikainen, R. 1988, "Dipolar coupling and solvent dependence of ^{1}H, ^{1}H spin-spin coupling in naphthalene - Analysis of spectra containing overlapping lines", *Journal of Magnetic Resonance*, vol. 78, no. 1, pp. 127-132.

Laatikainen, R., Niemitz, M., Korhonen, S.-P., Hassinen, T., & Venäläinen, T. PERCH NMR Software, see www.perchsolutions.com (1993-2013). 1993.

Laatikainen, R., Ratilainen, J., Sebastian, R., & Santa, H. 1995, "NMR-study of aromatic-aromatic interactions for benzene and some other fundamental aromatic systems using alignment of aromatics in strong magnetic field", *Journal of the American Chemical Society*, vol. 117, no. 44, pp. 11006-11010.

Laatikainen, R., Niemitz, M., Malaisse, W. J., Biesemans, M., & Willem, R. 1996a, "A computational strategy for the deconvolution of NMR spectra with multiplet structures and constraints: Analysis of overlapping ^{13}C-^{2}H multiplets of ^{13}C enriched metabolites from cell suspensions incubated in deuterated media", *Magnetic Resonance in Medicine*, vol. 36, no. 3, pp. 359-365.

Laatikainen, R., Niemitz, M., Weber, U., Sundelin, J., Hassinen, T., & Vepsäläinen, J. 1996b, "General strategies for total-lineshape-type spectral analysis of NMR spectra using integral-transform iterator", *Journal of Magnetic Resonance Series A*, vol. 120, no. 1, pp. 1-10.

Lazzeretti, P. 2000, "Ring currents", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 36, no. 1, pp. 1-88.

Levy, G. C. & Komoroski, R. A. 1974, "Paramagnetic relaxation reagents. Alternatives or complements to lanthanide shift reagents in nuclear magnetic resonance spectral analysis", *Journal of the American Chemical Society*, vol. 96, no. 3, pp. 678-681.

Live, D. H. & Chan, S. I. 1970, "Bulk susceptibility corrections in nuclear magnetic resonance experiments using superconducting solenoids", *Analytical Chemistry*, vol. 42, no. 7, pp. 791-792.

Ludwig, C., Ward, D. G., Martin, A., Viant, M. R., Ismail, T., Johnson, P. J., Wakelam, M. J. O., & Günther, U. L. 2009, "Fast targeted multidimensional NMR metabolomics of colorectal cancer", *Magnetic Resonance in Chemistry*, vol. 47, no. S1, p. S68-S73.

Ludwig, C. & Viant, M. R. 2010, "Two-dimensional *J*-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox", *Phytochemical Analysis*, vol. 21, no. 1, pp. 22-32.

Ludwig, C., Easton, J., Lodi, A., Tiziani, S., Manzoor, S., Southam, A., Byrne, J., Bishop, L., He, S., Arvanitis, T., Günther, U., & Viant, M. 2012, "Birmingham Metabolite Library: a publicly accessible database of 1-D ^{1}H and 2-D ^{1}H *J*-resolved NMR spectra of authentic metabolite standards (BML-NMR)", *Metabolomics*, vol. 8, no. 1, pp. 8-18.

Maaheimo, H., Fiaux, J., Cakar, Z. P., Bailey, J. E., Sauer, U., & Szyperski, T. 2001, "Central carbon metabolism of *Saccharomyces cerevisiae* explored by biosynthetic fractional ^{13}C labeling of common amino acids", *European Journal of Biochemistry*, vol. 268, no. 8, pp. 2464-2479.

Mäkelä, A. V., Heikkilä, O., Kilpeläinen, I., & Heikkinen, S. 2011, "ImatraNMR: Novel software for batch integration and analysis of quantitative NMR spectra", *Journal of Magnetic Resonance*, vol. 211, no. 2, pp. 186-194.

Martineau, E., Giraudeau, P., Tea, I., & Akoka, S. 2011, "Fast and precise quantitative analysis of metabolic mixtures by 2D ^{1}H INADEQUATE NMR", *Journal of Pharmaceutical and Biomedical Analysis*, vol. 54, no. 1, pp. 252-257.

McConnell, J. 1987, *The theory of nuclear magnetic relaxation in liquids,* Cambridge University Press, Cambridge.

Mendes, P. 2002, "Emerging bioinformatics for the metabolome", *Briefings in Bioinformatics*, vol. 3, no. 2, pp. 134-145.

Mercier, P., Lewis, M., Chang, D., Baker, D., & Wishart, D. 2011, "Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra", *Journal of Biomolecular NMR*, vol. 49, no. 3, pp. 307-323.

Michaleas, S. & Antoniadou-Vyza, E. 2006, "A new approach to quantitative NMR: Fluoroquinolones analysis by evaluating the chemical shift displacements", *Journal of Pharmaceutical and Biomedical Analysis*, vol. 42, no. 4, pp. 405-410.

Middleton, D. A. 2011, "Solid-state NMR detection of ^{14}N-^{13}C dipolar couplings between amino acid side groups provides constraints on amyloid fibril architecture", *Magnetic Resonance in Chemistry*, vol. 49, no. 2, pp. 65-69.

Minch, M. J. 1994, "Orientational dependence of vicinal proton-proton NMR coupling constants: The Karplus relationship", *Concepts in Magnetic Resonance*, vol. 6, no. 1, pp. 41-56.

Mitra, A., Seaton, P. J., Assarpour, R. A., & Williamson, T. 1998, "Unprecedented concentration dependent chemical shift variation in ^{1}H NMR studies: A caveat in the investigations of molecular recognition and structure elucidation", *Tetrahedron*, vol. 54, no. 51, pp. 15489-15498.

Nabuurs, S. B., Nederveen, A. J., Vranken, W., Doreleijers, J. F., Bonvin, A. M. J. J., Vuister, G. W., Vriend, G., & Spronk, C. A. E. M. 2004, "DRESS: A database of REfined solution NMR structures", *Proteins: Structure, Function, and Bioinformatics*, vol. 55, no. 3, pp. 483-486.

Nicholas, M. P., Eryilmaz, E., Ferrage, F., Cowburn, D., & Ghose, R. 2010, "Nuclear spin relaxation in isotropic and anisotropic media", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 57, no. 2, pp. 111-158.

Nicholson, G., Rantalainen, M., Li, J. V., Maher, A. D., Malmodin, D., Ahmadi, K. R., Faber, J. H., Barrett, A., Min, J. L., Rayner, N. W., Toft, H., Krestyaninova, M., Viksna, J., Neogi, S. G., Dumas, M. E., Sarkans, U., Donnelly, P., Illig, T., Adamski, J., Suhre, K., Allen, M., Zondervan, K. T., Spector, T. D., Nicholson, J. K., Lindon, J. C., Baunsgaard, D., Holmes, E., McCarthy, M. I., Holmes, C. C., & The MolPAGE Consortium 2011, "A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection", *PLoS Genetics*, vol. 7, no. 9, p. e1002270.

Nicholson, J. K., Foxall, P. J., Spraul, M., Farrant, R. D., & Lindon, J. C. 1995, "750 MHz ^{1}H and ^{1}H-^{13}C NMR spectroscopy of human blood plasma", *Analytical Chemistry*, vol. 67, no. 5, pp. 793-811.

Nord, L. I., Vaag, P., & Duus, J. Ø. 2004, "Quantification of organic and amino acids in beer by ^{1}H NMR spectroscopy", *Analytical Chemistry*, vol. 76, no. 16, pp. 4790-4798.

Pan, J. W., Hamm, J. R., Rothman, D. L., & Shulman, R. G. 1988, "Intracellular pH in human skeletal-muscle by ^{1}H NMR", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 21, pp. 7836-7839.

Pascal, R. A., Grossman, R. B., & Van Engen, D. 1987, "Synthesis of in-[34,10][7]metacyclophane: Projection of an aliphatic hydrogen toward the center of an aromatic ring", *Journal of the American Chemical Society*, vol. 109, no. 22, pp. 6878-6880.

Pauli, G. F., Jaki, B. U., & Lankin, D. C. 2004, "Quantitative ^{1}H NMR: Development and potential of a method for natural products analysis", *Journal of Natural Products*, vol. 68, no. 1, pp. 133-149.

Pauli, G. F., Jaki, B. U., & Lankin, D. C. 2007, "A routine experimental protocol for qHNMR illustrated with taxol", *Journal of Natural Products*, vol. 70, no. 4, pp. 589-595.

Pauli, G. F., Gödecke, T., Jaki, B. U., & Lankin, D. C. 2012, "Quantitative ^{1}H NMR. Development and potential of an analytical method: An update", *Journal of Natural Products*, vol. 75, no. 4, pp. 834-851.

Peterson, D. J. & Loening, N. M. 2007, "QQ-HSQC: a quick, quantitative heteronuclear correlation experiment for NMR spectroscopy", *Magnetic Resonance in Chemistry*, vol. 45, no. 11, pp. 937-941.

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J. G., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., Greiner, R., McManus, B., Newman, J. W., Goodfriend, T., & Wishart, D. S. 2011, "The human serum metabolome", *Plos One*, vol. 6, no. 2, e16957.

Quirt, A. R. & Martin, J. S. 1971, "NMR spectra of symmetric molecules. I. The spin Hamiltonian for twofold symmetry", *Journal of Magnetic Resonance (1969)*, vol. 5, no. 3, pp. 318-327.

Rabenstein, D. L. & Isab, A. A. 1982, "Determination of the intracellular pH of intact erythrocytes by ^{1}H NMR spectroscopy", *Analytical Biochemistry*, vol. 121, no. 2, pp. 423-432.

Reily, M. D., Robosky, L. C., Manning, M. L., Butler, A., Baker, J. D., & Winters, R. T. 2006, "DFTMP, an NMR reagent for assessing the near-neutral pH of biological samples", *Journal of the American Chemical Society*, vol. 128, no. 38, pp. 12360-12361.

Rizzo, V. & Pinciroli, V. 2005, "Quantitative NMR in synthetic and combinatorial chemistry", *Journal of Pharmaceutical and Biomedical Analysis*, vol. 38, no. 5, pp. 851-857.

Rothman, D. L., Behar, K. L., Prichard, J. W., & Petroff, O. A. C. 1997, "Homocarnosine and the measurement of neuronal pH in patients with epilepsy", *Magnetic Resonance in Medicine*, vol. 38, no. 6, pp. 924-929.

Ruan, K., Yang, S. T., Van Sant, K. A., & Likos, J. J. 2009, "Application of Hadamard spectroscopy to automated structure verification in high-throughput NMR", *Magnetic Resonance in Chemistry*, vol. 47, no. 8, pp. 693-700.

Rubtsov, D. V., Jenkins, H., Ludwig, C., Easton, J., Viant, M. R., Günther, U., Griffin, J. L., & Hardy, N. 2007, "Proposed reporting requirements for the description of NMR-based metabolomics experiments", *Metabolomics*, vol. 3, no. 3, pp. 223-229.

Rundlöf, T., Mathiasson, M., Bekiroglu, S., Hakkarainen, B., Bowden, T., & Arvidsson, T. 2010, "Survey and qualification of internal standards for quantification by ^{1}H NMR spectroscopy", *Journal of Pharmaceutical and Biomedical Analysis*, vol. 52, no. 5, pp. 645-651.

Sandusky, P., Appiah-Amponsah, E., & Raftery, D. 2011, "Use of optimized 1D TOCSY NMR for improved quantitation and metabolomic analysis of biofluids", *Journal of Biomolecular NMR*, vol. 49, no. 3, pp. 281-290.

Schaefer, T. 2007, "Stereochemistry and long range coupling constants," in *Encyclopedia of Magnetic Resonance*, John Wiley & Sons, Ltd.

Sitkoff, D. & Case, D. A. 1998, "Theories of chemical shift anisotropies in proteins and nucleic acids", *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 32, no. 2, pp. 165-190.

Soininen, P., Haarala, J., Vepsäläinen, J., Niemitz, M., & Laatikainen, R. 2005, "Strategies for organic impurity quantification by ^{1}H NMR spectroscopy: Constrained total-line-shape fitting", *Analytica Chimica Acta*, vol. 542, no. 2, pp. 178-185.

Soininen, P., Kangas, A. J., Würtz, P., Tukiainen, T., Tynkkynen, T., Laatikainen, R., Järvelin, M. R., Kähönen, M., Lehtimäki, T., Viikari, J., Raitakari, O. T., Savolainen, M. J., & Ala-Korpela, M. 2009, "High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism", *Analyst*, vol. 134, no. 9, pp. 1781-1785.

Stephenson, D. S. & Binsch, G. 1980, "Automated analysis of high-resolution NMR spectra. 2. Illustrative applications of the computer program Davins", *Journal of Magnetic Resonance*, vol. 37, no. 3, pp. 409-430.

Stoyanova, R., Nicholls, A. W., Nicholson, J. K., Lindon, J. C., & Brown, T. M. 2004, "Automatic alignment of individual peaks in large high-resolution spectral data sets", *Journal of Magnetic Resonance*, vol. 170, no. 2, pp. 329-335.

Suhre, K., Wallaschofski, H., Raffler, J., Friedrich, N., Haring, R., Michael, K., Wasner, C., Krebs, A., Kronenberg, F., Chang, D., Meisinger, C., Wichmann, H. E., Hoffmann, W., Völzke, H., Völker, U., Teumer, A., Biffar, R., Kocher, T., Felix, S. B., Illig, T., Kroemer, H. K., Gieger, C., Römisch-Margl, W., & Nauck, M. 2011, "A genome-wide association study of metabolic traits in human urine", *Nature Genetics*, vol. 43, no. 6, pp. 565-569.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W. M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily, M. D., Thaden, J. J., & Viant, M. R. 2007, "Proposed minimum reporting standards for chemical analysis", *Metabolomics*, vol. 3, no. 3, pp. 211-221.

Swalen, J. D. & Reilly, C. A. 1962, "Analysis of complex NMR spectra. An iterative method", *The Journal of Chemical Physics*, vol. 37, no. 1, pp. 21-29.

Szakács, Z. & Hägele, G. 2004, "Accurate determination of low pK values by ^{1}H NMR titration", *Talanta*, vol. 62, no. 4, pp. 819-825.

Szakács, Z., Hägele, G., & Tyka, R. 2004, "^{1}H/^{31}P NMR pH indicator series to eliminate the glass electrode in NMR spectroscopic pK_a determinations", *Analytica Chimica Acta*, vol. 522, no. 2, pp. 247-258.

Szyperski, T. 1998, "^{13}C NMR, MS and metabolic flux balancing in biotechnology research", *Quarterly Reviews of Biophysics*, vol. 31, no. 1, pp. 41-106.

Thiele, H., McLeod, G., Niemitz, M., & Kühn, T. 2011, "Structure verification of small molecules using mass spectrometry and NMR spectroscopy", *Monatshefte für Chemie / Chemical Monthly*, vol. 142, no. 7, pp. 717-730.

Torgrip, R., Lindberg, J., Linder, M., Karlberg, B., Jacobsson, S., Kolmert, J., Gustafsson, I., & Schuppe-Koistinen, I. 2006, "New modes of data partitioning based on PARS peak alignment for improved multivariate biomarker/biopattern detection in ^{1}H NMR spectroscopic metabolic profiling of urine", *Metabolomics*, vol. 2, no. 1, pp. 1-19.

Tormena, C. F., Rittner, R., Contreras, R. H., & Peralta, J. E. 2004, "Anomeric effect on geminal and vicinal *J*(HH) NMR coupling constants", *Journal of Physical Chemistry A*, vol. 108, no. 38, pp. 7762-7768.

Traficante, D. D. 2007, "Relaxation: An introduction," in *Encyclopedia of Magnetic Resonance*, John Wiley & Sons, Ltd.

Trygg, J., Holmes, E., & Lundstedt, T. 2006, "Chemometrics in metabonomics", *Journal of Proteome Research*, vol. 6, no. 2, pp. 469-479.

Tuttle, T., Gräfenstein, J., & Cremer, D. 2004, "Analysis of the NMR through-space coupling mechanism between ^{19}F atoms", *Chemical Physics Letters*, vol. 394, no. 1-3, pp. 5-13.

Tynkkynen, T., Hassinen, T., Tiainen, M., Soininen, P., & Laatikainen, R. 2012, "^{1}H NMR spectral analysis and conformational behavior of *n*-alkanes in different chemical environments", *Magnetic Resonance in Chemistry*, vol. 50, no. 9, pp. 598-607.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., & Markley, J. L. 2008, "BioMagResBank", *Nucleic Acids Research*, vol. 36, no. suppl 1, pp. D402-D408.

Vermathen, P., Capizzano, A. A., & Maudsley, A. A. 2000, "Administration and ^{1}H MRS detection of histidine in human brain: Application to *in vivo* pH measurement", *Magnetic Resonance in Medicine*, vol. 43, no. 5, pp. 665-675.

Vidossich, P., Piana, S., Miani, A., & Carloni, P. 2006, "Deuterium isotope effects in A : T and A : U base pairs: A computational NMR study", *Journal of the American Chemical Society*, vol. 128, no. 22, pp. 7215-7221.

Weber, U. & Thiele, H. 1998, *NMR-Spectroscopy: Modern Spectral Analysis* Wiley-VCH, Weinheim.

Weljie, A. M., Newton, J., Mercier, P., Carlson, E., & Slupsky, C. M. 2006, "Targeted profiling: Quantitative analysis of ^{1}H NMR metabolomics data", *Analytical Chemistry*, vol. 78, no. 13, pp. 4430-4442.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L. Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Miller, V., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R. L., Tatusova, T. A., Wagner, L., & Yaschenko, E. 2006, "Database resources of the National Center for Biotechnology Information", *Nucleic Acids Research*, vol. 35, no. suppl 1, pp. D5-D12.

Wider, G. & Dreier, L. 2006, "Measuring protein concentrations by NMR spectroscopy", *Journal of the American Chemical Society*, vol. 128, no. 8, pp. 2571-2576.

Wishart, D. S. 2008, "Quantitative metabolomics using NMR", *TRAC-Trends in Analytical Chemistry*, vol. 27, no. 3, pp. 228-237.

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., & Woolsey, J. 2006, "DrugBank: a comprehensive resource for *in silico* drug discovery and exploration", *Nucleic Acids Research*, vol. 34, no. suppl 1, pp. D668-D672.

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M. A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., MacInnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B. D., Vogel, H. J., & Querengesser, L. 2007, "HMDB: the human metabolome database", *Nucleic Acids Research*, vol. 35, pp. D521-D526.

Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J. G., Jia, L., Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y. P., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J., & Forsythe, I. 2009, "HMDB: A knowledgebase for the human metabolome", *Nucleic Acids Research*, vol. 37, pp. D603-D610.

Wu, A., Gräfenstein, J., & Cremer, D. 2003, "Analysis of the transmission mechanism of NMR spin-spin coupling constants using Fermi contact spin density distribution, partial spin polarization, and orbital currents: XH_n molecules", *The Journal of Physical Chemistry A*, vol. 107, no. 36, pp. 7043-7056.

Wu, J. F., An, Y. P., Yao, J. W., Wang, Y. L., & Tang, H. R. 2010, "An optimised sample preparation method for NMR-based faecal metabonomic analysis", *Analyst*, vol. 135, no. 5, pp. 1023-1030.

Xu, Q. W., Sachs, J. R., Wang, T. C., & Schaefer, W. H. 2006, "Quantification and identification of components in solution mixtures from 1D proton NMR spectra using singular value decomposition", *Analytical Chemistry*, vol. 78, no. 20, pp. 7175-7185.

Zhang, F., Bruschweiler-Li, L., Robinette, S. L., & Brüschweiler, R. 2008, "Self-consistent metabolic mixture analysis by heteronuclear NMR. Application to a human cancer cell line", *Analytical Chemistry*, vol. 80, no. 19, pp. 7549-7553.

Zhang, F. & Brüschweiler, R. 2007, "Robust deconvolution of complex mixtures by covariance TOCSY spectroscopy", *Angewandte Chemie International Edition*, vol. 46, no. 15, pp. 2639-2642.

Zubkov, S. V., Golotvin, S. S., & Chertkov, V. A. 2002, "VALISA: A new procedure for total lineshape analysis of NMR spectra. Conformational analysis of trans-1,2-dibromocyclopentane", *Russian Chemical Bulletin*, vol. 51, no. 7, pp. 1222-1230.

Mika Tiainen

Quantitative Quantum Mechanical Analysis of ^{1}H NMR Spectra

Applications and Strategies

Nuclear magnetic resonance (NMR) spectroscopy is widely used for profiling of a variety of complex biological samples. However, accurate quantification of compounds from a proton NMR spectrum of biological sample is a demanding task due to spectral complexity and overlap. The protocols and tools developed in this thesis, Adaptive Spectral Library (ASL) and quantitative Quantum Mechanical Spectral Analysis (qQMSA), enable accurate, robust and cost-effective way to quantify individual components from the proton NMR spectra of complex mixtures.

UNIVERSITY OF
EASTERN FINLAND

Publications of the University of Eastern Finland
Dissertations in Health Sciences

ISBN 978-952-61-1076-9