Fast 2D NMR spectroscopy for complex mixtures

Adrien Le Guennec

École Doctorale de l'X

Spécialité Chimie analytique

Année 2015 N° :

Thèse

Pour l'obtention du grade de Docteur

par

Adrien Le Guennec

Titre de la thèse:

Fast 2D NMR spectroscopy for complex mixtures

Soutenance le 17 Juillet 2015

Composition du jury:

Mme Corinne GOSMINI	Chercheur - Ecole Polytechnique, Palaiseau	Présidente
M. Burkhard LUY	Professeur - Karlsruhe Institute of Technology (KIT)	Rapporteur
M. Nicolas GIRAUD	Maître de Conférences - Université Paris-Sud	Rapporteur
M. Dominique ROLIN	Professeur - INRA/Université de Bordeaux	Examinateur
M. Jean-Nicolas DUMEZ	Chargé de recherche - ICSN, Gif-sur-Yvette	Examinateur
M. Stefano CALDARELLI	Professeur - ICSN, Gif-sur-Yvette	Directeur de thèse
M. Patrick GIRAUDEAU	Maître de Conférences - Université de Nantes	Co-directeur de thèse

Contents

Remerciements

Cette thèse a été effectuée au sein de l'équipe 64 du laboratoire de l'ICSN, à Gif-sur-Yvette, ainsi qu'à l'École Polytechnique de Palaiseau et le laboratoire CEISAM, à l'Université de Nantes. Je remercie le directeur de l'ICSN, **Max Malacria**, récemment remplacé par **Angela Marinetti**, ainsi que le chef de l'équipe du groupe RMN, **Éric Guittet**, pour m'avoir accueilli dans l'ICSN.

D'un point de vue professionnel, je souhaiterais remercier mes deux directeurs de thèse, **Stefano Caldarelli** et **Patrick Giraudeau**, pour m'avoir donné ma chance pour cette thèse. Leur disponibilité, même pour les détails les plus triviaux, a été très appréciée durant ces 3 années de thèse. Je voudrais également remercier d'anciens post-docs de Stefano Caldarelli; **Robert Evans** et **Manjunatha Reddy,** qui m'ont également aidé durant le début de ma thèse afin de m'habituer à certains paramètres de la RMN et de Topspin. **Jean-Nicolas Dumez**, qui est arrivé à l'ICSN au milieu de ma thèse, a été également d'une grande aide durant cette 2ème moitie de thèse.

Dans le laboratoire de l'ICSN, je souhaiterais également remercier tous les membres de l'équipe 64: **Ewen Lescop, Nelly Morellet, François Bontems, Christina Sizun, Carine Van-Heijenoort, Nadine Assrir, Éric Jaquet, Naima Nhiri Marion André, Yasmine Skanji, Gwladys Rivière, Arthur Besle, Séverine Moriau, Celia Deville, Safa Lassoued, Nelson Pereira, Ying-hui Yang, Oriane Frances, Prishila Ponien**, **Annie Moretto et Alda Da Costa**, avec qui j'ai pu avoir plusieurs conversations intéressantes qui m'ont aidé durant cette thèse, à la fois pour les recherches et pour la vie quotidienne du laboratoire.

Au laboratoire CEISAM, je remercie également tous les membres de l'équipe EBSI, tout particulièrement **Illa Tea**, qui a été respectivement encadrante et co-encadrante de mes deux premiers stages de laboratoire. Je souhaiterais également remercier **Ingrid Antheaume, Benoît Charrier, Virginie Silvestre, Denis Loquet, Estelle Martineau, Serge Akoka, Gérald Rémaud, Richard Robins, Meerakhan Pathan, Maxime Julien, Kévin Bayle, Ugo Bussy, François Kouamé, Didier Diomandé et Renaud Boisseau** pour leur conseils et leur aide, que ce soit durant les stages ou durant la thèse. Je souhaite également bonne chance à **Boris Gouilleux et Laetitia Rouger**, qui entament leur thèse.

Je souhaiterais également remercier l'équipe 09 de l'ICSN, dirigée par **Didier Stien** et **Véronique Épervier**, pour nous avoir laissé utiliser leurs laboratoire durant le déménagement du groupe de Stefano Caldarelli de l'École Polytechnique vers l'ICSN, ainsi que **Alain-Louis Joseph**, de l'école Polytechnique, qui m'a également permis de mieux comprendre comment fonctionne la RMN.

D'un point de vue personnel, en plus de toutes les personnes cités précédemment, je souhaiterais également remercier ma famille, tout particulièrement ma mère **Véronique**, ma sœur **Julie** et ma tante **Corinne**, qui m'ont permis de rendre ces 3 années de thèse agréables, que ce soit durant les moments positifs tout comme les moments frustrants de la thèse.

Notation and abbreviations

1D	One-dimensional
^{1}H	Proton
2D	Two-dimensional
ASAP	Acceleration by Sharing Adjacent Polarization
B_1	Radio-frequency pulses
BEST	Band-selective Excitation Short-Transient
COSY	COrrelation SpectroscopY
CPMG	Carr-Purcell-Meiboom-Gill
CS	Compressed Sensing
CV	Coefficiant of Variation
CV-ANOVA	Cross-Validated ANalysis Of VAriance
DANS	Differential Analysis by 2D NMR Spectroscopy
DIPSI	Decoupling In the Presence of Scalar Interactions
DNP	Dynamic Nuclear Polarization
DISSECT	DIagonal-Suppressed Spin-Echo Correlation specTroscopy
DoE	Design of Experiments

DOSY	Diffusion Ordered Spectroscopy
DQ	Double Quantum
DQF	Double Quantum Filter
DQS	Double Quantum Spectroscopy
DQSSY	Double Quantum Spectroscopy SYmmetrized
DSE	Double Spin Echo
ES	Excitation Sculpting
FD/SD	Fold Difference to Standard Deviation
FID	Free Induction Decay
FT	Fourier Transformation
GARP	Globally optimized Alternating phase Rectangular Pulse
GUI	Graphical User Interface
HMBC	Heteronuclear Multiple-Bond Correlation
HMDB	Human Metabolome DataBase
HMQC	Heteronuclear Multiple-Quantum Correlation
HR-MAS	High-Resolution Magic Angle Spinning
HSQC	Heteronuclear Single-Quantum Correlation

$HSQC_0$	Time zero Heteronuclear Single-Quantum Correlation
IFT	Inverse Fourier Transformation
ILT	Inverse Laplace Transformation
IMPACT HMBC	IMProved and Accelerated Constant-Time HMBC
INADEQUATE	Incredible Natural Abundance DoublE QUAntum Transfer Experiment
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
IR	Inversion Recovery
J-RES	J-RESolved
J_{C-H}	Carbon-Proton J-coupling
J_{H-H}	Proton-Proton J-coupling
M3S	Multi-Shot Single-Scan
MaxQ	Maximum-Quantum
MDD	Multi-Dimensional Decomposition
MQ	Multiple Quantum
MRI	Magnetic Resonance Imaging
MRSI	Magnetic Resonance Spectroscopic Imaging
N_1	Number of t_1 increments

NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NUS	Non-Uniform Sampling
OPLS-DA	Orthogonal Partial Least Squares-Differential Analysis
PCA	Principal Component Analysis
PGFSE	Pulse Gradient Field Spin Echo
PROJECT	Periodic Refocusing Of J Evolution by Coherence Transfer
PURGE	Presaturation Utilizing Relaxation Gradients and Echoes
Q-OCCA HSQC	Quantitative, Offset-Compensated, CPMG adjusted HSQC
R^2	Squared Pearson coefficient
SE	Spin Echo
SEDUCE	SElective Decoupling Using Crafted Excitation
SMART	SMAll Recovery Times
SNR	Signal-to-Noise Ratio
SNR_{max}	Signal-to-maximum Noise Ratio
SNR_{RMS}	Signal-to-Root-Mean-Square-Noise Ratio

SO-FAST	band-Selective Optimized Flip-Angle Short-Transient
SQ	Single Quantum
T_1	Longitudinal relaxation constant
T_2	Transversal relaxation constant
T_2*	Transversal relaxation constant with corrections for field inhomogeneities
TOCSY	TOtal Correlation Spectroscopy
TPPI	Time-Proportional Phase Incrementation
TR	Repetition time
UF	UltraFast
US	Uniform sampling
UV	Unit Variance
VIP	Variable Importance for the Projection
WATERGATE	WATER suppression by GrAdient-Tailored Excitation
WET	Water Suppression Enhanced through T_1 effects
zf	z-filter
ZQ	Zero Quantum

Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the main analytical tools for metabolomics studies, along with mass spectroscopy, thanks to its high reproducibility and because it is a powerful tool to unambiguously identify metabolites.[1, 2] NMR has become a common tool for metabolomics, thanks to a standardized strategy for sample preparation and analysis, based on one-dimensional (1D) proton (^{1}H) spectra[3] associated with statistical analysis.[4] However, 1D ^{1}H spectra of complex mixtures, like biological extracts, are characterized by extensive overlap that can conceal part of the information from potential biomarkers.[5] This extensive overlap also severely hinders absolute quantification, which is another important step for metabolomics studies.[6]

Alternative strategies have been used to limit the problem of overlap in 1D spectra. The use of other nuclei, like carbon-13 (^{13}C) have been shown to reduce greatly the problem of overlap, which gives an easier access to the information from most visible metabolites, as shown by the work from Wei et al.[7] and Clendinen et al.[8], but this reduction of overlap comes at the price of the experimental time, since ^{13}C at natural abundance is far less sensitive compared to ^{1}H. Another possibility is the deconvolution of 1D ^{1}H spectra, initiated by the work from Weljie et al.,[9] where the 1D spectrum is separated in different subspectra, each corresponding to the separate 1D spectrum of a metabolite,. But this analysis requires a good preliminary knowledge of the metabolites visible in the spectrum, in order to have a library of 1D spectra of individual metabolites recorded with conditions as close as possible to the conditions for the biological samples.

The last known possibility, which is explored in this thesis, is two-dimensional (2D) NMR. This technique, introduced in 1971 by Jeener[10] and first applied in 1976 by Aue, Bartholdi and Ernst,[11] makes it possible to spread peaks along two orthogonal frequency dimensions. Along with the lower risk of severe overlap, one of the most useful properties of 2D NMR is the added structural information, which makes this technique one of the cornerstones for structural elucidation. In addition, there is a great number of existing pulse sequences that allows the exploration of several types of connectivity, like J-couplings, Nuclear Overhauser Effects (NOE) or chemical exchange.

Since the introduction of NMR in metabolomics, several studies have been carried out to show the usefulness of 2D pulse sequences for metabolomics. Four pulse sequences in particular have been used for these studies. The first one is homonuclear J-RESolved (J-RES) spectroscopy, where J-couplings and chemical shifts are separated in two different dimensions. Since, in many cases, the overlap is caused by the J-couplings in 1D ^{1}H spectra, the separation makes it possible to reduce the overlap in the J-RES spectrum, even with biological samples.[12] The acquisition of J-RES spectra has been optimized, by reducing the experimental time to levels close to 1D ^{1}H spectra with close sensitivity, in particular by the work of Viant and Ludwig.[13] But the limited spectral width in the indirect dimension does not completely avoid the risk of overlap, artifacts from strong couplings can plague the J-RES spectrum[14] and errors in quantification can arise because of partial cancellation.[15]

The Correlation SpectroscopY (COSY) and the TOtal Correlation SpectroscopY (TOCSY) pulse sequences are two others pulse sequences giving similar spectra, where connectivity between protons is shown via J-couplings. But while COSY spectra show protons with direct J-couplings,[11] TOCSY spectra show extended networks of J-couplings.[16] So far, COSY in complex mixtures has been used in a recent strategy from the group of Schroeder at Cornell University, called Differentiation Analysis by Nuclear Spectroscopy, in order to discover more easily bioactive molecules in complex mixtures.[17] For the TOCSY pulse sequence, a study by Van et al.[18] has shown the potential of this pulse sequence to discover important metabolites by statistical analysis that would be difficult to identify by 1D ^{1}H NMR. Band selective TOCSY has also been used for metabolomics studies[19] and a variant has been used by the groups of Portais[20, 21] and Markley[22] to understand fluxes by ^{13}C enrichment. While these spectra usually have little overlap and are sensitive, one of their main shortcomings is the loss of information for some metabolites, particularly those without J_{H-H} couplings.

Finally, the HSQC pulse sequence shows correlations between ^{1}H and ^{13}C (or ^{15}N, depending on the system studied).[23] This is one of the most used 2D pulse sequence for complex mixture analysis,[24-26] since the ^{13}C dimension can give a higher spread of peaks and therefore reduces the risk of overlap. Studies by Simpson's group have also shown that 2D HSQC is a good complement to 1D ^{1}H spectra.[27-29] However, at natural abundance, HSQC is the less sensitive of these four 2D pulse sequences but suffers from the low natural abundance of ^{13}C. This disadvantage can however be decreased by ^{13}C enrichment, as long as it is commercially possible with the system under study, as shown for several studies by Kikuchi's group.[30, 31]

As seen from the description above, all the pulse sequences have strengths and weaknesses. But most of them share a similar problem: the experimental time for 2D spectra is much longer than 1D ^{1}H spectra because of the need to repeat hundreds of times the pulse sequence in order to construct the Free Induction Decay (FID) in the indirect dimension. In order to reduce the experimental time of 2D spectra, several approaches have been developed.

Two of them, in particular, have shown promises for reduction of 2D spectra of complex mixtures. The first one is Non-Uniform Sampling (NUS), which can reduce the number of t_1 increments by changing the method to acquire the indirect FID as well as the algorithm to reconstruct the 2D spectrum.[32] Recent studies by Rai and Sinha,[33] as well as Hyberts and al.,[34] have shown that NUS can be used for complex mixtures to reduce the experimental times needed to record 2D spectra of complex mixtures, both for identification (with the use of very high resolution spectra) and quantification purposes (by reducing the acquisition time). The other approach, called UltraFast (UF) 2D NMR, and proposed by Frydman and co-workers[35] is the only current approach that makes it possible to record a 2D spectrum in only one scan.[36] Despite compromises to make between spectral widths, resolution and sensitivity,[37] improvements have made possible the quantification of metabolites in complex mixtures with concentrations as low as a few hundreds of µM with experimental times of 20 min.[38]

At the beginning of this thesis, the pre-cited fast 2D NMR methods had not been evaluated for high-throughput metabolomics applications. Therefore, the goal of this thesis was to make a first evaluation of these fast 2D approaches for high-throughput metabolomics. In this manuscript, the main notions for the understanding of the subject will be introduced first, while the next chapters describe the strategies developed to reach this goal.

We made a first evaluation of these fast methods for untargeted metabolomics, using synthetic mixtures for metabolomics studies with 1D and 2D spectra. This study was separated in two parts: the first part was focused on the differences between 1D ^{1}H spectra and conventional 2D spectra with three different pulse sequences (COSY, J-RES and HSQC) by using the statistical analysis of concentrations as a model. This study showed that the statistical analysis of 2D spectra was closer to the statistical analysis of concentrations than the statistical analysis of 1D spectra, confirming the usefulness of 2D spectra for metabolomics. Then the same comparison was made for conventional 2D NMR and fast 2D NMR, with the UF variant of COSY and the NUS variant of DQF-COSY and HSQC. These three variants were indeed compatible with metabolomics studies, since they made it possible to retrieve the same information as the conventional 2D counterparts, with a reduced experimental time.

Then, methodological developments were carried out in order to increase the amount of information available from fast 2D spectra. First, a new UF pulse sequence was developed based on double-quantum spectroscopy (DQS), in order to suppress diagonal peaks that can overlap with some correlation peaks. This sequence was tested with a sucrose sample and was compared to the UF COSY spectrum. Then another variant of UF DQS was tested to produce UF COSY-like spectra without diagonal peaks. And finally these two pulse sequences were tested with a sample of several metabolites in H_2O/D_2O (90/10) to simulate biological extracts. A second optimization was done with the NUS approach. The use of time-equivalent NUS was performed with its two variants, sensitivity-enhanced NUS and resolution-enhanced NUS. For resolution-enhanced NUS, an analytical evaluation was done to verify the sensitivity, the repeatability and the noise behavior, compared to conventional 2D NMR for two pulse sequences, HSQC and TOCSY. Finally, new tools were adapted for the measurement of T_1 and T_2, which are important parameters in NMR since they impact the signal-to-noise ratio per unit of time. For that, a Diffusion-Ordered SpectroscopY (DOSY) Toolbox was modified for the calculation of T_1 and T_2 and this toolbox was tested for simple samples, followed by a complex mixture in H_2O/D_2O (90/10). For the last sample, a solvent suppression scheme was tested for good water suppression without changing the accuracy of T_1/T_2 measurements.

Part 1. Literature survey

The goal of this chapter is to introduce the main concepts, which served as a basis for the present work. The limitations of 1D NMR for metabolomics are first discussed. The mechanism of several 2D pulse sequences and the resulting 2D spectra are then analyzed, in order to highlight their potential for metabolomics. Finally, the alternatives to conventional 2D acquisition are presented, with their pros and cons. Existing results of 2D NMR and fast 2D NMR for the analysis of complex mixtures will be also presented.

In this chapter, the general knowledge about multi-pulse 1D NMR will be assumed to be already known. This general knowledge includes the Zeeman effect, the effects of radiofrequency pulses on magnetization, the chemical shift, the J-coupling, nuclear spin relaxation, the vector model and the coherence transfer pathway model to explain the purpose of several pulse sequences and the effect of spin echo and INEPT building block on magnetization. For more details about these parts, the reader should refer to other references.[39, 40] The use of product operators to explain the effect of two-dimensional (2D) pulse sequences is not indispensable, but can help to better understand the pulse sequences and their variants.

1. 1D NMR for metabolomics: principles and limitations

1.1. Metabolomics: goal and main tools

Metabolomics is a part of the 'omics analytical tools since at least 15 years, after the sequencing of the human genome [41] has shown its limits to fully understand the function of each gene.[42] Its goal is the global understanding of the metabolome, *i.e.* the exhaustive list of small molecules participating in chemical reactions in an organism. Nowadays, metabolomics is considered one of the main branch of the 'omics family, along with genomics and transcriptomics, which studies DNA and RNA, respectively,[43] and proteomics, which focuses on proteins.[44] Metabolomics has already shown to give complementary results to the other three 'omics techniques.[45] While potentially every organism can be studied, the more common samples used are biofluids, like plasma, serum or urine,[46, 47] and plants.[48, 49]

Because of the huge number of different chemical functions involved in organisms and the large dynamic range in concentrations, no analytical tool can achieve that goal with only one analysis. For that reason, metabolomics usually refers to any study where at least a subset of the metabolome is studied, for profiling or quantitative purposes.[6] Despite this lack of universal analysis for metabolomics, two analytical tools are most commonly used today: NMR spectroscopy and mass spectrometry, the latter usually coupled to gas chromatography or liquid chromatography.[50] These two analytical tools are the closest universal detectors currently available within the analytical tools and gives reproducible results.

A classic metabolomics study consists of obtaining samples from several groups, for example a control group and a diseased group in clinical research.[47] The difference of intensities

between spectra from each sample should reveal metabolites that vary between groups, which should bring answers about metabolic pathways perturbed by the disease. Between these two analytical tools, mass spectrometry is generally considered as more sensitive, but the intensity of each peak is dependent on their ionization and behavior during chromatography, which makes quantification difficult. NMR, on the other hand, is less sensitive, but identification and quantification are easier.[51] Finally, a sample used for mass spectrometry is destroyed, while a sample for NMR can be reused, for other NMR experiments or used for other analytical tools. Both are considered complementary techniques, with NMR for a good identification of the most abundant metabolites and mass spectrometry for detection of less concentrated metabolites.

1.2. Metabolomics with 1D NMR

1.2.1. General considerations

In most NMR studies for metabolomics, 1D proton (^{1}H) pulse sequences are used. The main reasons are the sensitivity of these spectra, the high reproducibility and because it is a primary analytical technique, which means that it is possible to measure the concentration from one experiment without the need of a calibration curve.[52, 53] Its use for metabolomics requires several steps already described more than 10 years ago,[2] and shown in figure 1.1.

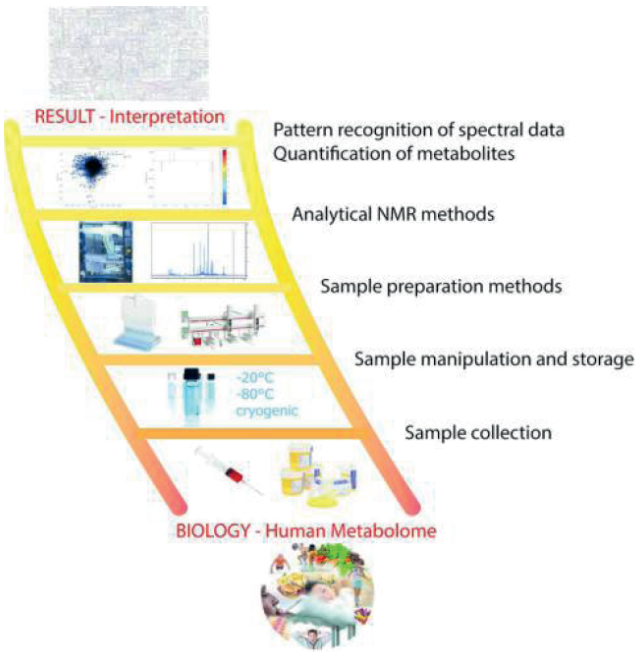


Figure 1.1. The key stages of a NMR metabolomics study in order to obtain relevant biological information. Reprinted with permission from A.C. Dona et al.[54]
Copyright 2014 American Chemical Society.

The reasoning behind these steps is to conserve all the biological information in the samples, by avoiding degradation of unstable metabolites or enzymatic activity and to limit fractionation by purification steps. Since one advantage of NMR is the possibility to analyze crude extracts without prior purification, the preparation of NMR samples are kept to a minimum, a part

that is standardized for some biofluids.[55] The protocol for NMR experiments is also highly standardized in order to harmonize results from several groups as much as possible.[3]

The extraction of intensities from the spectrum is done automatically by separating the spectrum into several regions, whose areas are used as the parameters for statistical analysis. The spectrum may be divided into regions of the same size, which is usually called binning, or regions of varying size depending on the peak size and overlap, which is usually called bucketing. However, the two terms can be used for either strategy in the literature.

While most of the steps stated above are automated, metabolomics studies are still dependent on the choice of the researcher. One of the most critical points in a metabolomics study is the sample collection. Since many factors influence the metabolome, like diet or age, the selection of patients for the study is important to limit as much as possible random modifications of metabolite concentrations between groups.[42, 56] A critical analysis of results from statistical analysis is also required to avoid possible pitfalls.[57]

Two of the main pulse sequences used for metabolomics, 1D NOESY with water presaturation (noesypr1d) and excitation sculpting, only differ by their water suppression scheme. Indeed, the water suppression is critical in 1D ^{1}H spectra for metabolomics. The main water suppression schemes will be explained in the next section, including the two schemes stated above.

1.2.2. Main statistical tools

Because of the large amount of data generated by NMR spectra of biological samples, multivariate analysis is used to extract meaningful information about the system studied. A description of the different multivariate analyses is outside the scope of this review, but the basics of one of the most used multivariate analysis, principal component analysis (PCA) will be detailed below. Once raw intensities are extracted from the spectrum (for NMR or mass spectrometry), generally by fragmenting the spectrum into several regions, two questions are especially important. The first one is to know if the data allows a clear separation between groups and if it is the case, which part of spectrum contributes the most to the separation.

With PCA, the goal is to answer these two questions with a two-dimensional plot instead of the direct observation of all regions. From the intensity of each region, a new set of parameters is created, which are linear contributions of the intensities, as explained in figure 1.2.a.[4] These parameters, called principal components, are created with the purpose of concentrating most of the total variance in the first new parameters. It is therefore possible to see most of the variance in a two-dimensional plot.

The first plot, called the score plot (figure 1.2.b), shows the position of each sample as a function of the two first principal components. It is a first start to see if the groups can be separated. The second plot (figure 1.1.c), called the loading plot, shows the coefficients of each region of the spectrum as a function of these two principal components. The bigger the coefficient, the more likely the region in question has an importance for the separation of the groups. The metabolites having the most importance in separation are the most likely to be important for the understanding of the system, and are often called biomarkers. Discovering these biomarkers is the

first step towards the understanding of the system studied and one of the main goals in metabolomics.

Other statistical tools exist to maximize the separation between groups, such as Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA). This statistical tool is also a multivariate analysis, and produces the same plots as PCA. However, the separation involved directly the inter-group variability, instead of the total variance. This difference will be expended in part 2. Several tests also exist to verify the robustness of the results and their predictability, but are not within the scope of this review.

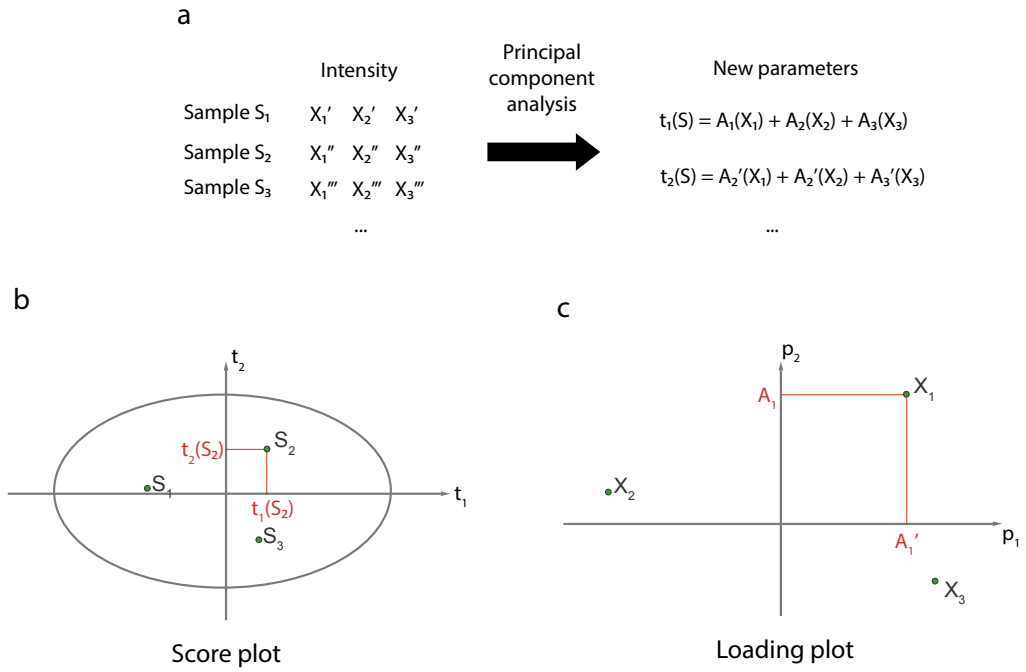


Figure 1.2. Scheme of principal component analysis (a), along with a scheme of the score plot (b) and the loading plot (c). In the score plot, the ellipse represents the 95% tolerance for outliers based on the Hotelling distribution T^2. Any point outside of this ellipse is a possible outlier, whose profile is very different of the other samples and may be excluded from the analysis.

1.3. Suppression of the solvent peak in 1D NMR of biological samples

Most of the samples in NMR are in highly protonated water, usually a mixture of H_2O/D_2O (90/10). A high concentration of salts is also typical, because of the biological sample itself or the use of a buffer to control the pH of the sample. This high concentration of protonated water will cause problems for NMR analysis if left unchecked, since 90% of protonated water means that the sample has roughly 50 M of H_2O, compared to the concentration of the metabolites, which is rarely higher than tens of mM.[46] Another problem is the "faraway" water, which is the water at the edge of the sample.[58] Since this water does not exactly experience the same magnetic field as the rest of the sample, its chemical shift is slightly different and radiofrequency pulses do not have the same effect, creating a different coherence transfer than the rest of the signal.[59] Finally, with these high concentrations, radiation damping can also occur,[60] especially with cryogenic probes commonly used for NMR studies of biological samples.[61] Radiation damping happens when the sample

magnetization (in this case, water magnetization) is coupled with the magnet magnetization, and acts as a selective pulse toward water, which returns the magnetization along the z-axis. In this case the water peak has a very short apparent T_1, causing a very short T_2* in the process and therefore a large linewidth in the spectrum.[62]

These different problems cause different shortcomings, if left unattended. Because of the large dynamic range, problems can arise during the analog to digital conversion of the NMR signal to properly represent the peaks of the metabolites. It forces to choose low receiver gains to avoid artifacts, reducing sensitivity in the process.[63] The combination of high concentrations with radiation damping means that most of the peaks in the spectrum will be altered by the water peak, by overlap or baseline distortion. The problem of "faraway" water will be important during the discussion on the different schemes to suppress water.

To avoid these problems, many schemes have been created to reduce as much as possible the water peak without altering the other peaks. The most used schemes for complex mixtures and diluted samples are represented in figure 1.3, along with two advanced schemes published recently that improve already existing schemes for cleaner spectra, especially for complex mixtures. All these schemes have pros and cons and no universal scheme exists, making each scheme more adapted for different cases, as explained in the next paragraphs.

Presaturation is one of the oldest solvent suppression schemes.[64] It relies on the continuous irradiation of one frequency during the inter-scan delay to saturate the peak with this Larmor frequency, destroying its magnetization (figure 1.3.a). This simple scheme offers great reductions of the solvent intensity, but has three main problems: it can produce lineshape distortions, depending on the intensity of the residual peak;[53] it works poorly on "faraway" water[59]; it reduces the intensity of protons in exchange with water, like the amide protons from proteins.[65] However, a recent modification of this scheme is also effective against "faraway" water.[66]

The first increment of the Nuclear Overhauser Effect SpectroscopY (NOESY) pulse sequence with presaturation during the mixing time (noesy1Dpr, figure 1.3.b) has been developed in the case of metabolomics[67] and has since then become one of the standard water suppression scheme in metabolomics,[3] since it requires little optimization.[68] Compared to presaturation alone, noesy1Dpr benefits from a better volume selection and phase cycling, which allows a good suppression of "faraway" water.[69] However, the scheme is sensitive to changes between samples (pH, salt concentrations...), which can cause different profiles for the water residual peak between samples.[70]

Water suppression Enhanced through T_1 effects (WET) has been designed to be used in localized NMR spectroscopy and imaging, where pulse power (B_1) inhomogeneities and longitudinal relaxation (T_1) of water can lead to distortions in images and differences in water suppression between spectra.[71] This scheme uses selective pulses to put only the water magnetization in the transverse plane, which is then defocused by gradients. The robustness of WET toward B_1 and T_1 effects comes from the difference in the tip angle for each selective pulse, along with the strength of each subsequent gradient (figure 1.3.c). The use of selective pulses

makes WET also adequate for liquid chromatography-NMR (LC-NMR), with the possibility to excite several frequencies simultaneously with one selective pulse.[72] Since LC-NMR often uses mixtures of protonated solvents, several peaks have to be suppressed in order to identify peaks from other molecules. However, WET alone does not seem to be effective for 1D NMR in metabolomics, since the water residual peak is still important, with lineshape distortions and poor control of "faraway" water.[70] Furthermore, it needs several optimizations (shape, spectral width and duration of the selective pulses) to eventually become efficient, which limits its use in metabolomics.

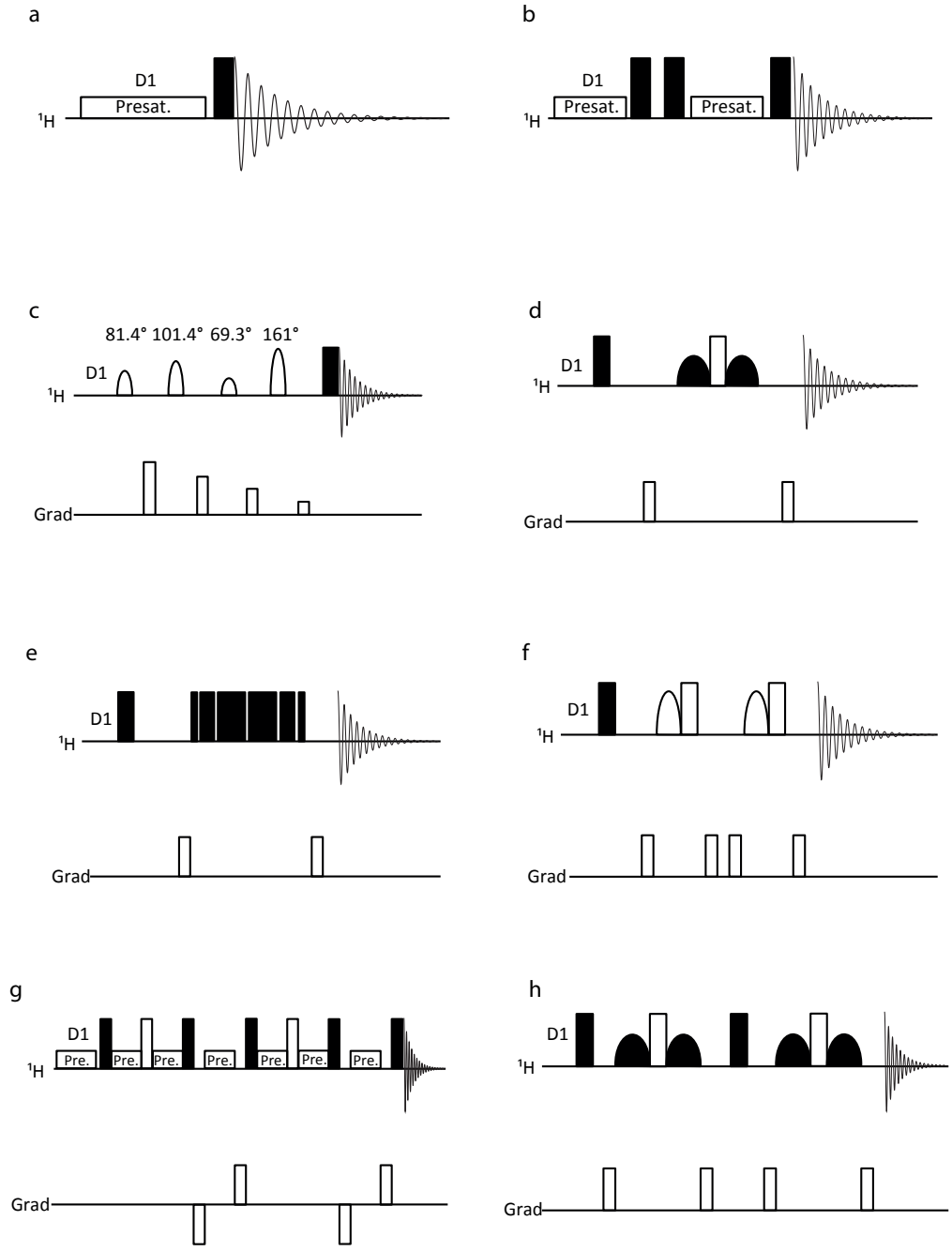


Figure 1.3. Basic solvent suppression schemes, with presaturation (a), first increment of the NOESY pulse sequence with presaturation (b), WET (c), WATERGATE with selective pulses (d) or composite pulses (e) and excitation sculpting (f). Some advent solvent suppression schemes are also shown, like PURGE (g) and "perfect echo" WATERGATE (h)

WATER suppression by GrAdient-Tailored Excitation (WATERGATE) are a series of schemes designed to change the relative angle of the water magnetization compared to those from other protons by the use of pulse field gradient spin echo (PFGSE), where an inversion is caused between two identical gradients. This can be achieved by selective pulses (figure 1.3.d)[73] or composite pulses (figure 1.3.e).[74] In both cases, the wanted coherence has been inverted, which allows it to be refocused by the last gradient, while the water peak experienced a 360° pulse as a whole, and thus is not refocused by the last gradient.

Since the water peak is not saturated, these schemes are adequate for protein NMR since they do not affect the signal of exchangeable protons. But the intensity of the nearby peaks is attenuated for each version, and the attenuation profile depends on the length of the selective pulses (selective WATERGATE) or the delay between pulses (composite WATERGATE). Furthermore, since the magnetization is inverted, WATERGATE is equivalent to a spin echo for this magnetization, causing the build up of anti-phase magnetization. Peak distortion is therefore found in the spectra, which can complicate the identification of the peaks in 1D spectra and limits its use for quantitative analysis. Baseline distortion due to the residual water peak also occurs, which can be suppressed by a double PGFSE, also called Excitation Sculpting (ES, figure 1.3.f).[75] While a shaped and a hard 180° pulses are used as a spin echo in this figure, any spin echo already used in WATERGATE can be used in ES. ES has been shown to give very reproducible water residual peaks and needs little optimization,[70] which makes this scheme adequate for metabolomics.

Finally, two advanced water suppression techniques are also shown. Presaturation Using Relaxation Gradients and Echoes (PURGE, figure 1.3.g)[76] is an extension of the noesy1Dpr with a spin echo between the first and the second pulse and presaturation between delays. Gradients are also added during the mixing time and the scheme is repeated for better phase properties, similarly to double PGFSE. "Perfect" spin echo WATERGATE (figure 1.3.h) is an extension of ES where a 90° pulse is introduced between the two PGFSE.[77] Any antiphase magnetization created during the first PGFSE is transferred by the 90° pulse and is reverted to in-phase magnetization by the second PGFSE, which is the mode of action for each "perfect" spin echo.[78] It can also be added that combination of schemes are possible, in order to suppress further the water peak.[59]

All these schemes are used in 1D ^{1}H spectra and can be integrated in two-dimensional pulse sequences or other multi-impulse pulse sequence.

1.4. Measurement of T_1/T_2 parameters

The longitudinal relaxation time (T_1) and transverse relaxation time (T_2) are two important parameters in NMR that can be used to set the inter-scan delay and the acquisition time, respectively. The knowledge of the relaxation times can be used to optimize both parameters for optimal signal-to-noise ratio (SNR) per unit for time. The knowledge of T_1 is also critical for quantitative 1D NMR, since the area of a peak is dependent on T_1 if complete relaxation has not occurred between scans. A common value is five times the longest T_1 to obtain an error of less than 1% because of incomplete relaxation.[52, 79]

For each relaxation time, several pulse sequences exist for its calculation. For T_1, the main pulse sequence is inversion-recovery (figure 1.4.a),[80] where the magnetization is first inverted by a 180° pulse. A delay allows relaxation of the magnetization, then a 90° pulse is used to tip the magnetization in the transverse plane. By varying the delay between the 180° and the 90° pulse, one can obtain a series of spectra with varying intensity depending on the T_1 parameter. Another possible pulse sequence is saturation-recovery. More details about T_1 measurements will be shown in part 6.

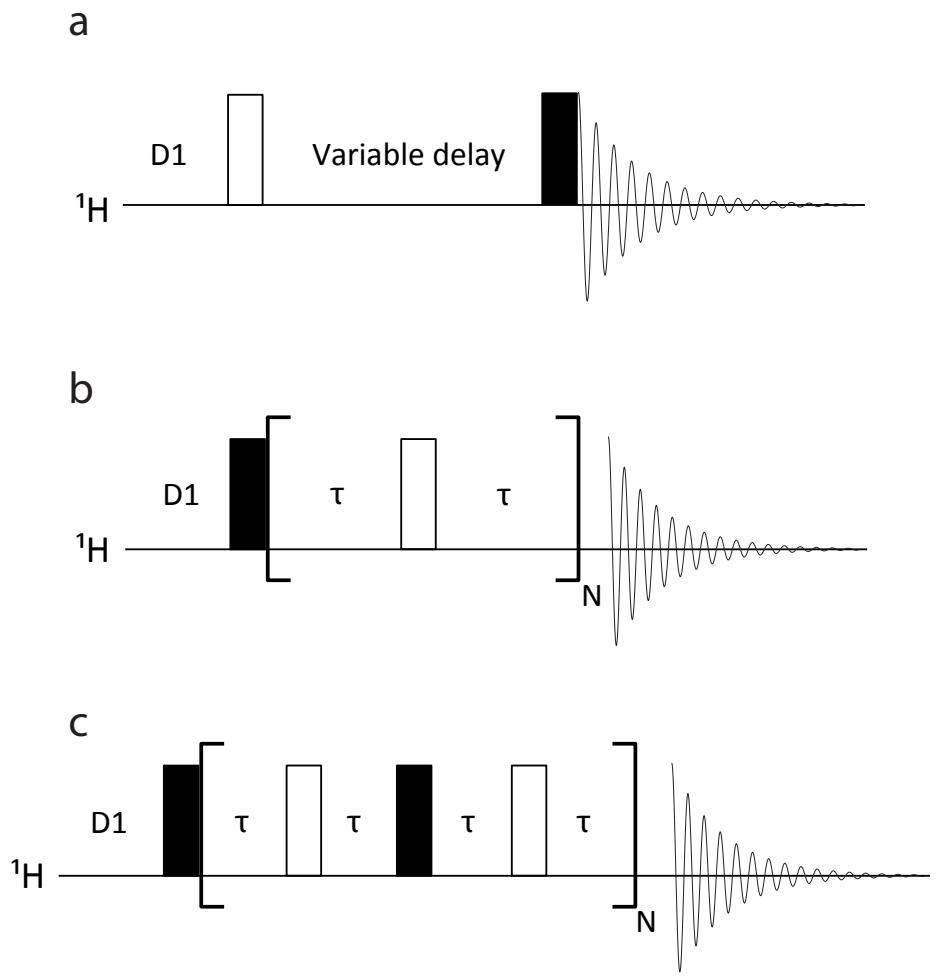


Figure 1.4. Pulse sequences commonly used for measurement of T_1 and T_2. Inversion-recovery (a) is used for T_1 measurement, while CPMG (b) is used for T_2 measurement. Recently, the PROJECT pulse sequence (c) has also been suggested for T_2 measurements. Full rectangles are 90° pulses, while empty rectangles are 180° pulses.

For T_2 measurements, the method of choice is the Carr-Purcell-Meiboom-Gill sequence[81] (CPMG, figure 1.4.b). After a 90° pulse, the magnetization undergoes several spin echoes. During these spin echoes, the chemical shift does not evolve, but relaxation occurs depending on T_2, more precisely on T_2*, which accounts for T_2 as well as the effects of magnetic field inhomogeneity in the sample. By varying the number of spin echoes, the resulting spectra have varying intensities depending on T_2*. More details about T_2* measurements from these spectra will also be given in part 6.

Recently, a new pulse sequence has been reported for T_2* measurements. This pulse sequence is called Periodic Refocusing of J Evolution by Coherence Transfer (PROJECT).[82] Like the "perfect echo" WATERGATE, the PROJECT uses "perfect" spin echo[78] to avoid the presence of anti-phase peaks in the spectrum, even with high delays between 180° pulses. A T_2 analysis has been recently published using this pulse sequence.[83]

Despite the existence of these pulse sequences, few studies in metabolomics or complex-mixture analysis mention a study of T_1 or T_2 in order to set the acquisition delay or the inter-scan delay,[84-86]. Among the possible explanations for this low interest in T_1 and T_2 are the highly congested 1D spectra of metabolomics mixtures, as detailed below, the fact that the inter-scan delay and the acquisition delay is already suggested in the standard procedure,[3], that T_1 and T_2 are generally considered constants in all samples during one study, and therefore have no impact for statistical analysis, or the lack of tools adapted for an automated measurement of the T_1 and T_2. More details about the last point will be given in part 6.

1.5. Limits of 1D ^{1}H NMR for metabolomics

While 1D ^{1}H spectra are the most sensitive spectra in NMR, the small spectral range of protons makes these spectra prone to peak overlap, especially in the case of complex mixtures, as shown in figure 1.5. While various degrees of complexity can arise from different biological media,[5, 46] 1D ^{1}H spectra are always characterized by a high degree of overlap between peaks. This overlap complicates the quantification of the metabolites, which is an important parameter for metabolomics.[6]

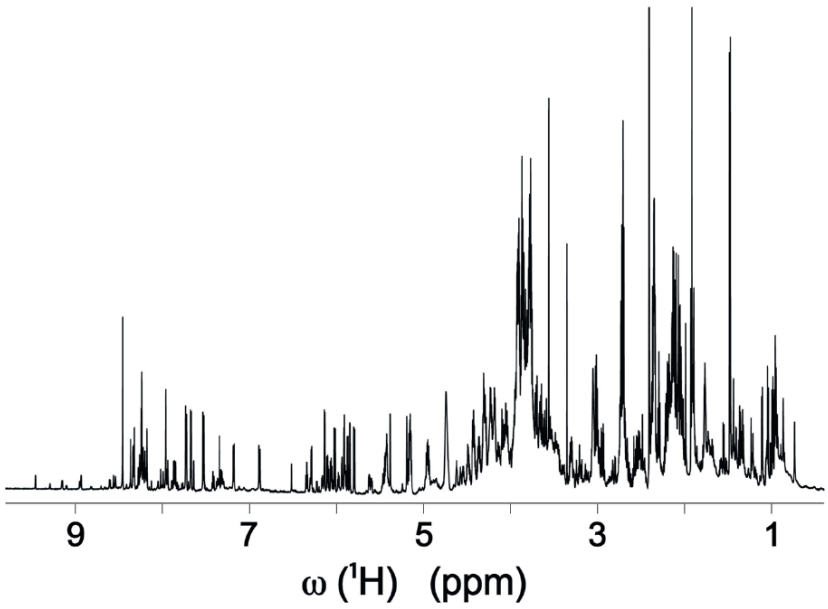


Figure 1.5. 1D ^{1}H spectrum of a cell lysate. Reprinted with permission from K.Bingol and R. Brüschweiller[5] Copyright 2014 American Chemical Society.

Even more importantly, this means that the regions delimited by spectral binning or bucketing can contain intensities from several peaks, and therefore several metabolites. This

complicates the statistical analysis and the identification of the biomarkers, since this brings uncertainty about the possible metabolites that are important for group separation. A more detailed explanation about the problems of peak overlap will be done in part 2.

Some methods to mitigate the problem of peak overlap have been developed in 1D spectra. The first is deconvolution of the 1D ^{1}H spectrum.[87] It consists on the separation of the 1D spectrum into several subspectra of the different metabolites.[9] This gives the possibility to quantify the metabolites and using directly the concentrations instead of peak intensities for statistical analysis. Since this method requires prior knowledge of the metabolites present in the mixture and focuses the study on these metabolites, deconvolution in metabolomics is often called metabolic profiling.

Despite this advantage, deconvolution has a few drawbacks, mainly the need for a library of the different metabolites to be quantified. Since the chemical shift of protons is sensitive to pH, salt concentration, magnetic field (for the chemical shift in Hz) and temperature, the library is also sensitive to these parameters. Also, since the peak intensity is dependent on acquisition and processing parameters, like acquisition time, inter-scan delay or the use of a window function (and therefore is similar to an external calibration),[52] a good quantification of the metabolites needs to replicate correctly the acquisition parameters of the individual metabolites. One last limitation is the increased difficulty to correctly deconvolute the spectrum when peak overlap becomes more preeminent, even with a good knowledge of the metabolites present in the mixture.[88] For all these reasons, deconvolution mainly works for the simplest biological samples, with a good prior knowledge of the metabolite content and a standardized protocol for sample preparation and spectrum acquisition, like plasma/serum[89] and, to a lesser extent, urine.[90]

Another possibility is the use of other nuclei for NMR than ^{1}H. The main target for these studies is ^{13}C, thanks to its near-universal presence in metabolites, similarly to ^{1}H. By using 1D ^{13}C pulse sequences with ^{1}H decoupling during acquisition, it is possible to obtain spectra with high spectral width (250 ppm instead of 10 ppm for ^{1}H) where every peak is a singlet.[91] Peak overlap is therefore less important in ^{13}C spectra compared to ^{1}H spectra, while the direct proportionality between peak area and concentration is the same as ^{1}H spectra. The increased resolution has proven to lead to better discrimination in multivariate analysis.[7, 8] Plus, ^{13}C chemical shifts are less sensitive to pH and salt concentration than ^{1}H,[7] making libraries of ^{13}C chemical shifts less sensitive on the medium studied.

However, this increased resolution comes with a reduced sensitivity, which arises from the reduced gyromagnetic ratio (around 1/4th of ^{1}H gyromagnetic ratio) and the low natural abundance of ^{13}C (around 1.1%). Because of this shortcoming, ^{13}C acquisition needs high concentrations or high acquisition times, which are not ideal for metabolomics. Another possibility is the use of microprobes optimized for ^{13}C direct dimension, which makes it possible to use less biological material in order to record a ^{13}C spectrum with acceptable experimental times for metabolomics.[92] Isotopic enrichment is also possible to enhance sensitivity, but the peaks in the resulting 1D ^{13}C spectrum are not necessarily singlets anymore, because of the involvement of J-couplings between ^{13}C. $^{1}J_{C\text{-}C}$, in particular, is around 40-60 Hz, which increases the risk of overlap in 1D ^{13}C spectra.[5] Other, more sensitive nuclei have been also used, like ^{19}F[93, 94] or ^{31}P[95], but they did not share the

same success as ^{1}H or ^{13}C, as they are present only in a limited number of metabolites. Nevertheless, these nuclei can be useful for targeting groups of metabolites.

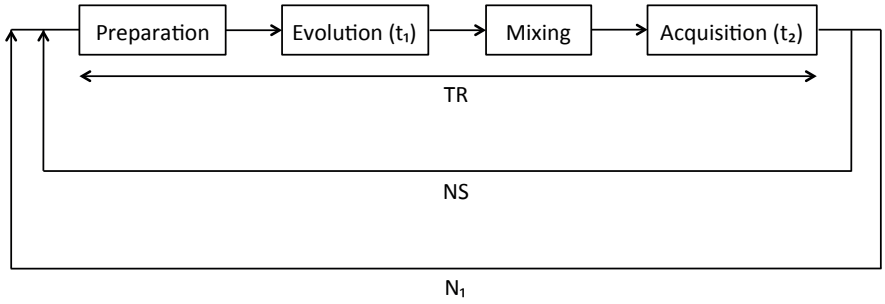
A last possibility, which will be discussed in the next section, is the use of multi-dimensional NMR, especially 2D NMR.

2. 2D NMR for complex-mixture analysis

2.1. Basics of 2D NMR

Two-dimensional (2D) NMR is a concept that has been developed in the seventies, which has revolutionized NMR, as explained below. It allows the acquisition of spectra with two frequency dimensions, with a wealth of different alternatives for these different frequencies. One of the biggest strength of 2D NMR is the possibility to directly see the connectivity between atoms, generally through bond or through space. This strength makes 2D NMR an important tool for the determination of molecular structure. Different 2D pulse sequences will be shown in later sections and will show at least a part of this wealth of information.

The general scheme for 2D NMR is shown in figure 1.6 and can be applied to describe almost any 2D pulse sequence with four subparts.



$$T_{exp} = TR \times N_1 \times NS$$

Figure 1.6. General scheme of a 2D pulse sequence.

The preparation aims to manipulate the magnetization in order to obtain the desired state for the evolution time. Generally, this subpart consists of the inter-scan delay and a 90° pulse to tip magnetization in the transverse plane, but more elaborate manipulations can occur for some pulse sequences.

The evolution is the first temporal dimension of the acquired FID. No acquisition is done during this time, but different scans with an incremented addition to this delay make it possible to obtain a FID, with each scan corresponding to a point in this FID, which are usually called t_1 increments.

The mixing consists in further manipulations of the magnetization in order to obtain the desired state for acquisition. During this mixing, a transfer of magnetization is achieved, whose nature (coherence, polarization...) will be different for each pulse sequence.

Finally, acquisition occurs, where a FID is recorded, similarly to 1D NMR. The time during which the probe records the signal is usually called t_2.

The experimental time T_{exp} of a 2D pulse sequence depends on 3 parameters: the time necessary to complete the pulse sequence once, including the inter-scan delay, called the repetition time (TR), the number of scans (NS) necessary to complete phase cycling or for sensitivity issues and the number of t_1 increments (N_1). Compared to 1D NMR, one new parameter is involved in the calculation of T_{exp}, which is N_1. Since several points are needed to correctly represent the indirect FID (usually several hundreds), the experimental time of 2D NMR is usually higher than 1D NMR for a similar sample, from several minutes to several hours for one spectrum. Each parameter has a minimum value: TR must be high enough to let enough magnetization return along the longitudinal axis between scans, for sensitivity issues as well as avoiding artifacts in certain cases, N_1 is dependent on the resolution needed in the indirect dimension and a minimal number of scans must be allowed for phase cycling, in order to suppress potential artifacts, or for sensitivity issues.

Once the bi-dimensional FID is acquired, Fourier transformation (FT) is used to obtain a 2D spectrum in the frequency domain. The strategy used in the FT algorithm in both dimensions is shown in figure 1.7, as used in most algorithms from commercial and public software.

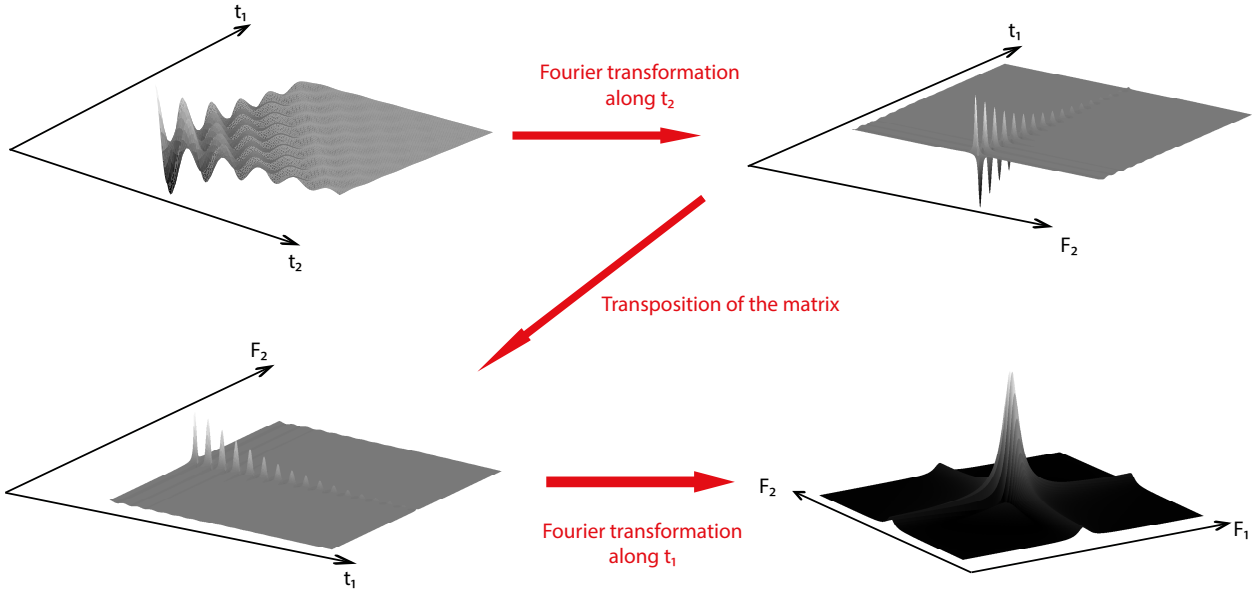


Figure 1.7. Strategy to use Fourier transformation in a two dimensional FID, as treated by the FT algorithm in commercial and public software.

From the 2D FID, which is stored as a matrix, FT is performed on one of the dimensions, in practice the direct dimension t_2. In order to let the algorithm use FT in the second dimension, which is the indirect dimension t_1, the data matrix is then transposed. The 2D spectrum in the frequency domain is then obtained. The F_2 dimension is often referred as the direct dimension, since the FID is directly recorded during acquisition, while F_1 is the indirect dimension, since the FID is recorded point by point with each t_1 increment.

While phase-sensitive detection in the direct dimension uses the same strategy as 1D spectra, phasing in the indirect dimension requires other methods, by shifts in pulse phases between t_1 increments (States,[96] Time-Proportional Phase Incrementation (TPPI)[97]) or by changes in the sign of gradients (echo-antiecho[98]). However, some 2D spectra are preferably displayed in magnitude mode, because of unfavorable phase properties. The content of a 2D spectrum will depend on the pulse sequence used to obtain the 2D FID. The next section will show some of the most used 2D pulse sequences, along with important improvements that have simplified the analysis of the spectra obtained from these pulse sequences.

2.2. Basic 2D pulse sequences and important improvements

For this section, only the pulse sequences used in later parts of the manuscript will be described. The spectrum shown for each pulse sequence has been acquired with a sample of 7 metabolites: lactate, alanine, asparagine, glycerol, serine, proline and citrate in H_2O/D_2O (90/10). This sample will also be used in part 4. For COSY, TOCSY, DQS, HMQC and HSQC, the product operator formalism will be used for a better understanding of each pulse sequence. Unless specified, all the pulses are along the x-axis.

2.2.1. Homonuclear J-resolved spectroscopy (J-RES)

J-RES is one of the oldest 2D pulse sequence developed[99] and is not a correlation experiment, but instead focuses on the separation of two parameters, the chemical shift and the J-coupling. The pulse sequence is shown in figure 1.8.a., which is one of the most simple 2D pulse sequences. After the 90° pulse, the evolution already starts, and a spin echo is used during the evolution time in order to let only the J-couplings evolve during t_1. No mixing is involved and the acquisition starts immediately after evolution.

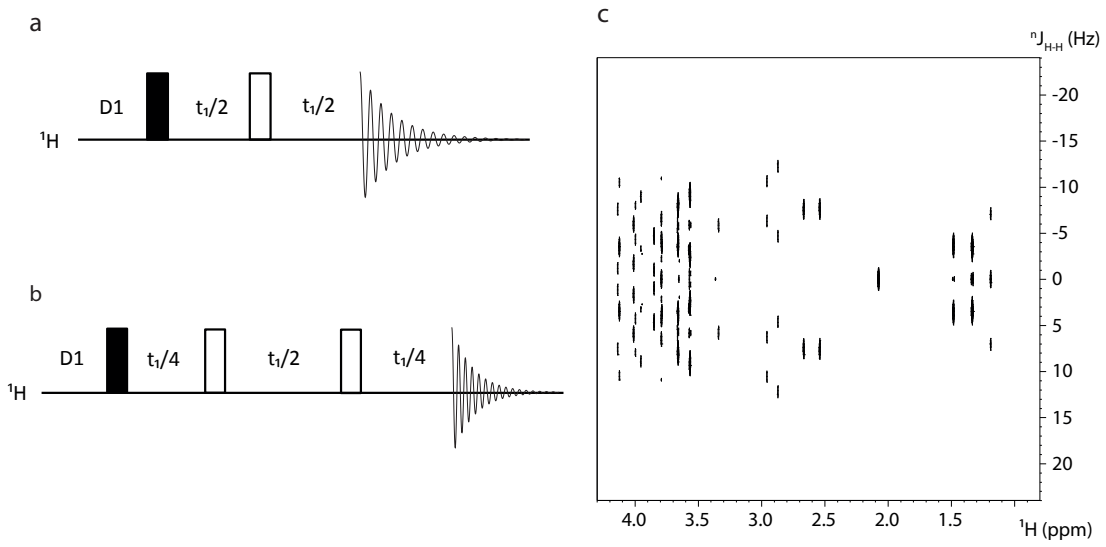


Figure 1.8. 2D pulse sequence for J-RES (a), along with J-RES DSE (b) and a J-RES DSE spectrum of a mixture of 7 metabolites in H_2O/D_2O (90/10). All the pulses are along the x-axis.

During acquisition, both the chemical shifts and the J-couplings evolve, which means that the spectrum obtained immediately after FT does not directly separate the two parameters. Each

peak has the coordinates (δ+J; J) along F_2 and F_1, respectively, with δ representing the chemical shift and J the J-couplings. In order to separate each parameter, an operation called tilting (also called shearing) can be used, which transforms each coordinate into a linear combination of these coordinates. For J-RES, each peak of coordinate (Ω_1; Ω_2) has the new coordinates (Ω_1-Ω_2; Ω_2), which finally gives the coordinates (δ; J) in the J-RES spectrum.[15] The peaks have the desired coordinates, but are distorted,[13] which can be resolved by symmetrization along F_1 at the axis J=0 Hz.

After this processing step, the J-RES spectrum is obtained, but this spectrum may contain artifacts from strong coupling.[14] Since these artifacts undergo the same coherence pathway as the expected peaks, neither phase cycling nor gradients can be used to suppress them. However, this suppression can be obtained by a change in the pulse sequence and processing, with the J-RES Double Spin Echo (DSE),[14] shown in figure 1.8.b. The main change between J-RES and J-RES DSE is a second 180° pulse that also acts as a spin echo. Since the genuine peaks and the artifacts have a different behavior during the second 180° pulse, it becomes possible to suppress the strong coupling artifacts by tilting and symmetrizing the spectrum.

The resulting spectrum can be seen in figure 1.8.c, and shows the complete separation of chemical shifts and J-couplings. This separation simplifies the analysis of the spectrum, especially in the region between 3.5 and 4 ppm. However, it also shows that the spectral width in the J-coupling dimension is small, not higher than 40 Hz, compared to several thousands of Hz in the chemical shift dimension. Therefore, this pulse sequence allows only a moderate increase in resolution compared to 1D ^{1}H spectra.

A heteronuclear version of the J-RES pulse sequence has been developed shortly after the homonuclear variant,[100] with the ^{1}H chemical shifts in the direct dimension and $J_{C\text{-}H}$ in the indirect dimension. Contrary to homonuclear J-RES, the heteronuclear variant does not need shearing or symmetrization to completely separate the two parameters.

2.2.2. COrrelation SpectroscopY (COSY)

With J-RES, COSY is one of the first 2D pulse sequence developed.[11] Its purpose is to show the direct neighbors of each proton through bond connectivity, in other words through J-couplings, as seen in figure 1.9.a. The pulse sequence, in its simplest form, is shown in figure 1.9.b. Only the magnetization from one spin I_1 will be considered in the next paragraph, but the analysis can be expended to the other proton spins of the molecule, like with the other pulse sequences presented in the next sections.

After the 90° pulse along x, which tilts the magnetization from I_{1z} (longitudinal axis) to $-I_{1y}$ (transverse plane), both the chemical shift and the J-couplings are evolving during t_1. With these evolutions, anti-phase magnetization $2I_{1x}I_{2z}$ is building up during t_1, with I_2 being another spin that is sharing a J-coupling with I_1. During the second 90° pulse, the anti-phase magnetization undergoes coherence transfer, from $2I_{1x}I_{2z}$ to $-2I_{1z}I_{2x}$: the anti-phase magnetization that was with I_1

before the 90° is transferred to I_2. Therefore, this magnetization does not evolve at the same Larmor frequency during acquisition compared to the evolution time. The peak, with coordinates (δ_2; δ_1), is a direct proof of an existing J-coupling between I_2 and I_1 and is called cross peaks (or correlation peaks). Other visible peaks in the 2D spectrum of COSY come from magnetization which was not transferred during the second 90°pulse. These peaks, of coordinates (δ_1; δ_1), are called diagonal peaks, and give no structural information about the molecule.

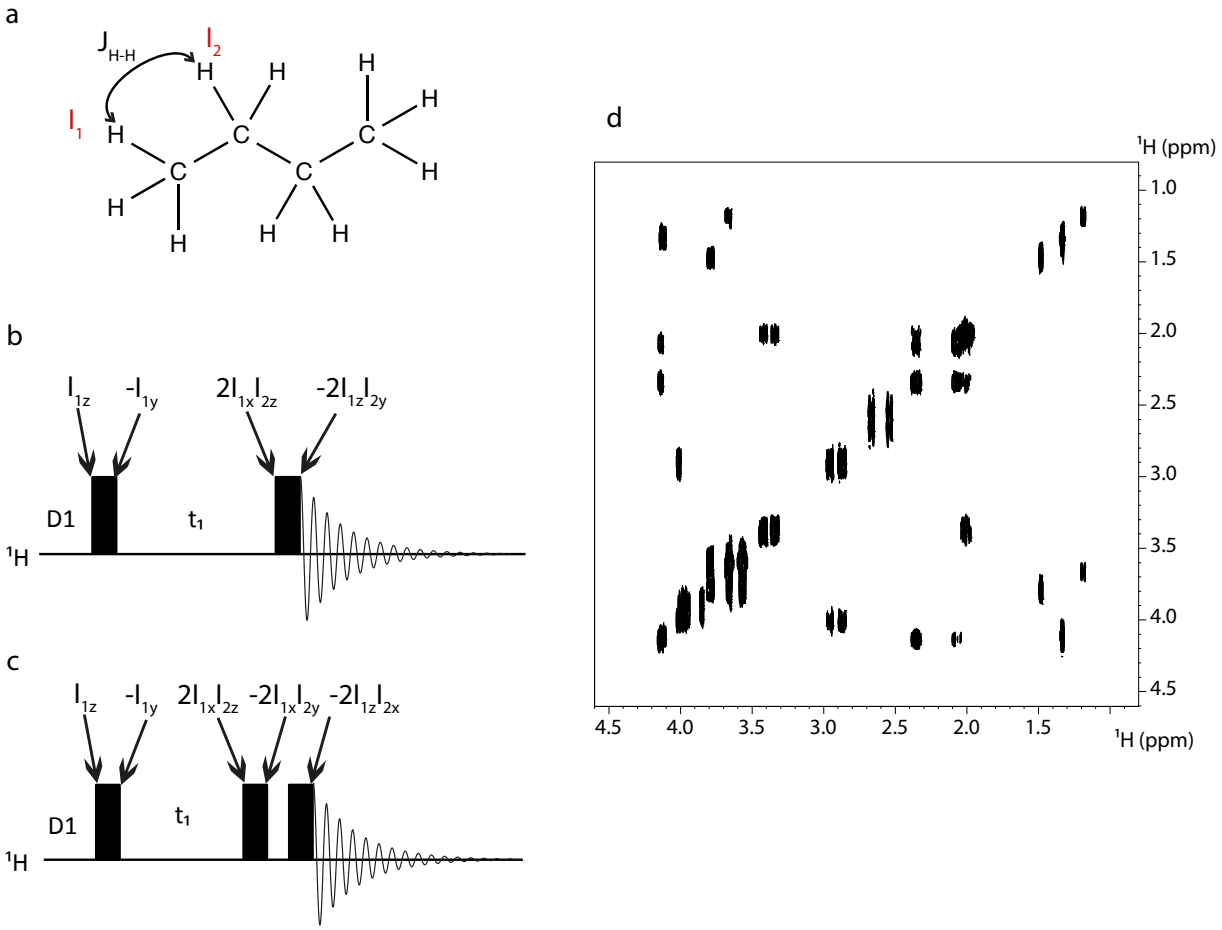


Figure 1.9. Correlations involved during the COSY pulse sequence (a), with the 2D pulse sequence for COSY (b), along with DQF-COSY (c) and a DQF-COSY spectrum (d) of a mixture of 7 metabolites in H_2O/D_2O (90/10). All the pulses are along the x-axis.

Unfortunately, the basic COSY experiment is hampered by the phase difference between diagonal and correlation peaks. If the lineshape of correlation peaks are absorptive, the lineshape of diagonal peaks are dispersive, and vice-versa.[11] This complicates the analysis of a COSY spectrum, since dispersive lineshapes for diagonal peaks means that it becomes difficult to visualize correlation peaks close to the diagonal, for example. For this reason, a variant of the COSY pulse sequence has been developed, called Double Quantum Filtered-COSY (DQF-COSY),[101] which is represented in figure 1.9.c. In this variant, a third 90° pulse is used just after the second one. Contrary to the COSY pulse sequence, the magnetization of interest is the one converted to double-quantum coherence $-2I_{1x}I_{2y}$ during the second pulse (another possible evolution of the anti-phase magnetization $2I_{1x}I_{2z}$ built up during t_1). This magnetization is reconverted to anti-phase magnetization $-2I_{1z}I_{2x}$ during the third 90° pulse, which is then acquired during t_2. The

resulting spectrum, shown in figure 1.9.d, also has diagonal and correlation peaks. But this time, the phase of these two peaks is the same, facilitating the analysis of the spectrum. Plus, the intensity of the diagonal peaks is reduced, especially those of singlets. However, the generation of the double-quantum coherence is done at the price of a reduction by a factor of two of SNR.[101]

This variant, along with COSY-45, are the most used variants of the COSY pulse sequence. However, when using DQF-COSY, it is important to reduce as much as possible the transverse magnetization between scans, in order to avoid certain artifacts to become visible in the spectrum, which are difficult to completely suppress by phase cycling.[102]

2.2.3. TOtal Correlation SpectroscopY (TOCSY)

The goal of the TOCSY pulse sequence is to show an extended network of J-couplings in a given molecule, with the example of three spins I_1 I_2 and I_3, with I_1 and I_2 sharing a J-coupling, along with I_2 and I_3, as shown in figure 1.10.a. The pulse sequence to generate this kind of spectrum is shown in figure 1.10.b,[103] which is already an improvement upon the original TOCSY pulse sequence, where only the first of the three 90° pulses is applied, which originally caused artifacts in the resulting spectrum.[16]

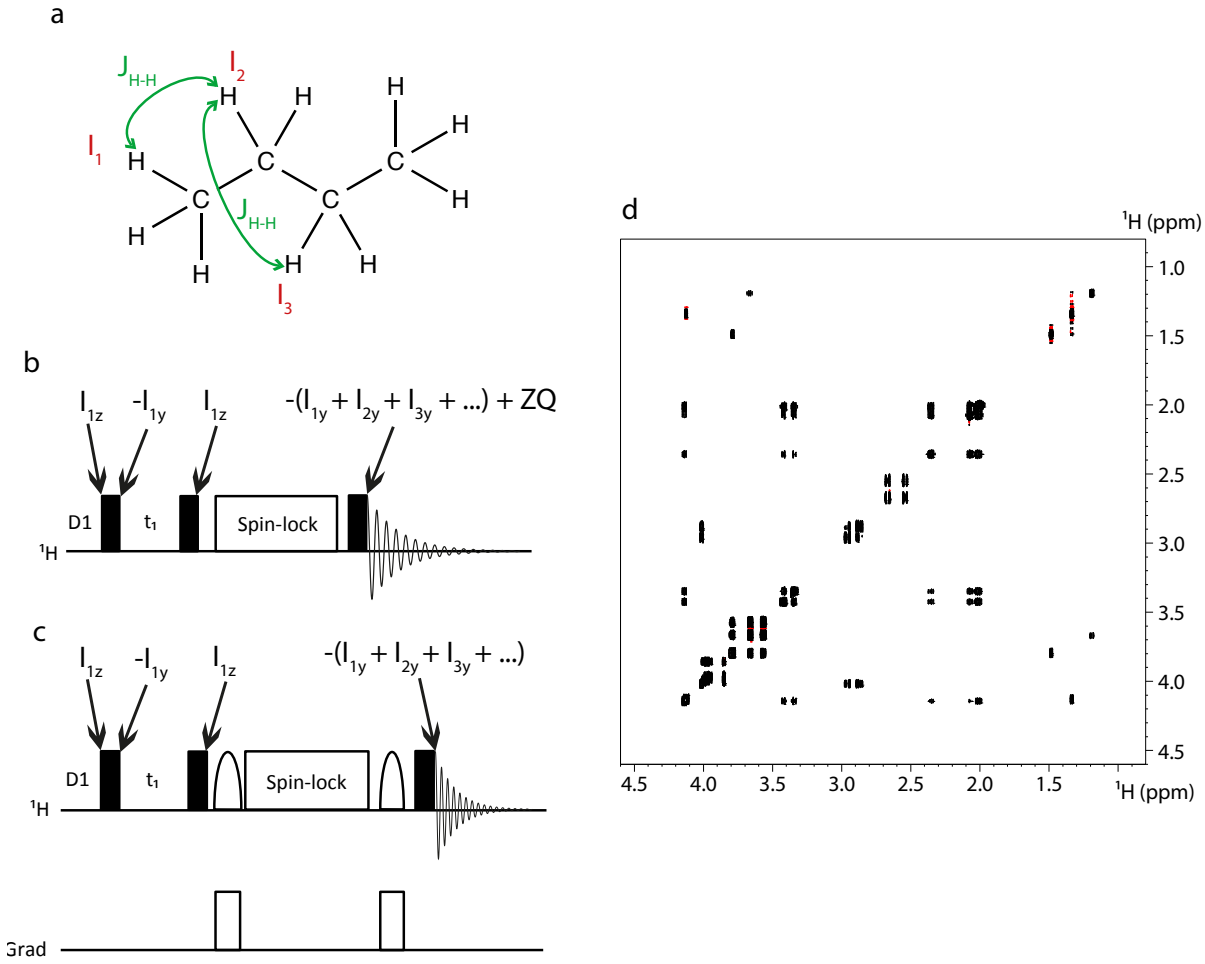


Figure 1.10. Correlations involved during the TOCSY pulse sequence (a), with the 2D pulse sequence for TOCSY (b), along with zf-TOCSY (c) and a zf-TOCSY spectrum (d) of a mixture of 7 metabolites in H_2O/D_2O (90/10). All the pulses are along the x-axis.

During this pulse sequence, after the 90° pulse, which tilted I_1 from I_{1z} to I_{1x}, and the evolution time, the second 90° pulse returns I_1 to I_{1z} and a spin-lock is used. A spin-lock can be seen as a delay during which many 180° pulses are performed. Therefore, the chemical shift does not evolve during the spin-lock, but the J-couplings do: the strong coupling condition is forced for all protons having at least one J-coupling. For magnetization in strong coupling, a 180° pulse can cause coherence transfer,[14] therefore the magnetization can undergo several coherence transfers during the spin-lock. At the beginning, the magnetization is only transferred to spins with a J-coupling, but if the spin-lock is long enough (generally between 60 and 80 ms), the coherence transfer is extended to protons with no direct J-coupling, but that are in the same network of coupling as the initial proton, like I_1 and I_3. The last 90° pulse tips all the spins in the transverse plane before acquisition. The resulting spectrum is similar to a COSY spectrum, with diagonal peaks and correlation peaks, but more correlations peaks are present to show the extended network of protons. It is possible, from one slice of the spectrum, to visualize this network.

But with the spin-lock, zero-quantum coherence (ZQ) is also generated, which can generate anti-phase dispersive magnetization after the last 90° pulse, with the possibility to obscure the correlation peaks. Since phase cycling or gradients do not affect this coherence, it is difficult to suppress. This suppression can be achieved by the addition of a z-filter (zf), as shown in the pulse sequence for zf-TOCSY in figure 1.10.c.[104] This z-filter consists in a shaped pulse (in this case, an adiabatic inversion pulse) used at the same time as a gradient. This element creates a phase for the zero-quantum coherence that is dependent on its position along the z axis. When detecting the whole sample, the different contributions cancel themselves and therefore no net signal from anti-phase peaks is seen during acquisition.

As seen in the resulting spectrum in figure 1.10.d, no dispersive peaks are present in the spectrum. The spectrum is also similar to the COSY spectrum in figure 1.9.d, but with additional correlation peaks corresponding to the extended network of proline peaks. One additional note is that the peaks are in-phase in TOCSY,[103] compared to anti-phase peaks in DQF-COSY, which will become important in later sections.

2.2.4. Double-quantum spectroscopy (DQS)

The interest in multiple-quantum spectroscopy, in particular DQS, lies in the simplification of the resulting spectrum. However, because multiple-quantum coherences are not directly detectable by NMR, their measurement was not possible with FT NMR until the development of 2D NMR.[105] Like COSY, the connectivity that is used by the pulse sequence is J-couplings between protons, as seen in figure 1.11.a, so the same notation will be used. However, the coherence pathway is different and leads to a different spectrum. The pulse sequence is shown in figure 1.11.b, and can work for every multiple-quantum spectroscopy by changing the phase cycling or by the use of gradients,[106] but in this section we will assume that double-quantum coherence is selected.

During the preparation, after the 90° pulse, a spin echo is used to generate anti-phase magnetization $-2I_{1x}I_{2z}$, which is converted to double-quantum coherence $2I_{1x}I_{2y}$ by the next 90° pulse. Since the creation of anti-phase magnetization requires at least two protons of different

Larmor frequencies δ_1 and δ_2 that share a J-coupling, singlets in 1D ^{1}H spectra cannot create double-quantum coherence and therefore are not present in DQS spectra. Ideally, the value of 1/(4J) during the delay of the spin echo should allow the complete conversion of in-phase magnetization into anti-phase magnetization. But if several values of J_{H-H} exist in a sample, a medium value should be used to limit losses. During evolution, the Larmor frequency for double-quantum spectroscopy is the sum of the two protons Larmor frequencies $\delta_1+\delta_2$.

After evolution, a last 90° pulse converts double-quantum coherence $2I_{1x}I_{2y}$ back into anti-phase magnetization $2I_{1z}I_{2x}$, which is recorded during acquisition. An example of a DQS spectrum is shown in figure 1.11.c. In this spectrum, there are no diagonal peaks, only correlation peaks, which reduce the risks of overlap. Correlation peaks with the same chemical shift in the double-quantum dimension share a J-coupling, and their position in the single-quantum dimension are the same as in the 1D ^{1}H spectrum, as seen for some metabolites in the spectrum.

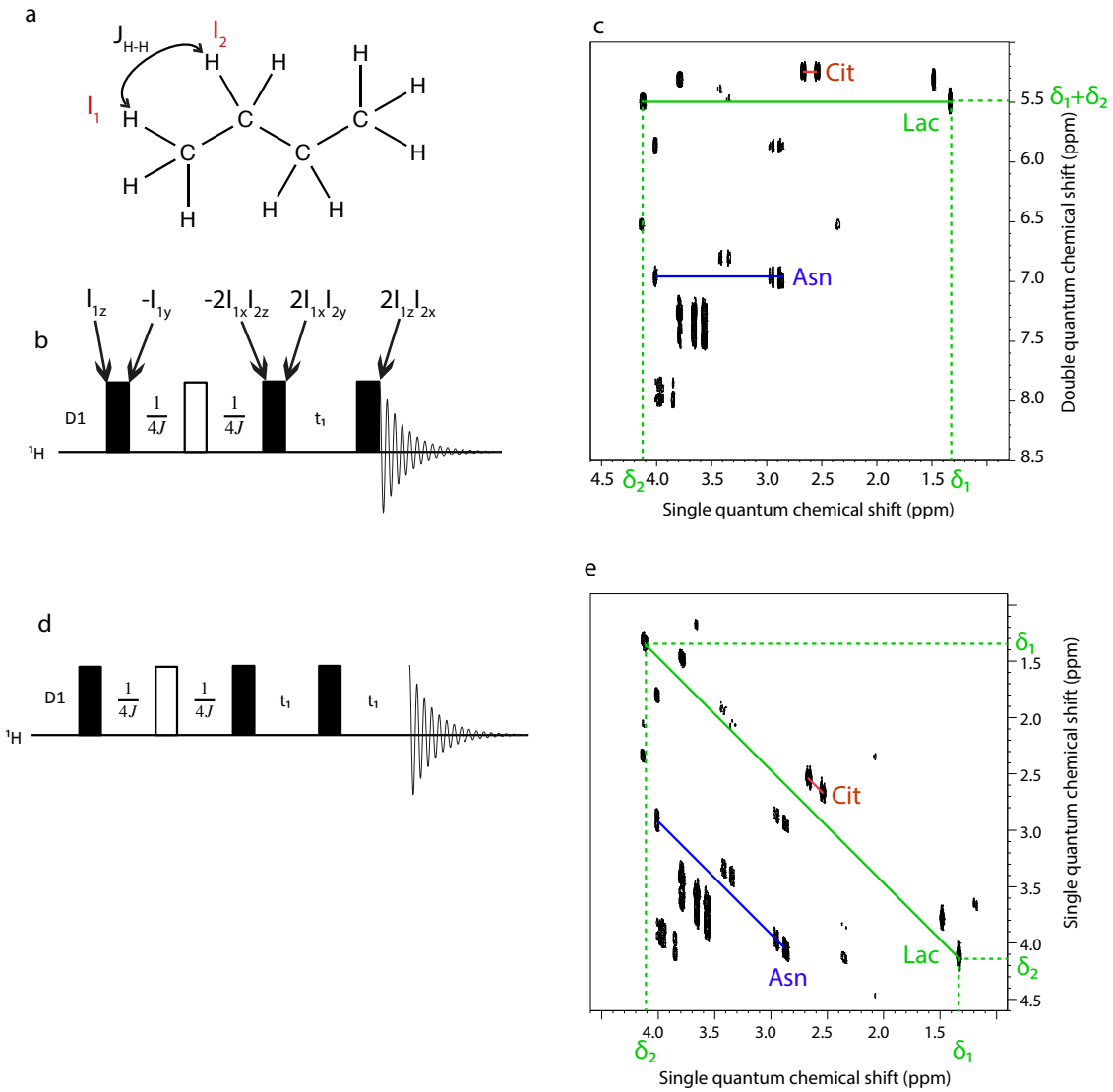


Figure 1.11. Correlations involved during the DQS pulse sequence (a), with the 2D pulse sequence for DQS (b), and a DQS spectrum (c) of a mixture of 7 metabolites in H_2O/D_2O (90/10), along with the assignment of a few peaks. The pulse sequence for DQSSY (d) and a spectrum (e) of the same sample is also shown. Abbreviations: Cit: citrate, Lac: lactate, Asn, asparagine. All the pulses are along the x-axis.

From this description, it would seem that there are two correlation peaks at most for each chemical shift in the indirect dimension, but in reality, this number may be higher, as seen for the asparagine correlation peaks. These peaks are called remote or relay peaks and their origin is explained elsewhere.[40] Their presence can give new information about the structure of the molecule,[107] but can also make the analysis of the DQS spectrum less straightforward.

While informative, the information content in DQS spectra is similar to COSY spectra, without the presence of diagonal peaks. It may be for this reason that DQS spectra for protons are not as widely used as COSY for structure elucidation. Another possible reason is the dependence of peak intensity on the delay for anti-phase build-up. Since $J_{H\text{-}H}$ can vary between around 15 and 1 Hz, the range of possible delays is important and a poor choice of this delay can result in important sensitivity penalties.

However, a variant of the DQS pulse sequence is often used in structure elucidation, which is basically the same pulse sequence at the frequency of ^{13}C: the Incredible Natural Abundance DoublE QUAntum Transfer Experiment (INADEQUATE) pulse sequence.[108] Using the DQS for ^{13}C instead of ^{1}H has several advantages:

-Knowing the ^{13}C connectivity is very important to confirm the molecular structure, and the COSY pulse sequence is not practical for natural abundance ^{13}C since the diagonal peaks constitutes at least 99% of the total intensity of the spectrum.

-$^{1}J_{C\text{-}C}$ is around 30-50 Hz for most molecules, which gives a narrower list of possible delays for anti-phase build-up.[109]

-Because of the very rare probability that a molecule has three ^{13}C at natural abundance, relay peaks are usually not seen in the INADEQUATE spectra. While they carry structural information, their absence makes the analysis of the INADEQUATE spectra straightforward.

Despite these advantages, the INADEQUATE pulse sequence is rarely used for diluted samples because of its very poor sensitivity, since only around 1 molecule in 10^4 has two adjacent ^{13}C.

A variant of the DQS pulse sequence exists to obtain COSY-like spectra, called the DQS SYmmetrized (DQSSY),[110] whose pulse sequence is shown in figure 1.11.d. The main difference between DQS and DQSSY is the presence of an incremented delay after the last 90° pulse, which lasts as long as the evolution time. During the first incremented delay (before the last 90° pulse), the magnetization from a correlation peak is into DQ coherence, then a 90° pulse is applied, shifting the coherence of a correlation peak from DQ to SQ. The second incremented delay transforms the term for evolution during t_1 from $e^{i(\Omega_1+\Omega_2)t_1}$ to $e^{i\Omega_1 t_1}$,[111] where Ω_1 is the Larmor frequency of I_1 and Ω_2 is the Larmor frequency of I_2. Since the final evolution term during t_1 is along I_1 and the magnetization is along I_2 before acquisition, the position of a correlation peak on the DQSSY spectrum is (Ω_1; Ω_2) in the (F_1; F_2) plane, as seen in the DQSSY spectrum, in figure 1.11.e. The resulting spectrum is the equivalent of a COSY spectrum without diagonal peaks, but with remote peaks. It can also be seen as a DQS spectrum that has been tilted by 45° along F_1, as

shown by the position of the peaks that shared a similar chemical shift in the double-quantum dimension of the DQS spectrum.

2.2.5. Heteronuclear Single-Quantum Coherence (HSQC) and Heteronuclear Multiple-Quantum Coherence (HMQC)

With HMQC and HSQC pulse sequences, it is possible to correlate protons with a different nucleus, usually ^{13}C. For these pulse sequences, only $^{1}J_{C-H}$ are used, as shown in figure 1.12.a, involving a proton spin called I and a ^{13}C spin called S sharing a bond with I. The most simple pulse sequence, in terms of the number of pulses, is the HMQC, shown in figure 1.12.b.

During this pulse sequence, after the 90° pulse, anti-phase magnetization $2I_xS_z$ is built up during a delay, then a 90° pulse in the ^{13}C channel transfers magnetization from the proton to a ^{13}C sharing a J-coupling. During t_1, magnetization evolves as heteronuclear multiple-quantum coherence $-2I_xS_y$ and both ^{1}H and ^{13}C chemical shifts can evolve. For this reason a spin echo is placed in the ^{1}H channel to avoid the evolution of ^{1}H chemical shifts and J_{C-H} couplings, so that only ^{13}C chemical shifts evolve during t_1. However, ^{1}H couplings are also evolving during t_1, which will have an importance in the resulting spectrum. After t_1, a second 90° pulse in the ^{13}C channel is used to retransfer magnetization back to the protons, as anti-phase magnetization $-2I_xS_z$, and a second delay is applied to completely refocus the ^{1}H chemical shifts before acquisition and obtain in-phase magnetization $-I_y$, which is recorded during acquisition. For maximum sensitivity, the two non-incremented delays should be fixed to $1/(2J_{C-H})$. The result, as seen in figure 1.12.b, is a spectrum with ^{1}H chemical shifts in the direct dimension and ^{13}C chemical shifts in the indirect dimension. Thanks to the high spectral width of ^{13}C, HMQC spectra have more resolution than the previous pulse sequences.

However, a similar spectrum can be obtained with the HSQC pulse sequence, shown in figure 1.12.c.[23] The pulse sequence is similar to HMQC, with the two non-incremented delays replaced by INEPT transfer blocks. During the first INEPT block, anti-phase magnetization $-I_xS_z$ is built up, then with the two last 90° pulses of the INEPT block, magnetization is transferred to S as anti-phase magnetization $-I_zS_y$. This magnetization evolves during t_1, and then the second INEPT block starts. The magnetization is first transferred back to I as anti-phase magnetization $-I_yS_z$, and then evolves as in-phase magnetization I_x by J-couplings.

For improved sensitivity with phased spectra, a variant of HSQC has been developed, called sensitivity-enhanced HSQC (figure 1.12.d).[112] Compared to the HSQC pulse sequence, a new INEPT block has been placed before acquisition. This allows acquiring the signal from two different pathways, allowing an improvement up to $\sqrt{2}$ of sensitivity. However, transverse relaxation also occurs during the INEPT block, so this pulse sequence is not recommended for molecules with low T_2*, like macromolecules.

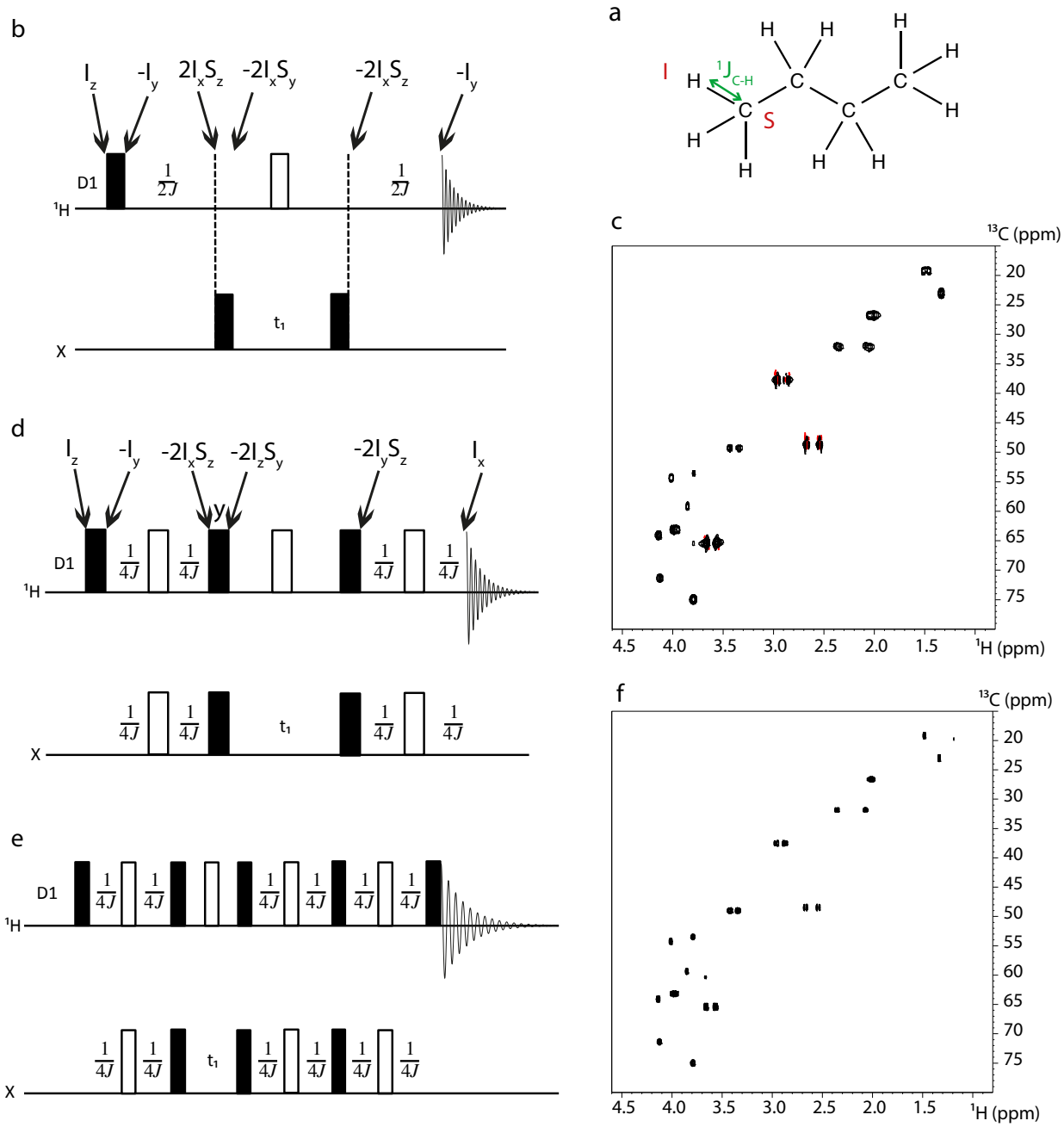


Figure 1.12. Correlations involved during the HSQC pulse sequence (a), with a 2D pulse sequence for HMQC (b), along with a HMQC spectrum (c), of a mixture of 7 metabolites in H_2O/D_2O (90/10). The HSQC (d) and sensitivity-enhanced HSQC (e) pulse sequence and are also presented, along with a HSQC spectrum (f) of a mixture of 7 metabolites in H_2O/D_2O (90/10). Unless specified, all the pulses are along the x-axis.

The HSQC spectrum of the same sample, shown in figure 1.12.e, is indeed very similar to the HMQC spectrum, but with a slightly better resolution, especially in the indirect dimension. This is caused by the evolution of J_{H-H} couplings during t_1 in the HMQC pulse sequence, which causes the multiplet pattern in the 1D ^{1}H spectrum to appear in the indirect dimension, and therefore making the peaks slightly larger in the HMQC spectrum. As long as the pulses in the two channels are well calibrated, the HSQC pulse sequence is also slightly more sensitive than HMQC, due to a slower relaxation of single-quantum coherence compared to multiple-quantum coherence.

2.3. Previous studies in 2D NMR for quantification in complex mixtures and metabolomics

2.3.1. Quantification

Compared to 1D NMR, 2D NMR is not considered a primary method for quantification, since the volume of a peak is sensitive to many peak-dependent factors, like the relaxation times T_1, T_2* or couplings, depending on the pulse sequence.[52] However, peak volumes are still proportional to the concentration –with the coefficient of proportionality being site-specific– and it is possible to extract the concentration from peak volumes, by using other strategies. Most of the existing studies have been done with the HSQC spectrum, but this can be in principle extended to any 2D pulse sequence.

As seen in figure 1.13, different strategies can be used for quantification from 2D spectra. The first is to estimate each parameter that influences the peak volumes. T_1 and T_2 can be measured from the pulse sequences mentioned in section 1.4, while J-couplings can be inferred from 1D pulse sequences or J-RES pulse sequences, both homonuclear and heteronuclear. This strategy has already been used in a study with the HSQC pulse sequence.[113]

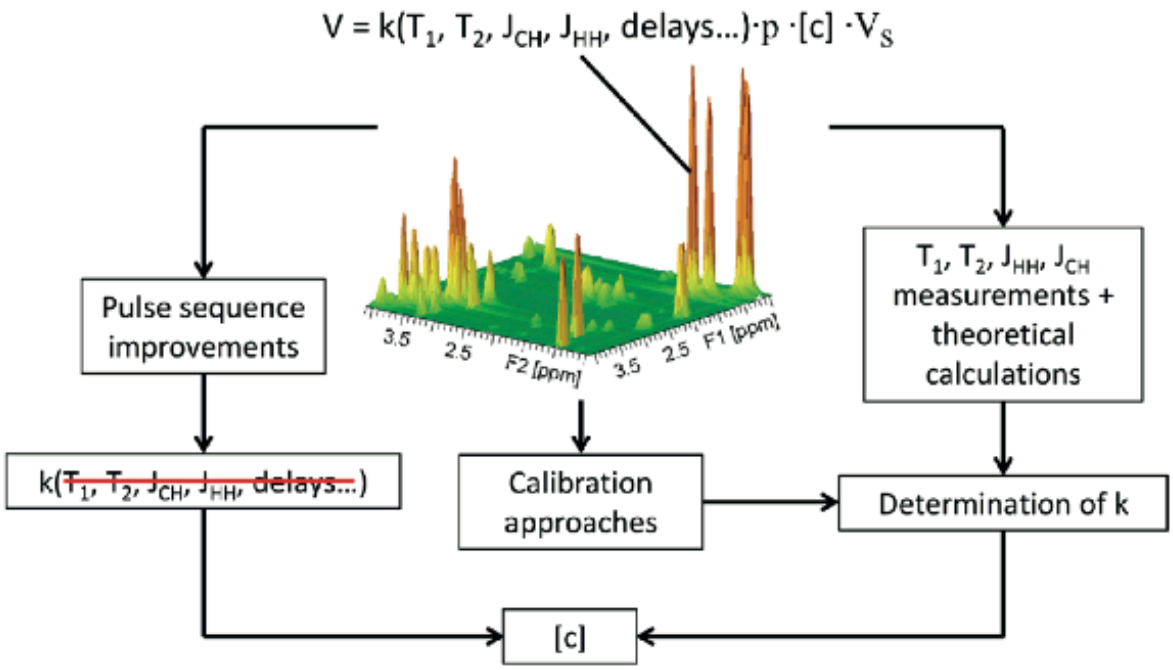


Figure 1.13. Different strategies to extract concentrations from 2D spectra. Reprinted with permission from P.Giraudeau[114] (c) Copyright (2014) John Wiley & Sons, Ltd.

The second is the development of new 2D pulse sequences that reduce or suppress the number of parameters changing the peak volumes. Two particular HSQC variants are been developed for this case: $HSQC_0$[115] and the Quantitative, Offset-Compensated, CPMG-Adjusted HSQC (Q-OCCA HSQC).[116]

The Q-OCCA HSQC is the result of several improvements on the HSQC pulse sequence [117-119] to make the peak volumes independent on the ^{13}C offset in the indirect dimension, the $^{1}J_{C-H}$ and the J_{H-H}. However, the peak volumes are still dependent on T_2* and T_1, which needs both complete relaxation and similar T_2* between peaks to avoid quantification errors. The $HSQC_0$ pulse sequence,[120] shown in figure 1.14, repeats up to three times the HSQC block, with three different sets of data recorded, one with only the $HSQC_1$ block applied, the second with $HSQC_1$ and $HSQC_2$ applied and the last one with the three blocks. The three resulting spectra have the same peaks, but with different intensities, and by linear extrapolation, it is possible to create a fourth spectrum, called the $HSQC_0$ spectrum, where peak intensities are those which would have been obtained with all evolution delays equal to zero. Therefore the resulting peak volumes are directly proportional to the concentrations, similarly to 1D NMR. This pulse sequence has already been tested on biological extracts and gives accuracy and precision generally lower than 10%.[115, 121, 122]

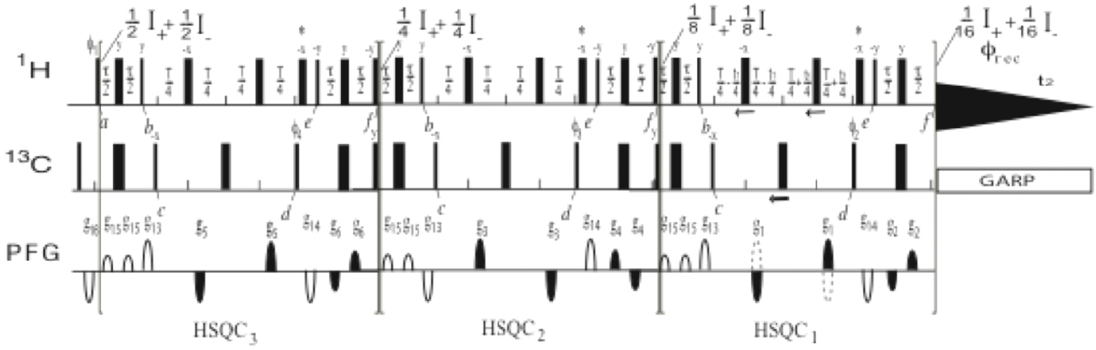


Figure 1.14. Pulse sequence for the $HSQC_0$,[115] reprinted with permission from K.Hu et al.[115] (c) Copyright (2012) American Chemical Society.

Finally, the last strategy is the use of calibration curves, similarly to commonly used strategies in other analytical disciplines such as optical spectroscopy or chromatography. This strategy requires the acquisition of several spectra to construct the calibration curves, but does not necessarily mean long experimental times, since multiple calibration curves can be recorded simultaneously if the standard samples are mixtures of all the targeted analytes. Different approaches can be used for building the calibration curves, but two are mainly used for quantification in complex mixtures. The first is external calibration, where 2D spectra of standard solutions containing a mixture of the targeted analytes are recorded with varying concentrations, a strategy which has been used in several studies with HSQC.[26, 123, 124] A non-linear dependence between concentration and peak volumes may occur in specific cases, but is not detrimental to the calibration[125] and one calibration curve can be used to quantify in all samples. However, this strategy could lead to errors if the molecule behavior is different between the biological sample and the samples for calibration, such as changes in T_1 or T_2 if partial saturation is used.

The second approach is standard addition, where the biological sample is spiked with a control solution containing the analytes in known concentrations. This strategy is less sensitive to changes in T_1 or T_2* since all the experiments are acquired with the same sample. However, this

approach is sensitive to possible non-linear dependency between concentration and peak volume,[125] and it is not appropriate for a large number of samples, since a standard addition curve must be drawn for each sample. This strategy has been used for HSQC and TOCSY spectra of plasma enriched in ^{13}C or ^{15}N,[126] and with a study of cell culture extract with the ^{1}H "INADEQUATE" pulse sequence, adapted for water suppression and fast analysis, thanks to the use of gradients that reduce the number of scans necessary to obtain an artifact-free spectrum.[127, 128] This pulse sequence is in fact a DQSSY pulse sequence with water presaturation and the use of gradients for coherence selection, in order to reduce the required NS for phase cycling. The pulse sequence is shown in figure 1.15.a, with a spectrum of a cell culture extract in figure 1.15.b.

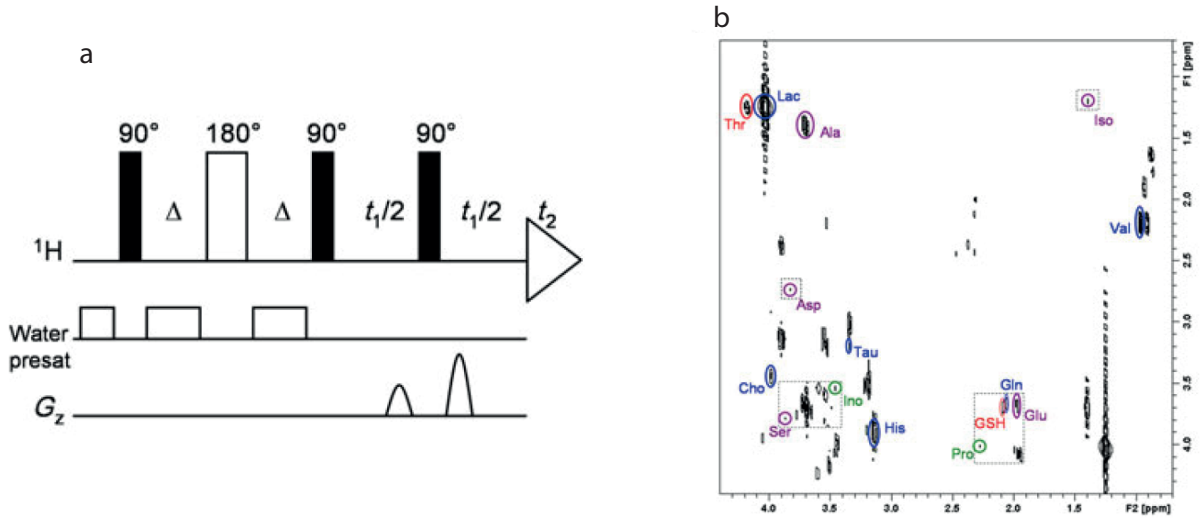


Figure 1.15. Pulse sequence for ^{1}H "INADEQUATE" adapted for diluted samples in water and faster acquisition (a) and a spectrum of a cell culture extract recorded in 15 min with this pulse sequence (b),[128] reprinted with permission from E. Martineau et al.[128] (c) Copyright (2012) John Wiley & Sons, Ltd.

2.3.2. Metabolomics

Several 2D pulse sequences have already been tested for metabolomics, the main ones being COSY, J-RES, HSQC and TOCSY. The main results for each pulse sequence will be detailed below.

For COSY, the main studies with biological samples do not involve statistical analysis, but a graphical analysis called Differential Analysis by 2D NMR Spectroscopy (DANS).[17] This analysis involves a subtraction between two spectra of non-fractioned extracts with Photoshop. The most intense peaks from the resulting spectrum are likely to come from a potential biomarker and can be used to start the structure elucidation. Therefore, this method is more suited for the discovery of natural products and biomarkers of unknown structure.[129-134] The DQF-COSY has been chosen for this analysis thanks to its sensitivity and its usefulness for the elucidation of most chemical groups, allowing the selection of a purification scheme for isolation and further analysis.[135]

The homonuclear J-RES pulse sequence has been used for metabolomics since 2003.[12] Compared to most 2D pulse sequences, the homonuclear J-RES has several advantages: it has one of the best sensitivity per unit of time of 2D pulse sequences,[13] it can be recorded with experimental times of tens of minutes, similarly to 1D ^{1}H spectra, with optimization of the parameters for acquisition and processing,[13] and all peaks visible in 1D ^{1}H spectra, both singlets and multiplets, are also visible in J-RES spectra, contrary to most homonuclear 2D pulse sequences like COSY or TOCSY, where singlets in 1D ^{1}H spectra have no correlation peak. A typical J-RES spectrum of a biological extract is shown in figure 1.16.a, which shows that, despite the complexity of the spectrum, the separation of chemical shifts and J-couplings is a benefit for identification of multiplets.

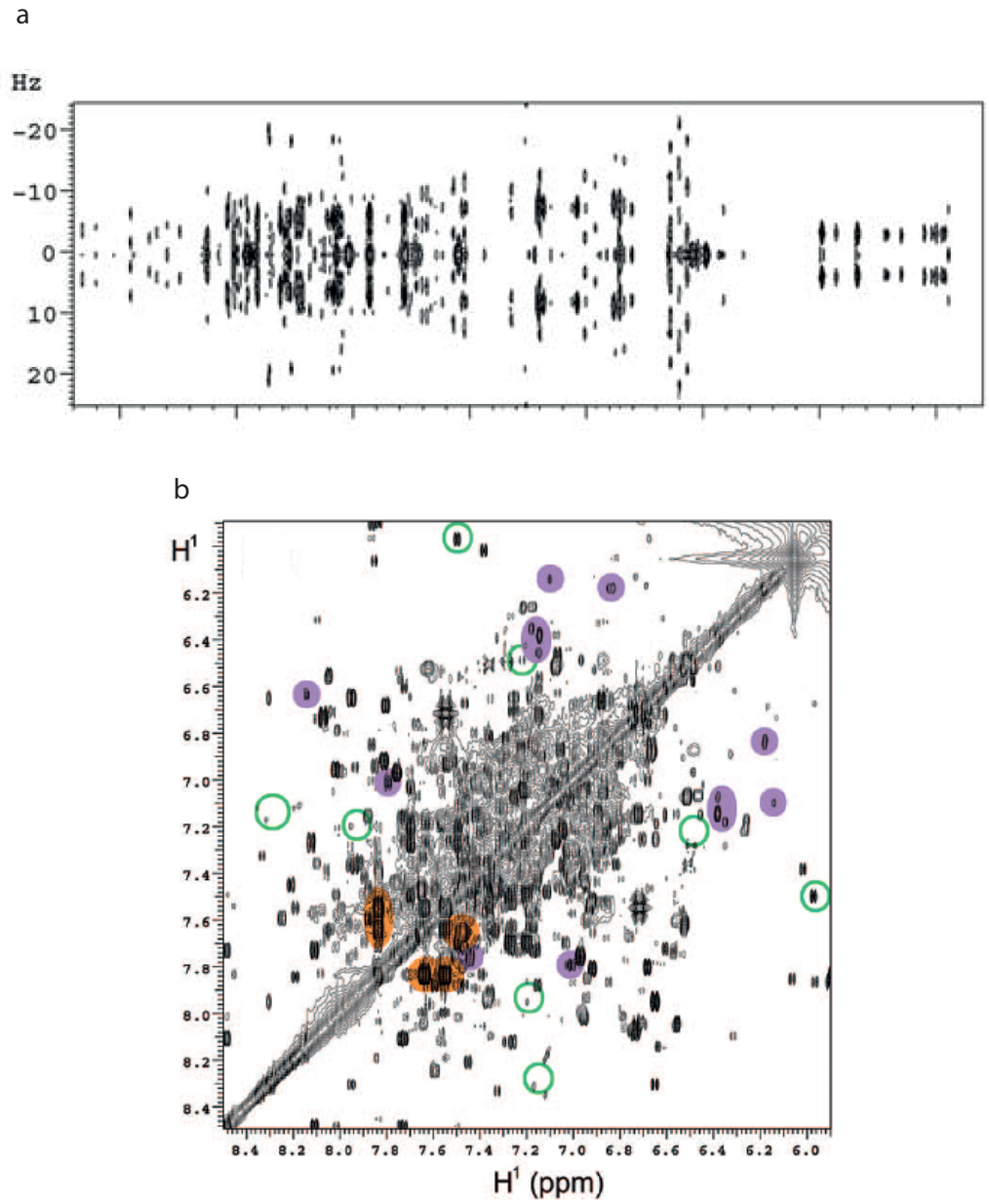


Figure 1.16. J-RES spectrum of a fish embryo extract (a), reprinted with permission from M.R. Viant,[12] Copyright 2003 Elsevier Inc.. TOCSY spectrum of a mouse urine sample (b), reprinted with permission from Q.N. Van et al.,[18] Copyright (2008) American Chemical Society.

However, most metabolomics studies that use the J-RES pulse sequence do not use the 2D spectrum but rather the projection along the direct dimension.[12, 136-139] This projection can be treated like a 1D ^{1}H spectrum with J_{H-H} decoupling, where every peak is a singlet, which simplifies greatly the 1D spectrum. However, like the 2D spectrum of J-RES, the projection is subject to artifacts due to strong coupling, whose number can be reduced by the use of the DSE J-RES pulse sequence. Another problem is that quantification errors due to partial cancellation can occur.[15] Errors up to 20% have been reported.[13, 15] Also, due to the unfavorable phase properties, the spectrum is generally processed in magnitude mode, which produces dispersive tail on each peak that can overlap with other peaks.

Similar to quantification in the previous section, the HSQC pulse sequence is the most used in metabolomics studies with 2D pulse sequences.[24, 25, 27-29, 140-147] Since the ^{13}C dimension gives a higher dispersion of peaks than the ^{1}H dimension, the HSQC pulse sequence gives the largest resolution compared to the other 2D pulse sequences shown thus far. But the increased resolution comes with a sensitivity penalty, especially at natural abundance. For this reason, it has been stated that statistical analysis with HSQC spectra is a good complement to statistical analysis with 1D ^{1}H spectra.[27] To increase the sensitivity of HSQC spectra, one possibility is to perform isotopic labeling with ^{13}C, in order to increase the abundance of ^{13}C. [30] This strategy has already been used in several studies about plants, with the use of $^{13}CO_2$ during the growing phase[148] or ^{13}C glucose.[31, 149-154] ^{13}C labeling can also be added after the sample collection, by chemical reactions.[155] Substantial time-savings during acquisition of HSQC spectra have been reported, albeit with larger costs due to the use of labeled cultures.[30]

It can also be noted that several studies have been conducted in order to automate as much as possible the identification of metabolites in HSQC spectra,[156] as well as automating the analysis of HSQC spectra for statistical analysis.[157, 158] This shows that there is an effort to integrate 2D spectra as a tool for metabolomics that could enhance the amount of information extracted from a metabolomics study.

Another 2D pulse sequence that shows great promises for metabolomics studies is the TOCSY pulse sequence. One of the main appeals of the TOCSY spectrum is the display along rows, where an extended network of protons can be visualized.[159] This extended network is close to the 1D ^{1}H spectrum of the isolated metabolite, giving an increased confidence for metabolite identification compared to some peaks in the 1D ^{1}H spectrum of the complex mixture. Another advantage compared to COSY is the fact that the correlation and diagonal peaks are in-phase, which means that the phased spectrum can be used for volume extraction instead of the magnitude spectrum in COSY or DQF-COSY spectra. As seen in figure 1.16.b, the resolution obtainable in TOCSY spectra allows the visualization of a wealth of peaks with little overlap, even close to the diagonal.

Comparison of statistical analysis between TOCSY and 1D ^{1}H spectra have shown that more potential biomarkers could be identified in the TOCSY spectrum compared to the overlapped 1D ^{1}H spectrum.[18] It has also been shown to be an efficient tool for analysis of minor metabolites

in food analysis, by using band-selective variants of the TOCSY pulse sequence.[19, 160] Like HSQC, a study has been made to enhance the result from statistical analysis of TOCSY spectra,[161] which focused on the alignment of TOCSY spectra in order to reduce as much as possible the area of the 2D bucket for each correlation peaks.

Finally, the TOCSY pulse sequence has also been used for ^{13}C enrichment calculations, in fluxomics studies.[20-22] With proper modification in the TOCSY pulse sequence, it is possible to directly measure the ^{13}C enrichment from the rows of the TOCSY spectrum, which can be used for a better understanding of metabolic pathways.

The DQS pulse sequence has not been used for metabolomics yet, to the best of our knowledge, but a variant of this pulse sequence has shown promises for separation of the molecules in NMR spectra of mixtures, called Maximum-Quantum (MaxQ).[106] The idea is to use the highest quantum coherence possible for a molecule, where only one chemical shift in the indirect dimension is possible, and the row corresponding to that chemical shift shows an extended network of spin, similarly to the TOCSY spectrum. Its potential for the separation of polyaromatic molecules has already been shown,[162, 163] by progressively decreasing the level of quantum coherence to identify different molecules with different levels of maximum quantum coherence. However, for a potential application to metabolomics, it will be necessary to address some shortcomings of MaxQ, particularly its low sensitivity because of the low probability for the magnetization to be in its maximum coherence.

From this analysis of the literature, it is possible to see that 2D NMR is already seen as a promising tool for metabolomics and that its use can bring information that is difficult to obtain with 1D NMR. However, the experimental times necessary to obtain clean 2D spectra are often prohibitive for metabolomics, especially when the number of samples to analyze is more than a hundred or even a thousand. While efforts have been made to optimize the parameters for a fast acquisition of 2D spectra,[13, 26, 164] new improvements could be made by the use of alternate schemes for acquisition of 2D spectra.

3. Strategies for the reduction of experimental times in multi-dimensional NMR

Since the introduction of 2D NMR, several attempts have been made to accelerate the acquisition of multidimensional spectra, by departing from the classic Jeener-Ernst approach. This chapter will list the main alternative schemes available today for multi-dimensional NMR, with a particular emphasis on 2D NMR. This list will be separated in three main groups, depending on the parameter modified by the alternative scheme:

-The first part will focus on schemes that aim to reduce as much as possible the inter-scan delay.

-The second part will show the various ways to reduce the number of t_1 increments to obtain a 2D spectrum without reducing resolution.

-The last part will show 2 schemes that rely on completely different acquisition schemes: Hadamard and ultrafast 2D NMR.

3.1. Reduction of the inter-scan delay

3.1.1. Band-Selective Optimized-Flip-Angle Short-Transient (SO-FAST) and Band-selective Excitation Short-Transient (BEST)

These two schemes have been developed for the NMR analysis of macromolecules like proteins. In most cases for protein analysis by NMR, ^{1}H-^{15}N HSQC or HMQC are used as a starting point. For this acquisition, protons linked to a carbon have no contribution to the acquired signal. Another important part is that protons linked to a ^{15}N are in a different region than the other protons in the ^{1}H spectrum. In other words, it is possible to use a selective pulse to excite only protons linked to a ^{15}N and let protons linked to a ^{13}C along the longitudinal axis, which allows a more efficient way to relax for the protons in the transverse plane.

This possibility is exploited in the SO-FAST HMQC pulse sequence, where the hard pulses for the proton channel are replaced by selective pulses, as seen in figure 1.17. By letting the magnetization of the protons linked to a ^{13}C along the longitudinal axis during the acquisition and the inter-scan delay, it becomes possible to use cross-relaxation to reduce the T_1 of protons linked to a ^{15}N.[165]

Cross-relaxation describes one of the two ways a nucleus with a dipolar coupling with another nucleus can return to the longitudinal axis. For macromolecules, this constant is always negative thanks to its dynamics. Within this configuration, one very interesting result is that, if only one nucleus from the dipolar coupling is in the transverse plane, its T_1 is reduced, as if the nucleus in the longitudinal axis "helps" the relaxation of the other nucleus.[166, 167]

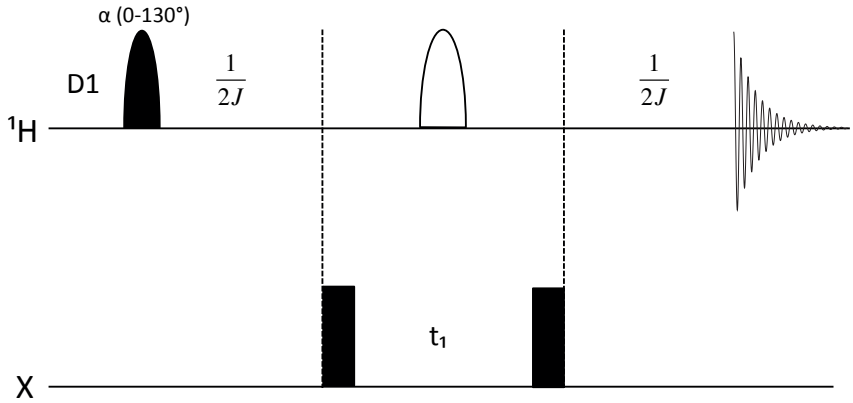


Figure 1.17. Pulse sequence for the SO-FAST HMQC.

Thanks to this scheme, the T_1 of protons linked to a ^{15}N can be very short, allowing a reduction of the inter-scan delay to values close to a hundred milliseconds. Between HSQC and HMQC, the HMQC pulse sequence was chosen since only two pulses are used in the proton channel, allowing little perturbation of protons linked to a ^{13}C if the pulses are calibrated correctly.[168] Another advantage of the HMQC pulse sequence is the possibility to use the Ernst angle for enhanced sensitivity,[169] which has already been used on its own in HMQC for the reduction of TR.[170]

This strategy to reduce the T_1 of macromolecules has been applied for 3D experiments, but without Ernst angle excitation because of the increased number of pulses in the proton channel: these variants are called BEST.[171, 172] Both the SO-FAST and the BEST strategies manage to reduce considerably the T_1 of macromolecules, but further improvements can be achieved with the addition of paramagnetic agents, especially disordered proteins, where paramagnetic agents usually have access to all amino acids of the protein.[173]

3.1.2. Acceleration by Sharing Adjacent Polarization (ASAP)

For small molecules, the scheme used in figure 1.17 will more likely increase the T_1 instead of decreasing it because of the different behaviors of cross-relaxation.[166] However, another scheme exists to decrease the apparent T_1 in heteronuclear experiments like HMQC, when nuclei with low abundance are used, like ^{13}C.[174] This scheme is illustrated in figure 1.18.

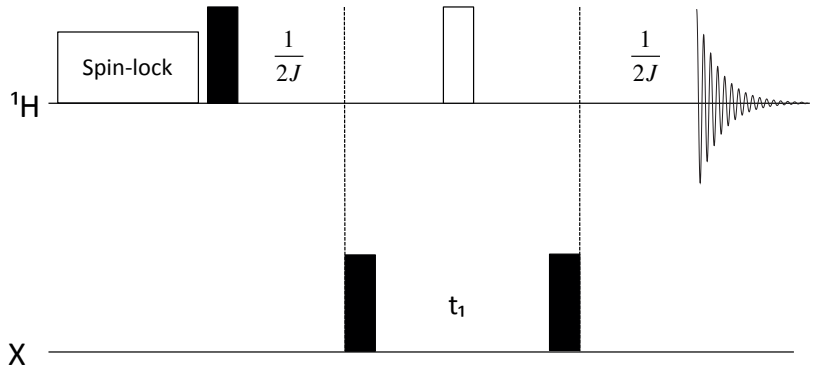


Figure 1.18. Basic pulse sequence for the ASAP HMQC

The main change in the pulse sequence compared to conventional HMQC is the substitution of the inter-scan delay by a spin-lock. The reasoning behind this choice will be explained by two main reasons. First, at ^{13}C natural abundance, the number of groups of protons that give a signal in the HMQC spectrum will likely not be more than one for each molecule. In other words, a part of the proton magnetization in one molecule will not be used during the HMQC pulse sequence. Also, by using the spin-lock between scans, it is possible to transfer the magnetization from unused protons in a molecule to the proton linked to a ^{13}C. Like TOCSY, the duration of the spin-lock rarely exceeds 80 ms.

Since the scheme relies on unused magnetization from other proton, it is necessary to adapt the pulse sequence in order to keep this magnetization along the z-axis during acquisition and not dephasing it during the pulse sequence by saturation or gradients. For this reason, this scheme was first developed for the HMQC pulse sequence.[174] Recently, this scheme has also been implemented in a variant of the HMBC pulse sequence called IMProved and Accelerated Constant-Time (IMPACT) HMBC.[175] For the HSQC, however, the ASAP was initially not considered because of the higher difficulty to keep the unused magnetization intact during the pulse sequence. This difficulty has been overcome recently,[176] thanks to a modification of the pulse sequence, along with the use of optimized adiabatic pulses. This variant of HSQC has already been used for slice-selective experiments.[177]

3.1.3. SMAll Repetition Times (SMART) NMR

When no strategy can be used to reduce the T_1 and very short inter-scan delays (also called fast pulsing since a non-negligible fraction of the experimental time is spent using pulses) are still employed, the resulting spectrum will usually have artifacts caused by the transverse magnetization, which has not relaxed between scans. The SMART NMR is a strategy for several pulse sequences, which aims at suppressing these fast-pulsing artifacts with the use of gradients.[178] This scheme gives important reductions for the experimental time, but has two shortcomings: it needs triple-axis gradients for several pulse sequences in order to properly suppress the magnetization, and since an important fraction of the magnetization is dephased by gradients, the sensitivity of these variants are lower than their conventional counterparts.

3.2. Reduction of the number of t_1 increments

For this chapter, two subgroups can be defined: linear methods, where the indirect FID is still collected the same way and the spectrum created from FT or a different algorithm that uses the same dataset, and non-linear methods, where the indirect FID is acquired differently and a different algorithm is used to obtain the spectrum.

3.2.1. Linear methods

3.2.1.1. Folding-aliasing

The resolution (in Hz) of a peak in a given dimension is related to the maximum evolution time in this dimension. In the indirect dimension, the maximum evolution time AQ_1 is given by:

$$AQ_1 = DW_1 \times N_1$$

where N_1 is the number of t_1 increments and DW the dwell time, which is the delay between two points in the FID. Therefore, reducing the number of t_1 increments can be done without changing the resolution by increasing the dwell time. However, DW_1 is also linked to the spectral width SW_1 because of the Nyquist condition (that can be used in both dimensions):

$$SW_1 = \frac{1}{2 \times DW_1}$$

In other words, increasing the dwell time decreases the spectral width. If the spectral width becomes too small to represent all the range of chemical shifts from a sample, some peaks will be observed at the wrong chemical shift. This phenomenon is called aliasing or folding, depending on the type of scheme used to phase the 2D spectrum.[179] By exploiting this property, it is possible to obtain a spectrum with a high resolution in the indirect dimension with few t_1 increments, but at the cost of a low spectral width, with peaks at the wrong frequency in the indirect dimension. Despite the increased risk of overlap, this strategy can be used to reduce the number of t_1 increments while keeping the same width at half-height[180, 181], as illustrated in figure 1.19. Another option is to increase the resolution and therefore reducing the risk of overlap between peaks with similar chemical shifts with the same experimental time, with strategies to deduce the real chemical shift from the aliased (or folded) spectrum.[182, 183]

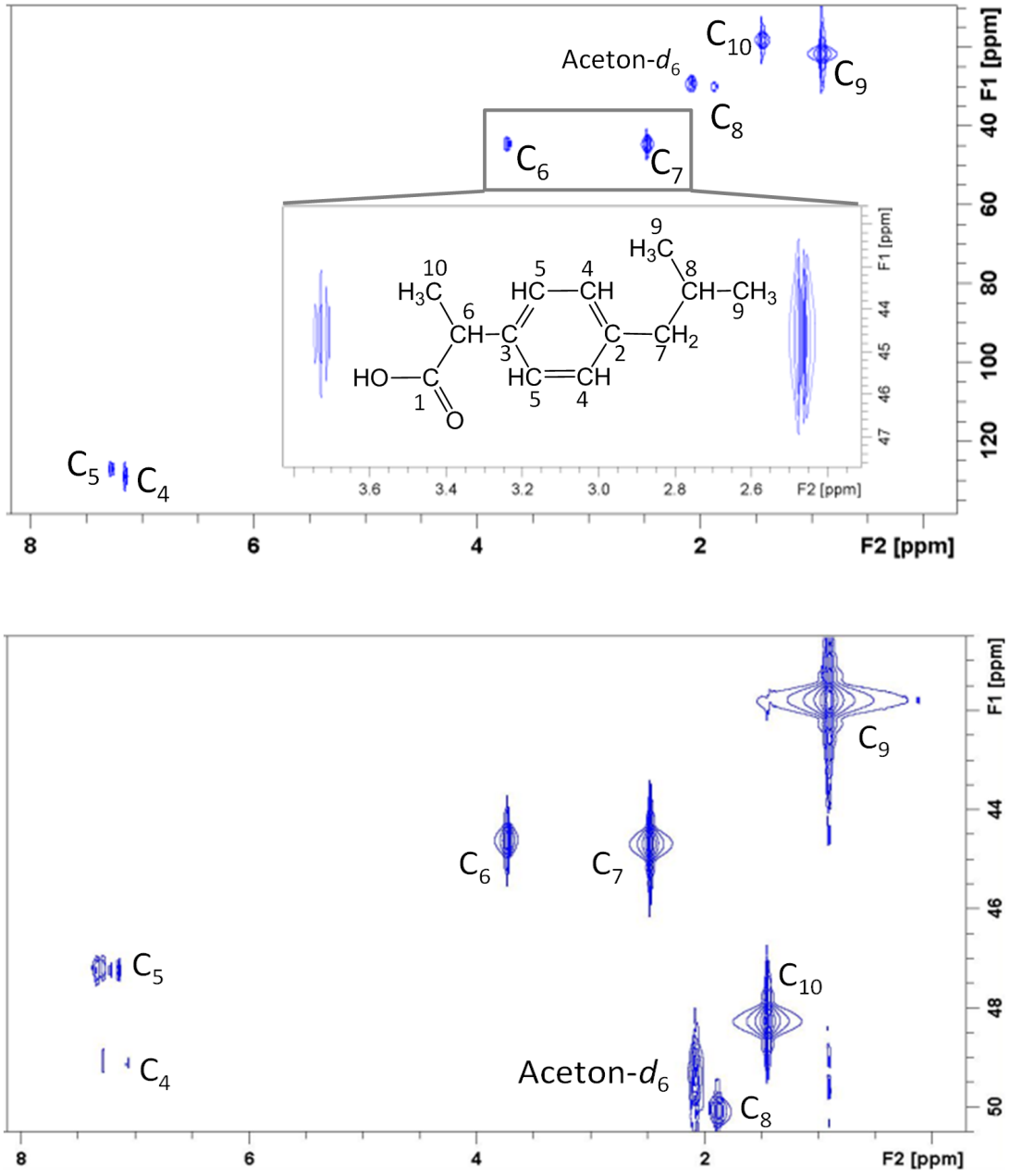


Figure 1.19. Conventional HSQC spectrum of ibuprofen (above), recorded with a spectral width of 129 ppm and 128 t_1 increments, along with the aliased HSQC spectrum (below), recorded with a spectral width of 10 ppm and a number of t_1 increments of 64, dividing by 2 the experimental time. Reprinted with permission from E. Martineau et al.,[181] Copyright 2013 American Chemical Society.

3.2.1.2. Linear prediction

A second possibility to reduce the number of t_1 increments is to use algorithms to guess the position of the next points of the FID from the already acquired points.[184] While this strategy is usually employed to increase the resolution in the indirect dimension,[184, 185] it can also be used to reduce the number of t_1 increments recorded during acquisition, and therefore the experimental time, while keeping the same spectral width and the same resolution.

From this description, it seems possible to use this strategy to draw the entire FID from a few t_1 increments. However, in practice, there is a limit on how much it is possible to expand the FID while keeping the same lineshape in the spectrum.[186] This limit seems to be from 2 to 4 times the number of t_1 increments for most spectra.[185, 187] Finally, with more extended use of linear prediction, or its use in very crowded spectra, the increase in the maximum evolution time does not necessarily leads to the separation of overlapping peaks.[186]

3.2.1.3. Covariance NMR

This method departs slightly from FT, since another algorithm is used to process the spectrum in the indirect dimension. Briefly, the covariance algorithm is close to a symmetrization process, where the resolution in the indirect dimensional is forced to match the resolution in the direct dimension, and is often called direct covariance NMR.[188, 189] Since the resolution in the direct dimension can be easily raised without an increase in the experimental time, this allows keeping the same resolution in the indirect dimension or increasing it with less t_1 increments. Modifications in the algorithm also allow the combination of spectra to produce a new spectrum that would be lengthy to directly acquire, like HSQC-TOCSY, which has been called indirect covariance NMR.[190] This method has already shown to produce high-resolution spectra with a small number of t_1 increments.[191]

It should be noted, however, that with direct covariance NMR, the algorithm keeps the peaks at the same position only if diagonal peaks are present.[192] A modified version of the algorithm that adds diagonal peaks in the spectrum during covariance and suppresses them after covariance has been developed recently.[192]

3.2.2. Non-linear methods

3.2.2.1. Non-Uniform Sampling (NUS)

While the Fourier transform has revolutionized NMR since its first use,[169] its main flaw is the need to have all points of the FID spaced equidistantly in order to use Discrete Fourier transform. While this is rarely a problem for 1D NMR, since the FID is recorded directly during acquisition and therefore extending the FID can be done without time penalties, with nD NMR, this requirement is a constraint for the indirect dimensions. Since the high spectral width usually needed for NMR spectra means that DW is low, a high number of t_1 increments are needed to obtain a good resolution in the indirect dimension in order to correctly separate peaks. For this reason, alternatives to the FT have been searched for a while, even for 1D spectra.[193] These alternative algorithms have been introduced in 2D NMR shortly after its first publications, but with data that could be also treated by FT.[194-197] It was not until 1987 that an alternative to conventional sampling, called exponential sampling, was proposed.[198, 199] This sampling scheme had a non-constant dwell time and therefore the resulting spectrum had artifacts if treated with FT. However, this partial FID could give clean spectra from other algorithms.

This combination of alternate sampling and alternate algorithm for obtaining the spectrum from the FID is called Non-Uniform Sampling (NUS), in contrast to the conventional sampling,

which is often called Uniform Sampling (US). This strategy is shown in figure 1.20 and shows that spectra can be obtained with the same characteristics (same resolution, same spectral width, same peak intensity) with less t_1 increments if we change the algorithm for the processing of the 2D spectrum. Usually, the NUS FID is seen as a fraction of the US FID with the same maximum evolution time. Therefore, 25% of NUS means the NUS FID contains only 25% of the t_1 increments from the US spectrum with the same maximum evolution time and dwell time.

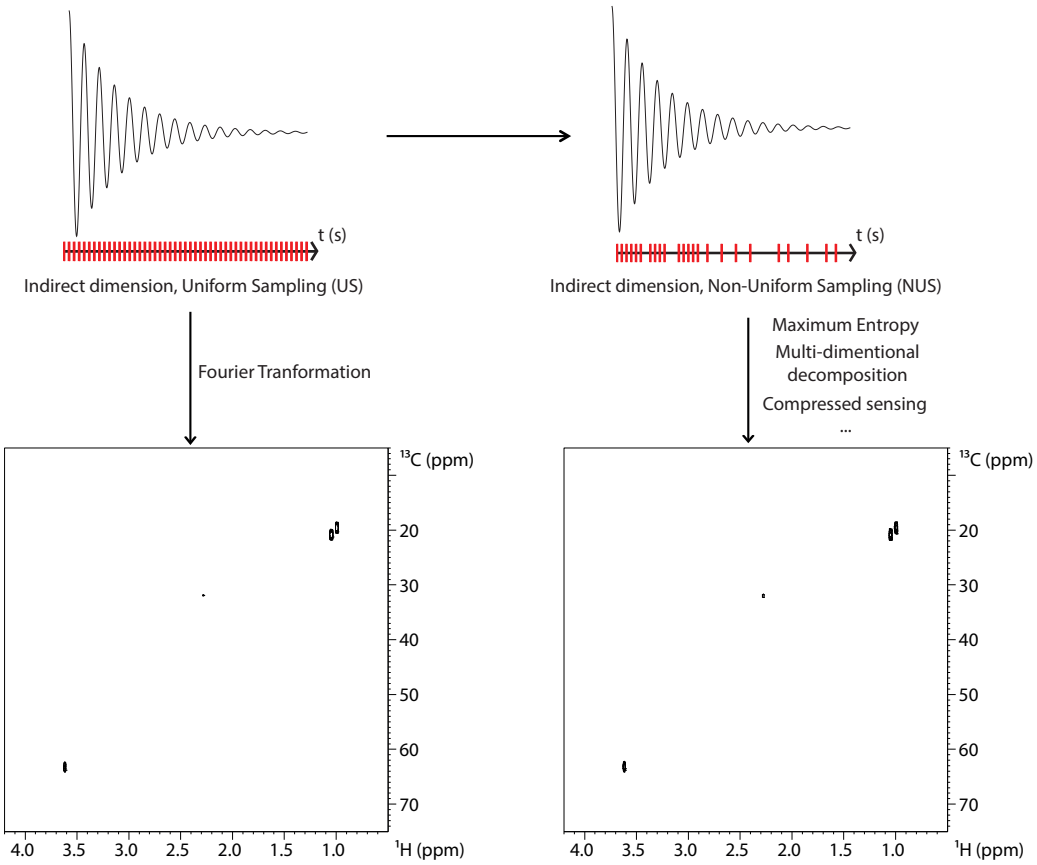


Figure 1.20. Scheme showing the difference between the acquisition in US and the acquisition in NUS. By changing the algorithm for the processing of the FID, it is possible to obtain similar spectra from US and NUS data, as seen for the HSQC contour plot of valine in H_2O/D_2O (90/10), with US (left) and 50% of NUS (right) at the same contour level.

Today, a range of possibilities exists for both the sampling scheme and the reconstruction algorithm. For the reconstruction algorithms, different approaches have been developed: maximum entropy[200] and its variants[201], Multi-Dimensional Decomposition (MDD)[202] and Compressed Sensing (CS)[203, 204]. For the sampling schedules, different algorithms have been used to create them, based on different equations for the probability of a t_1 increment to be recorded during acquisition. The most used today is still the exponential sampling, but several alternatives exists, such as Gaussian distribution[205] or Poisson-gap sampling.[206]

Several studies have already been done for the differences between US and NUS spectra, especially in terms of sensitivity. This part will be developed in part 5.

3.2.2.2. Radial sampling and projection-reconstruction

For 3D spectra and spectra of higher dimensions, the time constraints are even higher than for 2D spectra, since two or more indirect dimensions have to be sampled. These spectra, which need from hours to days to record with US, need drastic time reductions to be considered a viable tool. One of the first approaches that gave these time reductions is radial sampling.[207]

To explain this approach, the example of 3D spectra will be used, whose sampling schedule is shown in figure 1.21.a. For the acquisition of these spectra, two indirect time dimensions are needed, named t_1 and t_2, with t_3 being the dimension directly recorded during acquisition. At first, t_1 and t_2 are set to their minimal value, then all the increments from t_1 are recorded while t_2 is fixed. After t_1 is entirely sampled, the first time increment for t_2 is used and t_1 is entirely sampled once again. This procedure is repeated until all the t_2 increments are sampled. The principle of radial sampling, as shown in figure 1.21.b, is to increment t_1 and t_2 simultaneously, with $\Delta t_1=(\alpha/\beta)\Delta t_2$. α/β is a positive number, often called the projection angle, since this procedure can be seen as taking only a line from the rectangle represented by the (t_1, t_2) plane with (α/β) as the angle of the recorded line with the plane.[208]

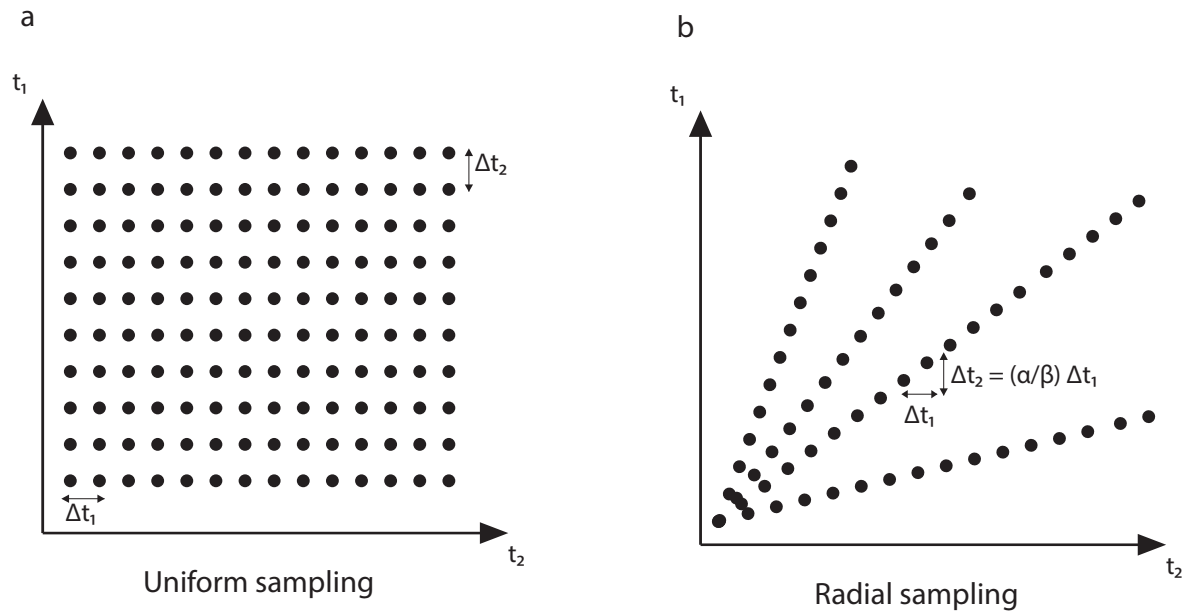


Figure 1.21. Principle of conventional sampling (a) and radial sampling. (b) In 3D NMR. Instead of sampling t_1 and t_2 independently with fixed Δt_1 and Δt_2, the two incremented delays are evolving simultaneously.

The result is a 2D spectrum whose indirect dimension changes with the projection angle. For example, if $\alpha=0$, then the indirect dimension is F_2. If $\beta=0$, then the indirect dimension is F_1. Between these two extremes, the other projection angles will give a mixture of F_1 and F_2 in the indirect dimension. From this plane, it is possible to reconstruct the 3D spectra by taking several planes at different projection angles, then using algorithms to reconstruct the two indirect dimensions from the 3D spectrum.[208, 209]

3.2.3. Alteration of the sampling scheme for the indirect dimension

3.2.3.1. Hadamard spectroscopy

Most 2D spectra are "sparse": an important part of the spectrum is only noise and the fraction of the spectrum containing peaks is small. For these cases, recording selective 1D spectra for the different chemical shifts may be quicker than recording the whole 2D spectrum. But in order to keep full sensitivity from multiplex acquisition, it is necessary to acquire all the selective 1D spectra simultaneously. Hadamard spectroscopy is based on this principle, by recording several selective 1D spectra simultaneously, by replacing the first 90° pulse by a selective pulse.[210] As said in the description of WET in section 1.3, it is possible to excite simultaneously several frequencies with one polychromatic pulse, which is used here to excite the different chemical shifts in the direct dimension. To isolate the result from the different chemical shifts, the experiment is repeated several times, with changes in the relative phases for the different chemical shifts excited by the selective pulse,[211] with the goal to create a Hadamard matrix from these different phases.[212] It becomes then possible to isolate the 1D selective spectrum for each chemical shift and display them as a pseudo-2D spectrum.[210]

With this strategy, the number of t_1 increments for the indirect FID in conventional 2D NMR is replaced by the number of chemical shifts with correlation peaks in the 2D spectrum for the calculation of the experimental time. If this number of chemical shifts is small, important reductions in the experimental time can be expected.

3.2.3.2. Ultrafast (UF) 2D NMR

UF 2D NMR is an approach that uses concepts from Magnetic Resonance Imaging. For this reason, a brief introduction to the k-space is necessary to understand the principle of UF 2D NMR. The k-space is the reciprocal of the Cartesian space and is the space sampled in NMR imaging. The image in the Cartesian space is obtained by FT of the image in the k-space.

In NMR imaging, the k-space is sampled with the use of gradients in the 3 directions of the Cartesian space, since the dephasing of the signal by a gradient in one dimension is the same thing as moving in the k-space in the same dimension.[213] Similarly to 2D NMR, the acquisition of a 2D image is usually lengthier than 1D images, due to the necessity to repeat several times the pulse sequence to map the k-space. However, a method called Echo Planar Imaging (EPI) was developed, which allowed the acquisition of a 2D image in one scan.[213] With this analogy, attempts have been made to transpose this idea for NMR spectroscopy by using variants of EPI, but were restricted to some pulse sequences.[214] A more general approach to obtain a 2D spectrum in 1 scan was developed more than 10 years ago,[213] and uses the k-space to map the chemical shifts, which departs considerably from imaging techniques. The step that allowed this mapping is called spatial encoding, which has been developed by Lucio Frydman and his team from the Weizmann Institute and is presented in figure 1.22.[215] It should be mentioned, however, that the spatial encoding shown in this figure is not the original discrete approach, which has been replaced by a continuous approach, developed by Pelupessy shortly after the first publication of UF 2D NMR.[216]

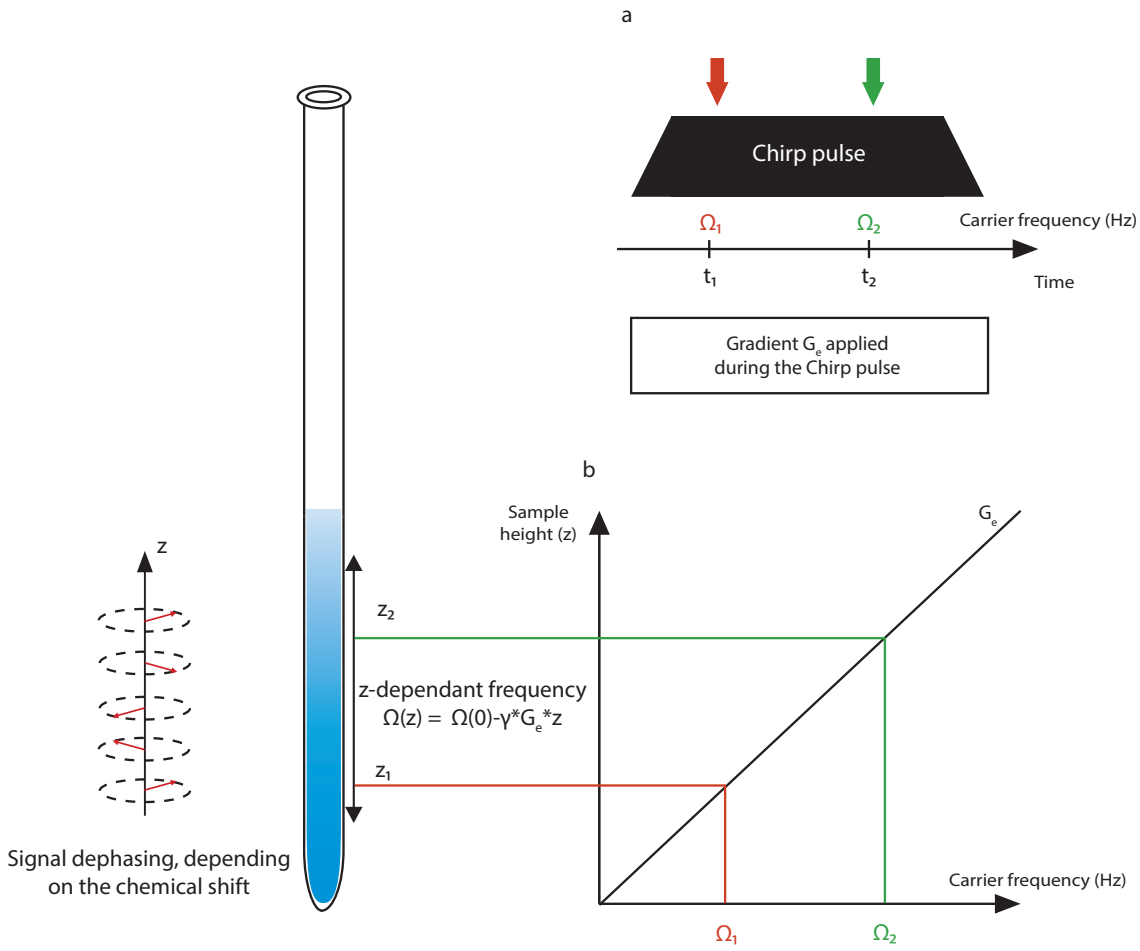


Figure 1.22. Principle of the spatial encoding with the use of a chirp pulse (a) at the same time as a gradient. This combination allows exciting a different part of the sample at different times (b). After spatial encoding, a magnetization helix is created, which is linearly dependent on the chemical shift, by using two chirp pulses applied together with gradients of opposite sign.

Spatial encoding relies on a chirp pulse applied at the same time as a gradient. A chirp pulse is a radiofrequency pulse whose carrier frequency changes linearly with time. When used with a gradient, a spatially-dependent shift is created, whose expression is $\Omega(t) = \Omega_0 + \gamma \cdot G \cdot z$, where Ω_0 is the chemical shift, γ the gyromagnetic ratio of the nucleus observed and G the strength of the gradient. With the combination of the two, it becomes possible to excite different parts of the volume at different times; therefore the result is a dephasing of the signal that is dependent on the chemical shift and the position in the volume. With only one chirp pulse, the dependence between the dephasing and z is quadratic. For this reason, two chirp pulses are used most of the time during spatial encoding. The gradient used at the same time as the second chirp pulse has an opposite sign to the gradient used at the time as the first chirp pulse, which cancels all quadratic terms for the dephasing while keeping the linear term. This makes it possible to obtain a linear dependency between the dephasing Φ and z of the form:

$$\Phi(z) = C \cdot z \cdot \Omega \qquad (1.1)$$

where C is a user-defined constant depending on the duration of the chirp pulse and the length of the coil, and Ω the chemical shift. The result of this dephasing is a magnetization helix, as shown in figure 1.22. Different types of spatial encoding exist,[216-219] most of them relying on the succession of two chirp pulses of same duration with gradients applied at the same time, in order to obtain a linear dephasing with the position in the volume. But the most used currently is the constant-time scheme proposed by Pelupessy,[216] which gave the best results for sensitivity and is overall easier to implement.[220] This spatial encoding scheme is shown in figure 1.23.

This scheme is composed of a hard 90° pulse, to tip the magnetization in the transverse plane, followed by two 180° chirp pulses at the same time as bipolar gradients of opposite times, which allows the spatial encoding. The term "constant-time" comes from the fact that the magnetization spends the same time in the transverse plane, regardless of its position in the z-axis, which is not the case from the other spatial encoding schemes.

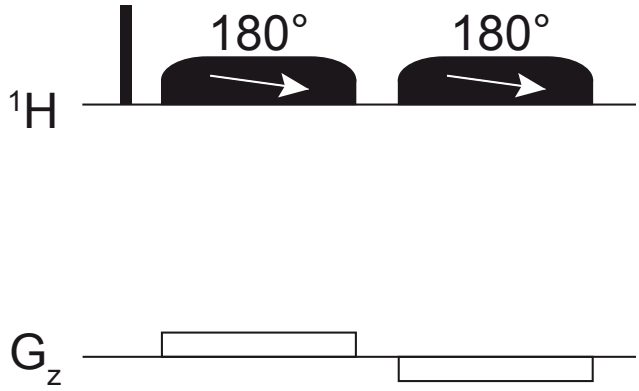


Figure 1.23. Constant-time spatial encoding, as developed by Pelupessy.[216]

Since the signal is dephased, it is necessary to use another gradient to rephase it prior to acquisition. In EPI, this gradient is placed during acquisition, which is the key to record several points in the k-space in one scan. The effect of the gradient on the dephased signal is shown in figure 1.24.

During this gradient, the magnetization helix formed by the spatial encoding is unfolding until a time where the magnetization becomes coherent once again. At this time, a signal is recorded, then for the rest of the acquisition gradient, the magnetization becomes dephased once again. Since the dephasing is dependent on the chemical shift, as seen in equation 1.1, the time when the signal is rephased is also dependent on the chemical shift. In other words, the position of the peaks in k-space is dependent on the chemical shift: k-space in this case is the equivalent of a NMR spectrum in the frequency domain, which does not need FT.

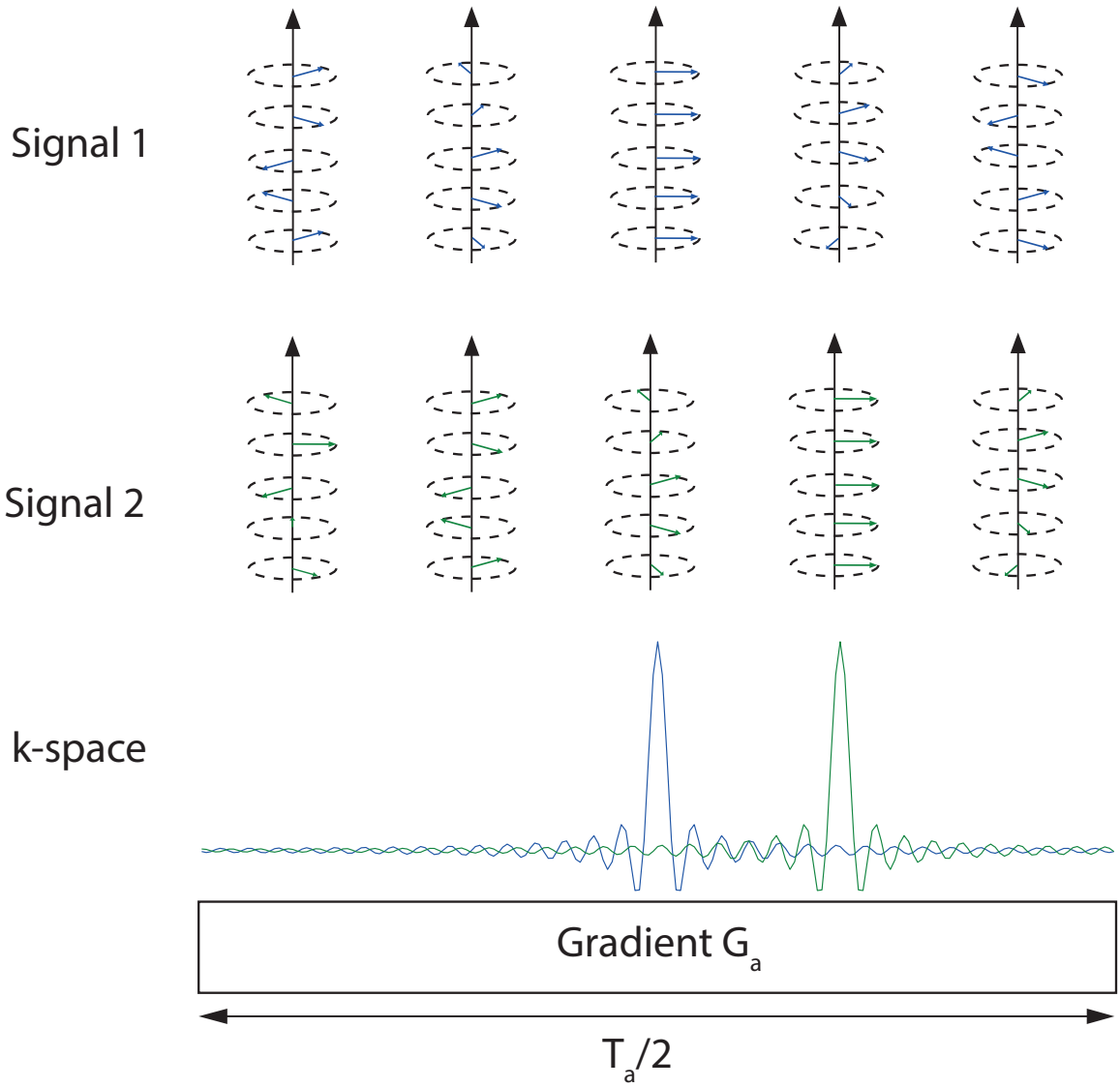


Figure 1.24. Effect of the gradient during acquisition on the magnetization helix from spatial encoding. The helix unfolds until the time where the signal becomes coherent: a peak is formed at this time and then a new helix will be formed. Since the form of the magnetization helix is dependent on the chemical shift, the position of the peak will also depend on the chemical shift.

The k-space becomes one of the frequency domains, but this step alone will not produce the 2D spectrum. In order to produce a 2D spectrum, it is necessary to use several gradients during acquisition. The method, which comes from Echo Planar Spectroscopic Imaging (EPSI),[221] consists in a series of bipolar gradients of opposite signs, but with the same absolute value and length, as shown in the UF COSY pulse sequence, in figure 1.25.a.

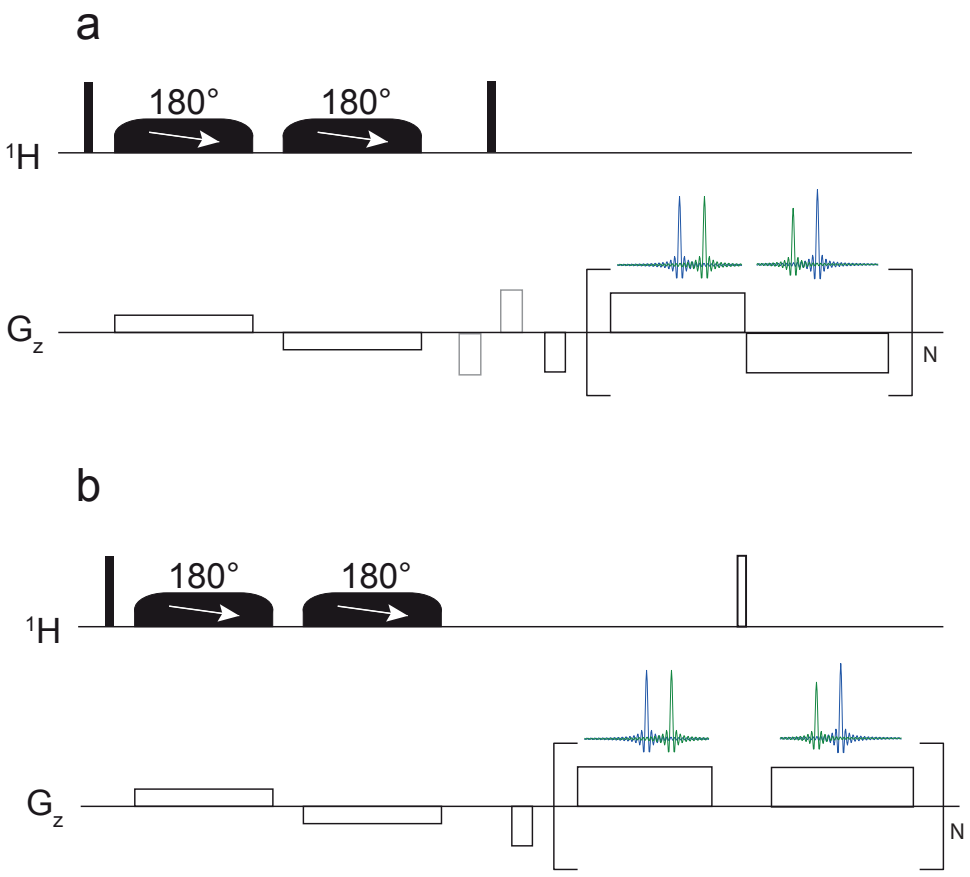


Figure 1.25. Pulse sequence for UF COSY (a) and UF J-RES (b).

This strategy, which exploits the reversibility of the dephasing by a gradient, makes it possible to obtain a second k-space spectrum, which is a mirror from the first k-space spectrum. However, the intensities between the two spectra are different, due to the evolution of magnetization from chemical shifts, J-couplings and relaxation times, which is the key to obtain the second dimension. At the end of the second gradient, the dephasing is the same as before the application of the first gradient. Therefore, it is possible to apply several times this succession of two gradients of opposite signs. During this succession of gradients, the magnetization continues to evolve because of chemical shifts, J-coupling and relaxation. The successive echoes acquired during this series of gradients can be viewed as the successive points of a FID. An example of the signal obtained at the end of the acquisition is shown in figure 1.26.a.

This dataset can be seen as a series of opposite trajectories in the k-space, with evolution of magnetization during the application of gradients, as seen in figure 1.26.b. Therefore, the repetition of trajectories sets a FID that constitutes the second dimension. Since this evolution is the one that should be displayed in the direct dimension in conventional 2D NMR, this dimension is often referred as t_2 or conventional dimension. For the k-space spectrum, since it usually shows the content from the indirect dimension in conventional 2D NMR, it is often referred as F_1 or ultrafast dimension.

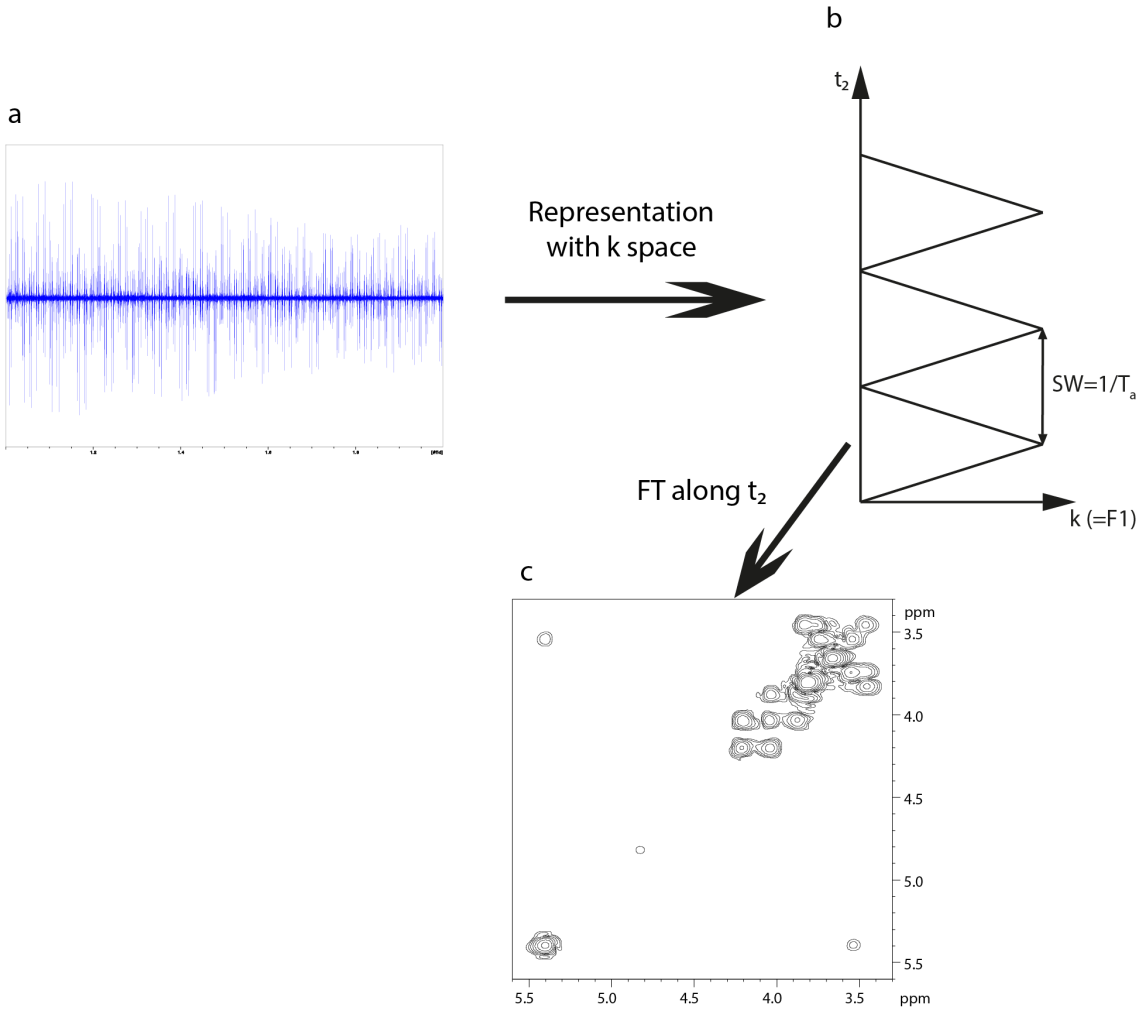


Figure 1.26. Recorded signal for a UF COSY of a sucrose sample (100 mM) (a), which can be represented as a succession of trajectories along k (which is equivalent to F_1 in the conventional dimension) with evolution along t_2 at the same time, which is similar to a FID along t_2 (b), with the spectral width linked to the length of the two gradients used during acquisition. By separating the data from + and - gradients into 2 datasets, and performing FT along t_2, it is possible to obtain the UF spectrum (c). The two datasets can be added for a better sensitivity after correcting for a possible offset between them.

To obtain the 2D UF spectrum from these data, the signal is first separated in two datasets, one obtained by positive gradients and the other by negative gradients. For each dataset, the UF spectrum is obtained by FT along t_2, as shown in figure 1.26.c. After FT, the two datasets can be combined to obtain a better sensitivity.

For the UF J-RES spectrum, the acquisition, as seen in figure 1.25.b, is slightly different. Instead of two gradients with opposite signs, the two gradients have the same sign and a 180° pulse in inserted between the two gradients. Despite this change, the same type of trajectory in the k-space happens, since the 180° reverses the dephasing present at the end of the first gradient, so the second gradient cancels the effects of the first gradient in terms of k-space trajectory. The main difference between the two versions is a change of parameters that affects the evolution of spins during acquisition, which will be discussed in part 3.

While figure 1.25 shows only two UF 2D experiments, UF 2D NMR can be extended to any type of correlation, homonuclear or heteronuclear. We can cite, for example, UF TOCSY,[215] heteronuclear UF J-RES,[222] UF HSQC and UF HMBC.[223]

With this strategy, the 2D spectrum is obtained in only one scan. This makes it possible to use UF 2D NMR for applications where conventional 2D NMR can be unsuitable, such as fast kinetics,[224] observation of intermediates in reactions,[225-227] combination of NMR with liquid chromatography[228] or hyperpolarization.[223]

However, despite this great advantage, UF 2D NMR has some drawbacks. The main disadvantage comes from sensitivity penalties: aside from the obvious losses from having the 2D spectrum in one scan instead of hundreds or thousands of repetitions of the pulse sequence, there are losses arise from diffusion[229, 230] and from the introduction of more noise, since the filter bandwidth needs to be opened more than in conventional 2D NMR.[215] The resolution in the ultrafast dimension can be improved with the length of the chirp pulses during spatial encoding, but this will introduce more sensitivity losses from diffusion,[220] while the resolution in the conventional dimension is limited by the time that the gradients can hold during acquisition.[231]

Finally, the spectral width is limited in both dimensions, and trying to improve the spectral width in one dimension will be at the expense of the spectral width in the other dimension.[231] The compromise between both spectral widths is aggravated when higher magnetic fields are applied, so most UF experiments are done on NMR spectrometers below 600 MHz or for the identification of compounds with low spectral widths needed to see the full spectrum.[228] However, several improvements have been done for acquisition and processing of UF 2D NMR, in order to obtain cleaner spectra,[232] to increase resolution, either by folding gradients,[233] spectral-spatial pulses[234] or interleaving,[215] or to reduce the constraints for spectral widths, along with making UF 2D NMR more user friendly.[37]

Despite this separated presentation of these approaches, several of them can be combined to further decrease the experimental time, such as Hadamard and SO-FAST,[235] or for new applications, such as UF 2D NMR and SO-FAST, for fast kinetics in proteins.[236, 237] Another interesting strategy is hybrid UF 2D NMR, which departs from the single-scan approach, but has interesting characteristics for complex-mixture analysis.[238] This part will be explained in the next section.

3.2.4. Applications of fast 2D NMR approaches for complex mixtures

So far, fast 2D NMR approaches are far from being routine, but more studies have been done lately to show the advantages of fast 2D NMR in complex-mixture analysis. NUS has already been studied to obtain very-high resolution 2D spectra in a relatively short acquisition time,[34] or for quantitative studies.[33] Hadamard has also been employed for targeted metabolomics,[239] and projection of a 3D spectrum to obtain a 2D spectrum with maximum distance between peaks.[240]

As stated in the last section, hybrid variants of UF 2D NMR have a great potential for complex-mixture analysis. The first idea is to increase NS. While departing from the single-scan

approach, it allows a better SNR than single-scan while keeping a low experimental time,[241] which can be tuned as easily as a 1D spectrum depending on the required sensitivity. This approach has been called Multi-Shot Single-Scan (M3S), which has been proved from one of my master thesis to give a better repeatability than a conventional counterpart acquired with the same experimental time, [38] as seen in figure 1.27. This approach has already shown to give valuable data on biological extracts, such as breast cancer cell extracts[38] or tomato fruit pericarp extracts.[242]

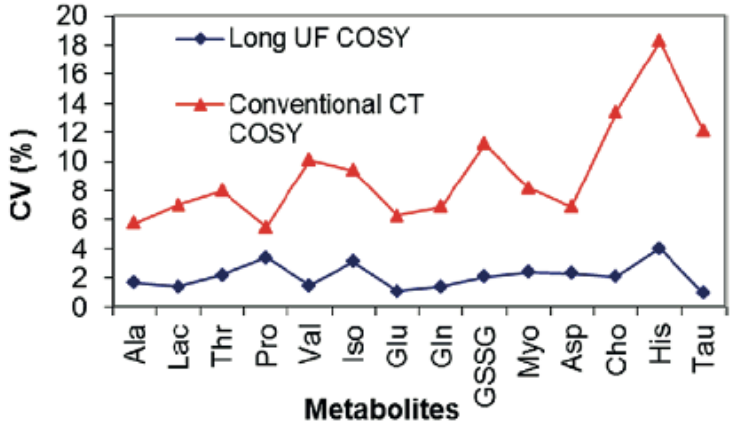


Figure 1.27. Repeatability of the volume for several peaks from 5 spectra obtained with M3S UF COSY and conventional constant-time COSY, recorded in 10 min. Reprinted with permission from Le Guennec et al.,[38] Copyright 2012 American Chemical Society.

UF TOCSY has also been used for fluxomics,[243] using the resulting spectrum in a similar way to the conventional TOCSY spectra for fluxomics mentioned in section 2.3.2. However, the limited resolution makes the UF TOCSY spectrum limited for very crowded spectra. For this reason, another hybrid UF approach has been developed: a tri-dimensional experiment where two dimensions are recorded in an UF 2D NMR manner, while the third one is incremented, like an indirect dimension in conventional 2D NMR.[244] This approach has been shown to work in fluxomics with an hybrid J-RES-COSY pulse sequence.[245]

Finally, covariance processing has been used with ^{1}H-^{1}H TOCSY spectra for a more automated identification of metabolites.[246-250] Recent studies have also been focused on covariance spectra from constant-time ^{13}C-^{13}C TOCSY, in order to obtain as much resolution as possible[251, 252] with potential quantification from these spectra.[253]

From the analysis of the literature, we can see that fast 2D NMR has been mostly used for identification or quantification of metabolites in complex mixtures. But to the best of our knowledge, no fast 2D approaches have been evaluated for non-targeted metabolomics yet.

Part 2. Testing conventional 1D and 2D NMR experiments for metabolomics

Goal: To verify the relevance of 2D NMR for metabolomics with the use of synthetic samples designed for a model metabolomics study between a control group and a diseased group. To reach this goal, statistical analyses from 1D ^{1}H and various 2D spectra were compared to the statistical analysis of the concentrations.

1. Preparation of synthetic samples relevant for metabolomics

In order to correctly evaluate the impact of pulse sequences in NMR for metabolomics studies, it is necessary to have a model statistical analysis as a reference. For metabolomics, this reference is the concentration of metabolites, since this parameter is the one evaluated from the peak areas or volumes. Showing the similarities and differences between the statistical analysis of concentrations and the statistical analysis of the areas (1D NMR) or volumes (2D NMR) should offer insight about the impact of overlap in 1D ^{1}H spectra and the use of 2D NMR. It is therefore necessary to create samples with known concentrations for that study. The design of these samples will be described in the first part, and then the statistical analysis of the concentrations will be performed.

1.1. Designing the synthetic samples for statistical analysis

As mentioned in section 1.5 of part 1, quantification is rarely used in metabolomics. However, thanks to the adaptation of spectral deconvolution for metabolomics,[9, 88] quantification of complex mixtures has become easier. In particular, one of the best-characterized biological samples is serum, thanks to its easy collection and rather simple 1D spectrum compared to many other biological samples. One review for this biofluid provided a list of the most concentrated metabolites, their mean concentration and their standard deviation.[46] From this list, around 30 of the most concentrated metabolites were chosen and their mean concentration multiplied by 20, in order to obtain a synthetic serum with a large dynamic range while keeping the acquisition time reasonable for the 2D spectra, especially for the HSQC. 20 samples were created from these data and constituted the control group. Since, as explained later, the parameter important in the statistical analysis used in metabolomics is the ratio between inter-group changes in concentration to the intra-group concentration changes, as long as these parameters are kept constant, the changes of concentration have no impact and only factors the experimental time of the NMR experiments.

For the diseased group, the quantitative data were taken from another publication, where quantitative differences between control and diseased groups (colorectal cancer) were presented.[254] The changes between groups, called “fold change” in the publication, were used to change the mean concentration of the metabolites in the diseased group and 20 other samples were created with the same standard deviation. The metabolites whose mean concentrations are different in the controlled and the diseased group were first considered as “biomarkers” for this study, which

means that they are considered to have a weight in the group separation in a statistical analysis. To summarize this information, table I shows the mean concentration of each metabolite in the control group, the standard deviation and the fold change between the control and the diseased groups.

In order to create the 40 samples from the information in table I, the following protocol was applied:

- 1 L of a solution of H_2O/D_2O (90/10) was prepared, along with NaH_2PO_4 and Na_2HPO_4, in order to obtain a phosphate buffer of 100 mM with a pH of 7. Sodium azide (around 4 mM) was also added to avoid bacterial proliferation and 0.5 mM of TrimetylSilylPropionate (TSP) for internal calibration of the chemical shifts.

- From this solution, 30 stock solutions were prepared, each containing one metabolite at a concentration 30 times higher than the mean concentration given in table I.

- The 40 synthetic serum samples were obtained by dilutions from the 30 stock samples. Each individual concentration for a sample was obtained by multiplying the mean concentration by a random number, with all the random numbers for one metabolite centered around 1 and respecting the standard deviation for this metabolite.

- The NMR samples were obtained by putting 600 μL of a synthetic serum into a 5 mm NMR tube from Wilmad Lab Glass. The 40 samples were stored in a refrigerator at 5°C until further use.

For most biological samples, concentrations from some metabolites can be correlated for each sample, since they are involved in the same metabolite flux. But for these samples, each concentration was calculated independently from the others. Therefore, there is no correlation between concentrations of the metabolites from the statistical point of view. Another point is that the two groups have the same metabolites: the only differences are the concentration levels. With these samples, the relevance of each metabolite as a potential biomarker in this study could be assessed.

1.2. Statistical analysis of the concentrations

Despite having designed 12 metabolites as biomarkers (see table I), it was obvious that some of them were not going to be relevant for group separation, because of an important standard deviation compared to the fold change. One visible example is creatine, where the standard deviation is around 80% of the mean concentration, while the fold change is around 50% of the mean concentration. Furthermore, since the standard deviations in the two papers we found were probably different, there was no guarantee that some of the biomarkers found in the study with colorectal cancer would be relevant in our case.

Table I. Mean concentration of metabolites in the control group, standard deviation and fold change between concentrations in control and diseased groups. The expected biomarkers have a fold change different from 1. Fold difference (difference between the mean concentration of control and diseased groups, in mM), and fold difference to standard deviation ratio (FD/SD) are also added for a better understanding of figure 2.1. All FC/SD superior to 1 are bold.

Metabolite	Mean Concentration in synthetic serum of healthy patients (mM)	Standard deviation (mM)	Fold change	Fold difference (mM)	Fold change to standard deviation ratio (FD/SD)
3-hydroxybutyrate	1.54	1.33	1.6	0.6	0.47
Acetate	0.84	0.30	1.56	0.4	**1.48**
Alanine	8.54	1.69	0.829	-1.8	**-1.04**
Arginine	2.27	0.29	1	0.0	0.14
Asparagine	1.65	0.15	1	0.1	0.69
Aspartate	0.42	0.12	1	0.0	-0.01
Betaine	1.44	0.45	1	-0.2	-0.35
Carnitine	0.91	0.23	1	0.1	0.43
Citrate	2.28	0.54	0.847	-0.3	-0.52
Creatine	0.73	0.57	0.525	-0.2	-0.30
Creatinine	1.73	0.38	1	0.2	0.41
Glutamate	1.95	0.26	1	-0.2	-0.64
Glutamine	10.21	2.36	0.835	-2.4	**-1.01**
Glycerol	8.63	2.01	1	-0.3	-0.13
Glycine	6.51	2.54	1	0.0	-0.01
Histidine	2.62	0.75	1	0.1	0.20
Isoleucine	1.21	0.37	1	0.0	0.09
Lactate	29.79	7.42	0.807	-2.9	-0.39
Leucine	1.97	0.23	0.852	-0.3	**-1.15**
Lysine	3.57	1.16	1	0.0	0.00
Methanol	1.55	0.33	1	0.1	0.16
Methionine	0.60	0.13	1	0.0	0.08
Ornithine	1.34	0.31	1	-0.1	-0.46
Phenylalanine	1.56	0.41	1.09	0.2	0.58
Proline	3.97	1.30	1.76	3.0	**2.30**
Pyruvate	0.69	0.50	0.801	0.0	0.09
Serine	3.20	0.53	1	0.0	-0.06
Threonine	2.55	0.82	1	0.0	0.00
Tryptophan	1.09	0.19	1	0.0	0.16
Valine	4.25	1.23	0.852	-0.6	-0.50

In order to compare as fairly as possible 1D and 2D spectra, a statistical analysis of the concentrations was necessary. Since studies in metabolomics use areas or volumes of NMR spectra as estimators for the concentration, the direct analysis of concentrations will give an unbiased view of concentration variance in the samples. The table containing the concentrations was exported to SIMCA, where the data were scaled with Unit Variance (UV). Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) was used for statistical analysis in order to maximize the separation between the two groups. The score plot and the loading plot for the OPLS-DA of concentrations are presented in figure 2.1.

From the score plot, we can see a clear separation between the two groups along the first dimension and no visible outlier is detected. The loading plot shows that five metabolites are mainly responsible for the separation: proline, acetate, glutamine, alanine and leucine. These metabolites were among those that were chosen as potential biomarkers. Meanwhile, the seven other potential biomarkers have a negligible impact on separation, even some with high fold change, like 3-hydroxybutyrate. A Variable Importance for the Projection (VIP) plot was used to show the parameters that had a real impact in the OPLS-DA model.[255] It confirmed that only the five cited metabolites were important for the separation of the two groups.

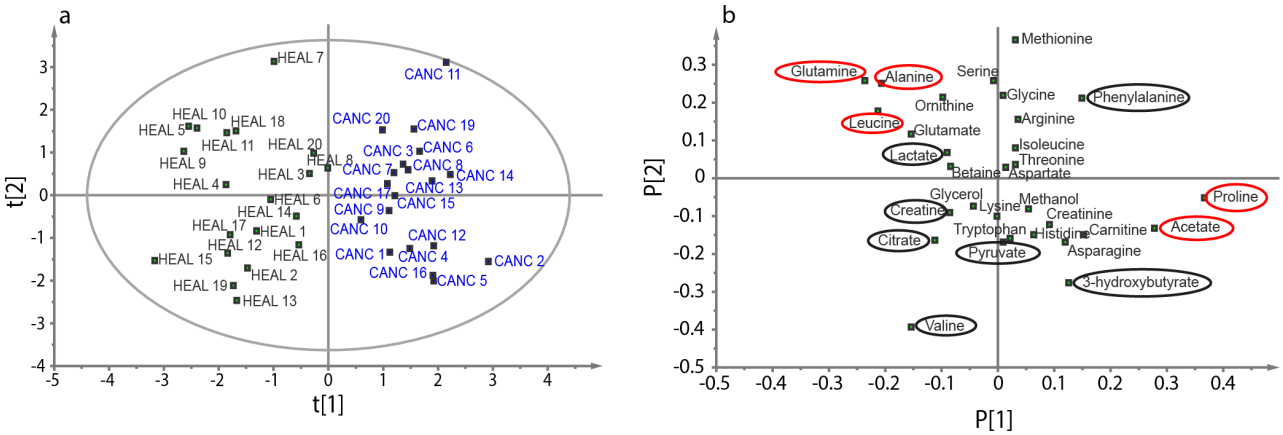


Figure 2.1. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the concentration of 40 synthetic serum samples. In (a) , the 95% tolerance for possible outliers is shown based on Hotelling's T^2 distribution. In (b), the metabolites circled in red are the most important for the separation of control (HEAL) and diseased groups (CANC). The metabolites circled in black are the other potential biomarkers defined in table I, with no importance in the separation. The importance of each metabolite in the separation was analyzed with Variable Importance for the Projection (VIP) plot from SIMCA. The first dimension explains 9.67% of the total variance while the second dimension explains 6.58% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $9.72 \cdot 10^{-4}$.

To explain the difference between fold change and importance in the separation in metabolomics, the standard deviation has to be taken into account. As seen in table II, the 5 metabolites which are important for the separation have a fold difference to standard deviation ratio (FD/SD) higher than 1. Therefore, a metabolite must have a fold difference higher than its standard

deviation to have an impact during statistical analysis. In other words, OPLS-DA helps uncovering metabolites that have higher inter-group variability than intra-group variability. One interesting note is that the coefficient in the first dimension of figure 2.1 seems to be directly correlated to the FD/SD value. Therefore, in the following chapters, changes in the coefficient of the first dimension will be interpreted in terms of changes in the FD/SD value. The importance of standard deviation explains also why the design of experiments (DoE) is critical in metabolomics. Since many parameters can impact the metabolite concentration, like diet or age, a careful design of sample acquisition can reduce drastically standard deviation within a group.

Thanks to this analysis, we know that there are differences between the two groups that have been built, and the important biomarkers for the separation (proline, acetate, glutamine, leucine and alanine). We also have a reference for the statistical analysis of these samples. Now we can judge the effect of different pulse sequences on the statistical analysis.

2. One dimensional spectra, statistical analysis and limits of 1D NMR for metabolite identification

Most metabolomics studies with NMR use a 1D ^{1}H pulse sequence to obtain the spectra used in the statistical analysis, with the aim of comparing concentrations. It is therefore necessary to compare the statistical analysis of a set of 1D ^{1}H spectra with the statistical analysis of the concentration. It will allow us to verify if there are limits to the 1D ^{1}H spectra that make relevant the use of 2D spectra. The 1D ^{1}H pulse sequence will be discussed at first, and then the 1D ^{1}H spectra will be analyzed, to draw conclusions about the samples that we used. Finally, the statistical analysis of the 1D ^{1}H spectra will be discussed to verify if there is a need for 2D spectra for the statistical analysis.

2.1. Acquisition parameters

All the experiments described in this chapter have been acquired on a Bruker Avance I 600 MHz equipped with a 5 mm TXI cryoprobe triple resonance ($^{1}H/^{13}C/^{15}N$). For this study, the Bruker pulse sequence "noesypr1d" was used, since it is one of the most commonly used pulse sequence in metabolomics, as said in section 1.1 of part 1.[67-69] The mixing time was 1 ms, contrary to most cases where the mixing time is usually 100 to 150 ms. This value was chosen from a review of that pulse sequence, where short values were recommended.[69] Each spectrum was acquired with 4 dummy scans, 8 scans, 32768 points during 2.2 s and an inter-scan delay of 3 s for relaxation and presaturation of the water peak. The spectral width was 12 ppm. The experimental time was 1 min and 3 s for each spectrum.

Before Fourier transformation (FT), the FID was weighted by an exponential function of 0.3 Hz, but no zero-filling was applied. After FT and phase correction, the baseline was automatically corrected by a polynomial function of order 2 and the chemical shifts were calibrated to the TSP. An example of spectrum after processing is shown in figure 2.2.

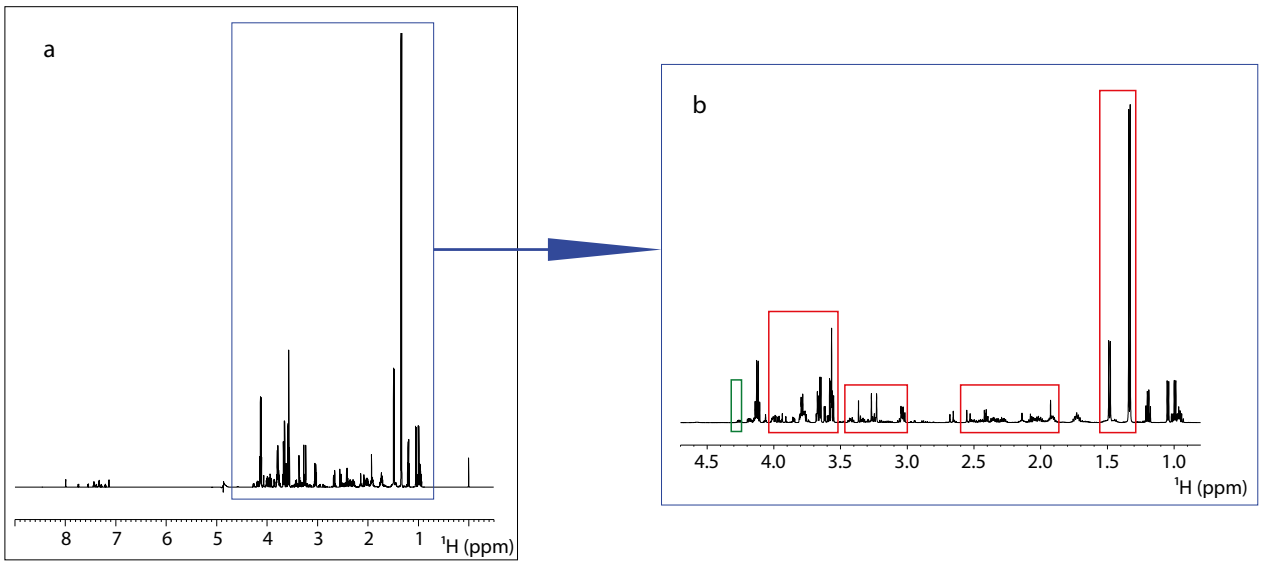


Figure 2.2. 1D ^{1}H spectrum of one of the synthetic serum acquired with the "noesypr1d" Bruker sequence (a) with a zoom on the aliphatic portion of the spectrum (b). The zones framed in red are regions of the spectrum with a high overlap of peaks. The peak framed in green is an isolated peak from pyroglutamate.

2.2. Analysis of the 1D spectrum, discussions and conclusions for the composition of the synthetic serums

We can see in this spectrum that the water suppression was efficient and faraway water has no significant contribution in the spectrum. The phase of the residual water peak (at 4.85 ppm) was not completely the same for all spectra, but no phase distortion was seen below 4.65 ppm. Since the closest peak is at 4.57 ppm, we can conclude that the water peak does not distort the spectrum. As explained in section 1.5 of part 1, 1D spectra of complex mixtures are characterized by a high overlap between peaks.[5] More than half of the 1D spectrum contains a high number of overlapping peaks, from 3 to 8 in the same region. Since many metabolites in serum are amino acids, one of the most crowded regions is between 3 and 4 ppm, where -CH groups from the backbone are found.

From the 1D spectrum, we can see at least one peak (framed in green in figure 2.2), which does not correspond to one of the defined metabolites. After analysis from 2D spectra, it turned out that this peak is from pyroglutamate, a degradation product of glutamine. It is well-known that in solution, free glutamine degrades into ammonia and pyroglutamate.[256] Usually the degradation kinetics is slow as long as the solution is stored in a freezer below 10°C. But this reaction is catalyzed by phosphates, which are used as buffer in our samples.[257] Since several months have passed between sample preparation and NMR analysis, a non-negligible fraction of glutamine has been already converted into pyroglutamate. Because glutamine was one of the biomarkers in our samples, pyroglutamate was included in the list of potential biomarkers.

With these samples, all expected peaks were detected with a signal-to-root mean square noise ratio (S/N_{RMS}) higher than 30 in most cases with 8 scans. For this reason, the delay between

scans was not optimized for sensitivity. It must be noted, however, that the choice of the acquisition time depends on the range of T_2* and that for samples with lower T_2*, like samples with paramagnetic ions, shorter acquisition times are recommended. From the analysis of lineshapes, it is likely that most T_2* were higher than 1 s, therefore most T_1 were likely higher than 1 s, which means we were probably under partial saturation conditions for many peaks. However, it had been shown before that the T_1 were not different under a range of concentrations,[38] so the modulation of peak areas by T_1 will not factor during the statistical analysis.

2.3. Bucketing and statistical analysis

Since all the expected peaks were visible, even for the more diluted metabolites (aspartate and carnitine, for example), statistical analysis was carried out with these spectra. Areas were extracted by manual bucketing into the integration tool of Topspin. Thanks to the DQF-COSY and the HSQC spectra, the position of each multiplet was known, and this knowledge was used during bucketing. Each time that a region of the spectrum contained an isolated multiplet or an isolated part of a multiplet, a bucket was centered on that region. Otherwise, a bucket was centered in a region with overlapping peaks, while still trying to reduce as much as possible the number of overlapping peaks in that bucket. With this mindset, 82 buckets were defined. The raw areas were normalized relatively to the total area, and then exported to SIMCA, where the analysis was similar to the one with the concentrations. The score plot and the loading plot for the OPLS-DA are presented in figure 2.3.

For this statistical analysis, 1 of the 40 samples was a clear outlier. After analysis of the data and the spectra, it turned out that this sample had a leucine concentration twice higher than intended. It has been probably an error during sample preparation, and the most likely reason is that during the dilution of the stock solutions, the sample has been spiked twice with the leucine solution. As a result, that sample was not included in any statistical analysis and only 39 of the 40 samples were used.

As seen in figure 2.3.a, there is still a clear separation of the two groups in the score plot with OPLS-DA. Therefore, there are buckets with a clear difference in areas between the control group and the diseased group. The loading plot, in figure 2.3.b, shows that the buckets from five metabolites are the main cause for separation: pyroglutamate, alanine, leucine, acetate and proline. While most of these buckets contain areas from at least two metabolites, it was possible to see clear differences between the two groups, justifying the use of 1D NMR for metabolomics even with peaks overlap.

As stated before, pyroglutamate is one of the biomarkers in this study. In fact, its relative weight for the separation compared to alanine and leucine is similar to the weight of glutamine in the concentration study. Therefore, we can conclude that pyroglutamate is probably a biomarker with a similar strength than glutamine. However, there is no bucket from glutamine, which does not appear to have any weight in the separation. Four buckets contains areas from glutamine peaks, with one of them containing areas from a pyroglutamate peak, another one with alanine and 6 other metabolites, and the last two have heavy overlap with other metabolites. These last two have little

impact in the group separation; therefore it is impossible to say if glutamine can be still considered a biomarker.

This analysis of glutamine shows the limits of 1D NMR for metabolomics. Since many peaks overlap, there is an increased difficulty to identity which metabolites contributes for group separation in a given bucket. Another risk is the dilution of information, since each overlapping peak acts as an additional source of random variation in the area of the bucket. This additional variation can create an increased standard deviation and therefore losing potential information since the FD/SD can become too low to be relevant in the OPLS-DA.

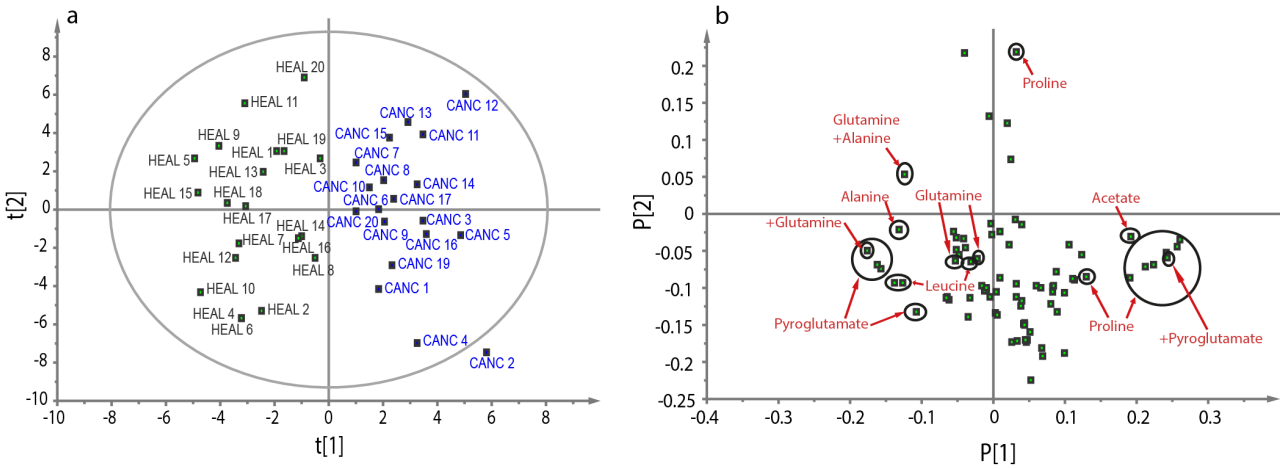


Figure 2.3. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the areas of peaks from the 1D spectra of 39 synthetic serum samples. In (b), the buckets circled in black contains areas of peaks from the biomarkers defined in chapter 1.2 plus pyroglutamate (see text). Overlapping between biomarkers is indicated, but not between biomarkers and other metabolites, for the sake of clarity in the figure. The first dimension explains 11.9% of the total variance while the second dimension explains 15.8% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $8.73 \cdot 10^{-7}$.

Another proof of that loss of information is the analysis of proline peaks. As shown in section 1.2, proline is the metabolite with the highest FD/SD. But the buckets containing areas from proline peaks have a variety of coefficients in the first dimension, due to various degrees of overlap with other peaks. In particular, two buckets have a significantly low coefficient, at least too low to be proved important in a VIP analysis. The first bucket contains only a fraction of the multiplet from the proline peak. The other one, which has the highest coefficient in the second dimension, overlaps with a lactate peak. Since lactate is the most concentrated metabolite in the sample, the variations from the lactate peak cause the fold change from the proline peak to become negligible before the standard deviation from that bucket. The same logic applies to a bucket containing areas from a leucine peak, which overlap with 3 other peaks.

Despite the good separation of the two groups, we see that the heavy overlap between peaks can induce an ambiguity for metabolite identification. This limit justifies the use of 2D NMR within the context of this study.

3. Two dimensional spectra, statistical analysis and comparison of different pulse sequences

Since the resolution is an important parameter for the statistical analysis, the increased resolution of 2D NMR should be beneficial for statistical analysis. In order to see more efficiently the effect of the increased resolution, 3 different pulse sequences were chosen for the study, with different types of spectral widths in the indirect dimension. The first was homonuclear J-RESolved (J-RES) spectroscopy, where the indirect dimension contains the $^{n}J_{H-H}$ couplings, which are rarely above 15 Hz. The second was a homonuclear correlation pulse sequence. The choice was between COSY and TOCSY, and for this first study, COSY was chosen since the fast 2D NMR counterparts used in part 3 were more efficient than the TOCSY counterparts. The last one was the HSQC, which gives a higher resolution thanks to the ^{13}C spectral width spreading over more than 200 ppm. The parameters in the 2D pulse sequences (number of scans, t_1 increments and inter-scan delay) were generally chosen to reduce as much as possible the acquisition time while keeping a good resolution and avoiding the formation of artifacts.

3.1. J-resolved pulse sequence

J-RES spectroscopy is one of the easiest pulse sequences to optimize for fast acquisition without the use of fast 2D approaches.[13] For this study, the Double-Spin Echo variant (DSE J-RES)[14] was used to reduce the number of artifacts caused by strong coupling. Water suppression was carried out by WET followed by presaturation. WET allowed suppressing as much as possible faraway water without sensitivity penalties. Acquisition in the indirect dimension was done by single channel detection, as usual for homonuclear J-RES.

3.1.1. Parameters for acquisition and processing

The DSE-J-RES pulse sequence was obtained by modifying the Bruker pulse sequence with the addition of a second 180° hard pulse during the mixing time and a change in delays to agree with the pulse sequence shown in figure 1.8. All the DSE J-RES spectra were acquired with 17480 points in the direct dimension during 1.37 s and 32 points in the indirect dimension, with a spectral width of 11 ppm in the direct dimension and 50 Hz in the indirect dimension. 16 dummy scans were used for magnetization equilibrium and 8 scans for cycle phasing. The delay between scans was 3 s, which was also used for presaturation of the water peak. The shaped pulses were 90° SElective Decoupling Using Crafted Excitation (SEDUCE) pulses,[258] which lasted 41 ms in order to target the water peak. The experimental time for each spectrum was 21 min and 30 s.

After acquisition, the FID were weighted with a $\pi/2$ shifted sine-bell in each dimension, then the number of t_1 increments was doubled by linear prediction to reduce the width of the peaks in the indirect dimension, making the bucketing process easier. The FID was then zero-filled to 32768 points in the direct dimension and 128 points in the indirect dimension. After FT, tilting and symmetrization were applied to suppress the majority of the strong coupling artifacts. The chemical shifts were calibrated to the TSP and a baseline correction of order 1 was applied. The spectrum of a synthetic serum after this procedure is shown in figure 2.4.

3.1.2. Results and statistical analysis

We can see from this spectrum that while isolated peaks are more numerous in the J-RES spectrum, there are still several highly congested zones. It shows that while J-RES allows for a better discrimination of peaks, there are still risks of a high overlap of peaks. The strong coupling artifacts are mostly suppressed, but for the more strongly coupled protons, their resulting peaks still contain artifacts. For these samples, the most common protons causing these artifacts are non-equivalent protons from a -CH_2- group. Metabolites containing such groups include serine, citrate, asparagine, proline and phenylalanine.

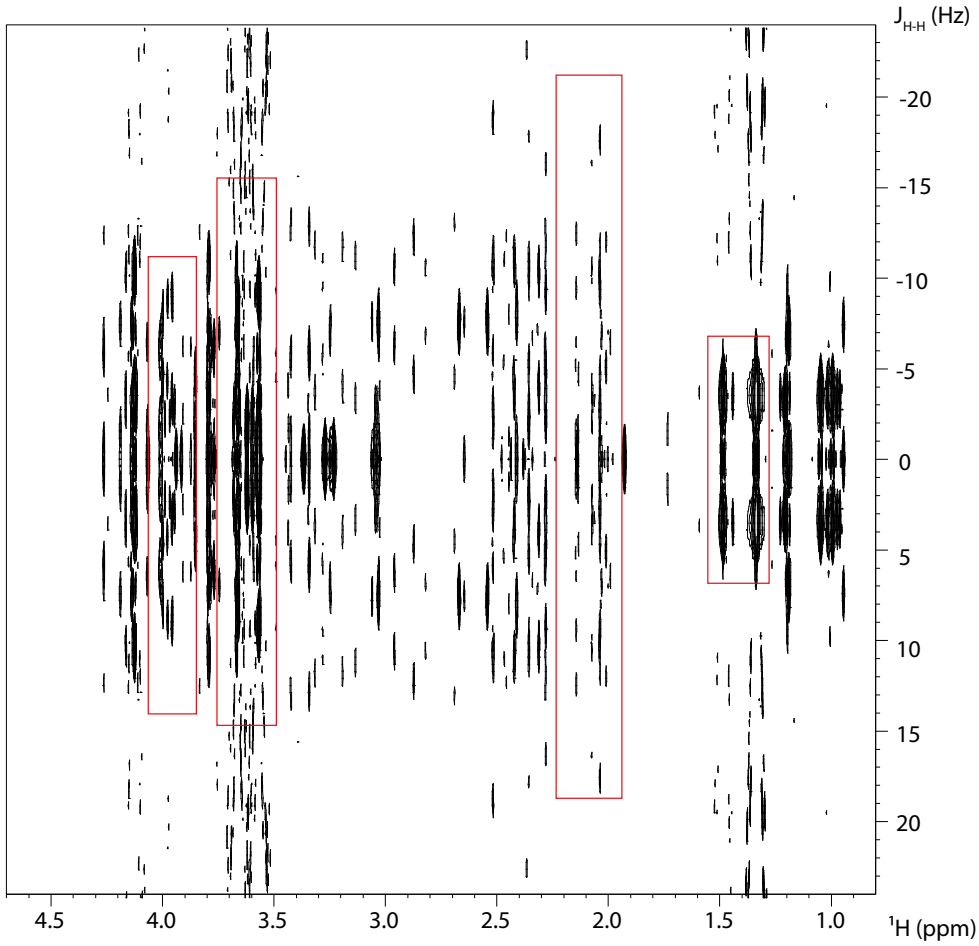


Figure 2.4. Homonuclear DSE J-RES spectrum of one of the synthetic serum with a zoom on the aliphatic portion of the spectrum. The zones framed in red are regions of the spectrum with a high overlap of peaks

Despite these shortcomings, identification of peaks was easier than in the 1D spectrum, with the exception of the very strongly coupled peaks mentioned above. The water peak did not interfere with any peak from the metabolite, and since a combination of presaturation and WET was used, the peak volume of metabolites was not affected by the water suppression scheme. In fact, this combination was more efficient in the J-RES pulse sequence than with 1D ^{1}H spectra, perhaps because of the phase cycling of DSE J-RES is more adapted to this water suppression scheme, compared to the 1D ^{1}H pulse sequence.

For statistical analysis, the peak volumes were extracted by manual bucketing in the integration tool of Topspin. Then the data were normalized relatively to the total volume and exported to SIMCA. Data were scaled to the UV and analyzed with OPLS-DA. The score and loading plots are presented in figure 2.5.

The manual bucketing was done with a similar mindset to the manual bucketing of the 1D spectrum. When a correlation peak was completely resolved, the entirety of the peak was taken as one bucket. A rectangular zone containing all the peak volume above the noise level was used to delimit the bucket. In case of overlap, if it was partial, the focus was to delimit a zone of the spectrum containing the volume from only one correlation peak. That fraction needed to include at least the center of the peak. The overlapped fraction constituted another bucket. If it was impossible to separate even a fraction of a correlation peak, the entirety of the overlapping signal was considered as only one bucket. Eighty-six buckets were defined this way in the J-RES spectrum.

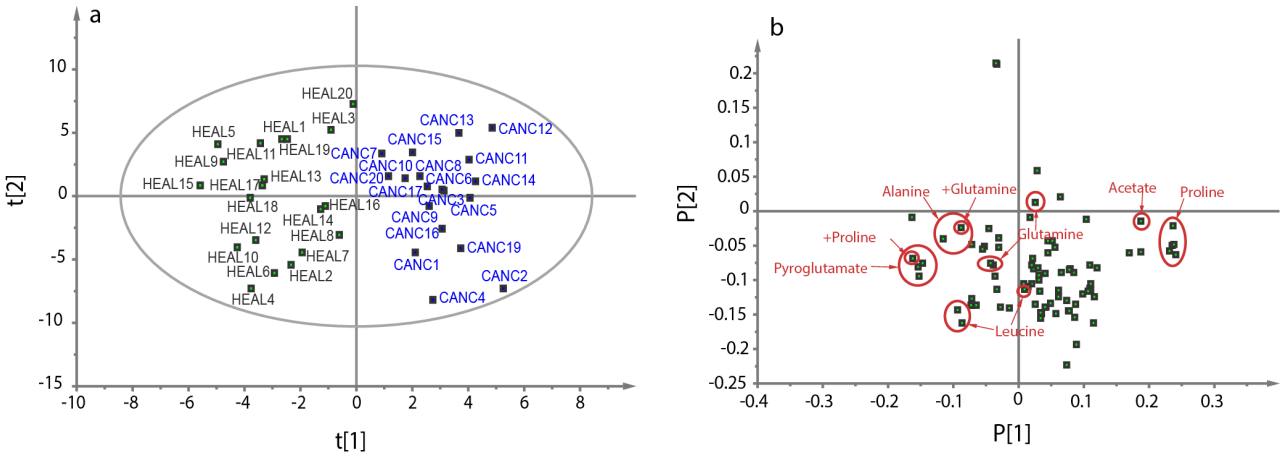


Figure 2.5. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the volumes peaks from the J-RES spectra of 39 synthetic serum samples. In (b), the buckets circled in black contain volumes of peaks from the biomarkers defined in chapter 1.2 plus pyroglutamate (see text in chapter 2.1). Overlapping between biomarkers is indicated, but not between biomarkers and other metabolites, for the sake of clarity in the figure. The first dimension explains 12.4% of the total variance while the second dimension explains 26.6% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $8.73 \cdot 10^{-7}$.

Once again, we can see a clear separation of the two groups in the score plot, and in the loading plot, the buckets responsible for the group separation come from the same biomarkers identified in section 2.3. One improvement over the 1D spectra is that the buckets coming from one metabolite are closer in the loading plot, making an identification of the potential biomarker from chemical shifts easier. Despite these improvements, some buckets containing the volume from a biomarker still suffer from overlap and a loss of information occurs because of this overlap. All the glutamine peaks are still concerned from this limitation, along with one peak from leucine.

With J-RES spectra, we obtain an improvement in biomarker identification, but some ambiguity remains towards some biomarkers, especially glutamine. Other 2D pulse sequences were tested to obtain more discrimination between peaks.

3.2. COSY pulse sequence

The COSY pulse sequence is one of the most used 2D pulse sequences and a wealth of variants has been developed. Among these variants, we chose the Double-Quantum Filtered COSY (DQF-COSY),[101] since it reduces the volume of diagonal peaks and makes the phase from diagonal and correlation peaks similar. The latter advantage makes it possible to phase the 2D spectrum without dispersive peaks, which is helpful in the case of metabolomics since it increases the effective resolution in the spectrum.

3.2.1. Parameters for acquisition and processing

The Bruker sequence used for this study was "cosydfphpr", with the phase cycling for the presaturation pulse taken from the pulse sequence "cosydfgpph19", for reasons explained in the next section. 4096 points were acquired during 0.31 s, with a spectral width of 11 ppm for both dimensions. Sixteen dummy scans were executed for magnetization equilibrium, then 256 t_1 increments were acquired with and 8 scans per increment. The delay between pulses was 3 s, which was also used for presaturation of the water peak. The experimental time for each spectrum was 1 h 55 min and 19 s.

After acquisition, the FID was weighted with a sine-bell function in each dimension, then the number of t_1 increments was doubled by linear prediction. The FID was then zero-filled to 8192 points in the direct dimension and 1024 points in the indirect dimension. After FT, the chemical shifts were calibrated to the TSP and an automatic baseline correction of order 2 was applied in both dimensions. Phase correction was applied, but the spectrum was presented in magnitude mode in order to avoid the cancellation of positive and negative peaks in the total volume for the statistical analysis. The spectrum of a synthetic serum after this procedure is shown in figure 2.6.

3.2.2. Results and discussion

As seen in this figure, there is t_1 noise in the spectrum, resulting from the more intense diagonal peaks in the spectrum, such as lactate and glycerol peaks. This results from the high concentration of these metabolites (see table I) along with the long experimental time, which made the spectrum sensitive to the spectrometer instabilities. Despite the t_1 noise, all expected correlation peaks were visible in the spectrum. Because of the high dynamic range in the sample, the t_1 noise from the most intense peaks have the same order of magnitude as the least intense correlation peaks, which explains why we had to show the t_1 noise in figure 2.6 in order to see all the correlation peaks.

With the inter-scan delay chosen, the pulse sequence is also sensitive to artifacts caused by the magnetization that has not relaxed at the end of the delay (also called rapid-pulsing artifacts).[102] Our first choice was the use of gradients, which made us choose the pulse sequence "cosygpmfph" with the inclusion of presaturation during the inter-scan delay. Unfortunately, we found out that control of the faraway water was difficult with this sequence and some correlation peaks from carnitine were not visible. Since we wanted to see all the correlation peaks, we tested another pulse sequence.

Our second choice was the pulse sequence "cosydfphpr", which gave spectra with a better control of the faraway water peak. But since no gradients were in the pulse sequence, the entire phase cycling was needed to suppress the rapid pulsing artifacts, despite not being limited by sensitivity. We had then to compromise between the need to suppress the artifacts efficiently and suppressing the water residual peak. Several approaches were tested, including the use of WATERGATE, bipolar gradients during t_1,[259] and changes in the cycle phasing. The only change left was the phase cycling. Strangely, adding a phase cycle for the presaturation peak helped suppressing the artifacts, while presaturation is not expected to have a role in the formation of artifacts. This phase cycle was already used in a Bruker pulse sequence called "cosydfgpph19".

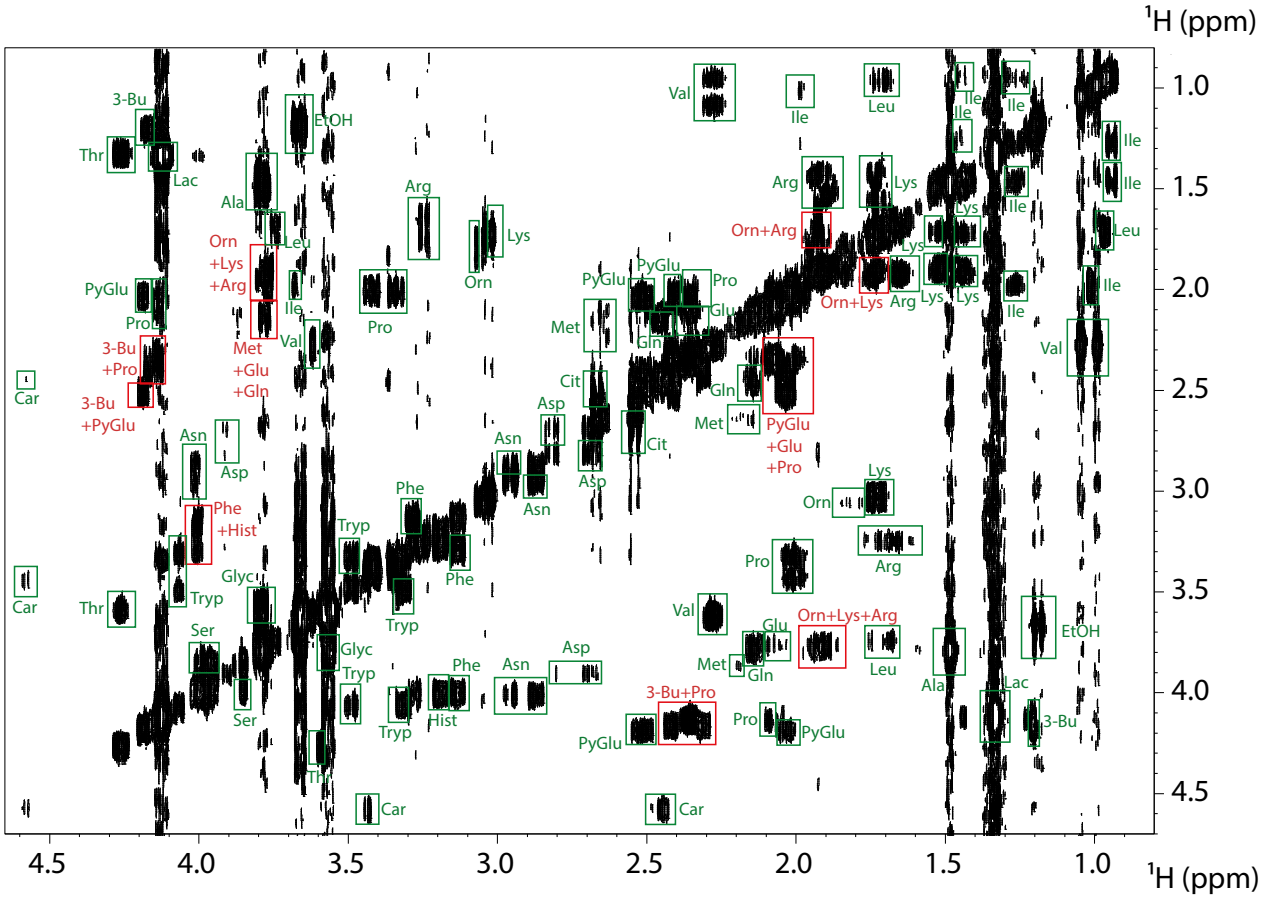


Figure 2.6. Homonuclear DQF-COSY spectrum of one of the synthetic serum with a zoom on the aliphatic portion of the spectrum. The buckets defined for the statistical analysis are shown, with the following legend: Ile: isoleucine, Leu: leucine, Val: valine, EtOH: ethanol, 3-Bu: 3-hydroxybutyrate, Lac: lactate, Ala: alanine, Lys: lysine, Orn: ornithine, Arg: arginine, Pro: proline, Glu: glutamate, Gln: Glutamine, Pyglu: pyroglutamate, Met: methionine, Cit: citrate, Asp: aspartate, Asn: asparagine, Car: carnitine, Phe: phenylalanine, Hist: histidine, Tryp: tryptophane, Glyc: glycerol, Ser: serine, Thr: threonine.

3.2.3. Statistical analysis

In the resulting spectrum, most of the correlation peaks are now well resolved. Most of the remaining overlap arises from metabolites with similar structures, mainly from arginine-lysine-ornithine and pyroglutamate-proline. After identification of all the peaks, the peaks volumes were extracted by manual bucketing in the integration tool of Topspin, with the definition of

117 buckets. Then the data were normalized relatively to the total volume and exported to SIMCA. Data were scaled to the UV and analyzed with OPLS-DA. The score and loading plots are presented in figure 2.7.

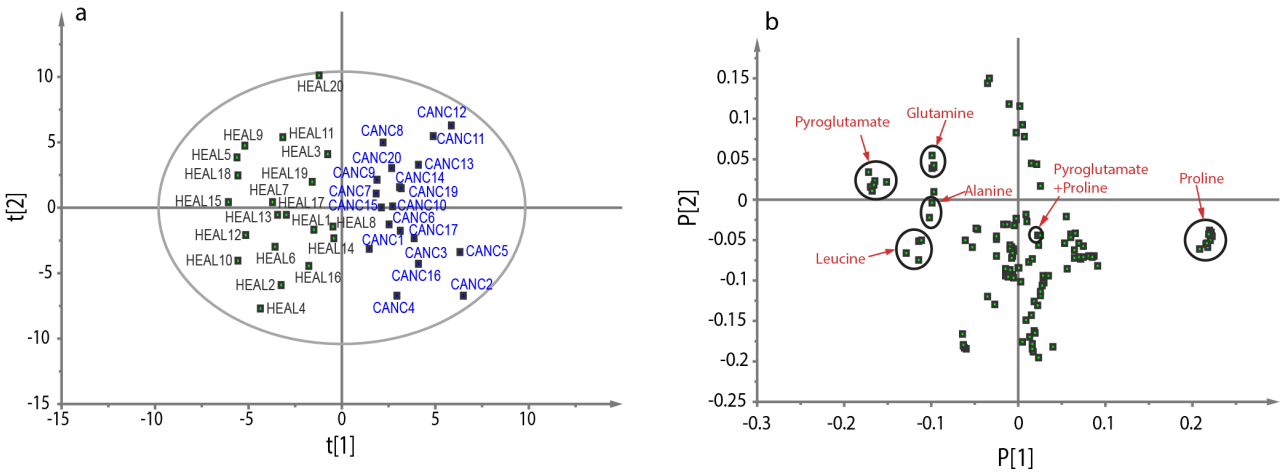


Figure 2.7. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the volume of peaks from the DQF-COSY spectra of 39 synthetic serum samples. In (b), the buckets circled in black contain volumes of peaks from the biomarkers defined in chapter 1.2 plus pyroglutamate (see text in chapter 2.1). The first dimension explains 12.3% of the total variance while the second dimension explains 13.9% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $4.13 \cdot 10^{-6}$.

Once again, the two groups are well separated in the score plot, but the real differences between DQF-COSY and the two other pulse sequences are in the loading plot. In this plot, we can see that the separation of the two groups is mainly caused by 5 metabolites: proline, glutamine, alanine, leucine and pyroglutamate. Acetate has no weight in the separation since it has no correlation peak in the DQF-COSY spectrum. This is one of the drawbacks of COSY: the lack of correlation peaks for metabolites without scalar coupling. Other examples in the mixture include pyruvate, glycine and methanol.

For the remaining biomarkers, nearly all the buckets containing volume from a given metabolite are close in the plot. Since nearly all the buckets have little to no overlapping volume, the FD/SD of the bucket is close to the FD/SD of the metabolite concentration. Therefore, the loading plot for the DQF-COSY is closer to the loading plot of the concentration than the other two pulse sequences. As stated earlier, glutamine is one of the biomarkers for the analysis with DQF-COSY. From this loading plot, it seems that while glutamine is still important for group separation, its FD/SD is lower than for its initial concentration. It also proves that the impossibility of 1D ^{1}H NMR and homonuclear J-RES to identify glutamine as a biomarker is caused indeed by the important overlap of all glutamine peaks.

From this analysis, we showed that the good separation of peaks impacts the FD/SD, similarly to sensitivity (to avoid the instrument noise, which increases the standard deviation) or DoE (to reduce standard deviation from other biological parameters). One more example of the loss of information is the only overlap between two biomarkers, proline and pyroglutamate, whose

corresponding bucket have no impact in the group separation. Now that the benefits of well-resolved peaks in statistical analysis are established, the HSQC can be tested to see its impact on statistical analysis.

3.3. HSQC pulse sequence

As for COSY, HSQC is one of the most used 2D pulse sequences, for small molecules and macromolecules. Thus many variants have been created for its different uses. Similar to the DQF-COSY, one of the most basic variants of HSQC has been chosen for this study, "hsqcgpph", with the addition of presaturation during the inter-scan delay. While it is not the most sensitive variant of the HSQC pulse sequence, it allows a better control of the residual water peak than most other variants.

3.3.1. Parameters for acquisition and processing

During acquisition, 4096 points were recorded in the direct dimension during 0.31 s, with a spectral width of 11 ppm. 16 dummy scans were executed to obtain a steady state magnetization, then 8 scans were acquired for 128 t_1 increments. The spectral width for the ^{13}C dimension was 165 ppm. The inter-scan delay was 3 s, which was also used for presaturation. The delays during the INEPT transfer were 1.72 ms, which corresponds to an average $^1J_{C\text{-}H}$ of 145 Hz. $J_{C\text{-}H}$ decoupling during acquisition was done with the Globally optimized Alternating phase Rectangular Pulse (GARP) supercycle. The total experimental time was 57 min and 37 s.

After acquisition the FID were weighted with a $\pi/2$ shifted squared sine-bell in each dimension, then the number of t_1 increments was doubled by linear prediction. Then the FID was zero-filled to 1024 points in the indirect dimension. After FT, the chemical shifts were calibrated to the TSP, a baseline correction of order 2 was applied and the resulting spectrum was phase corrected. The spectrum of a synthetic serum after this procedure is shown in figure 2.8.

3.3.2. Results and discussion

As mentioned earlier, many variants of the HSQC pulse sequences exist and we tested several of them, for example the inclusion of flipbacks[260], WATERGATE or bipolar gradients during t_1 and the INEPT delays[259]. It turned out that one of the simplest variants was the most effective for the control of the residual water peak with this spectrometer. This pulse sequence also proved to be repeatable, probably because of its high symmetry.[181]

Compared to the DQF-COSY, the sensitivity of the HSQC spectrum is lower, mainly because of the low natural abundance of ^{13}C, but t_1 noise is almost absent and even protons without $^nJ_{H\text{-}H}$ coupling have a correlation peak in HSQC. Therefore, all the metabolites defined in table I have at least one correlation peak in the HSQC spectrum.

Similar to the DQF-COSY, an experimental time higher than 1 h proved to be detrimental to the control of the residual water peak, thus the experimental time was fixed to only 1 h, by reducing the number of t_1 increments. However, as far as we know, this limitation was only present

with this cryogenic probe. Despite the low resolution in the ^{13}C dimension, very little overlap is present in the final spectrum and every metabolite have at least one well-resolved peak.

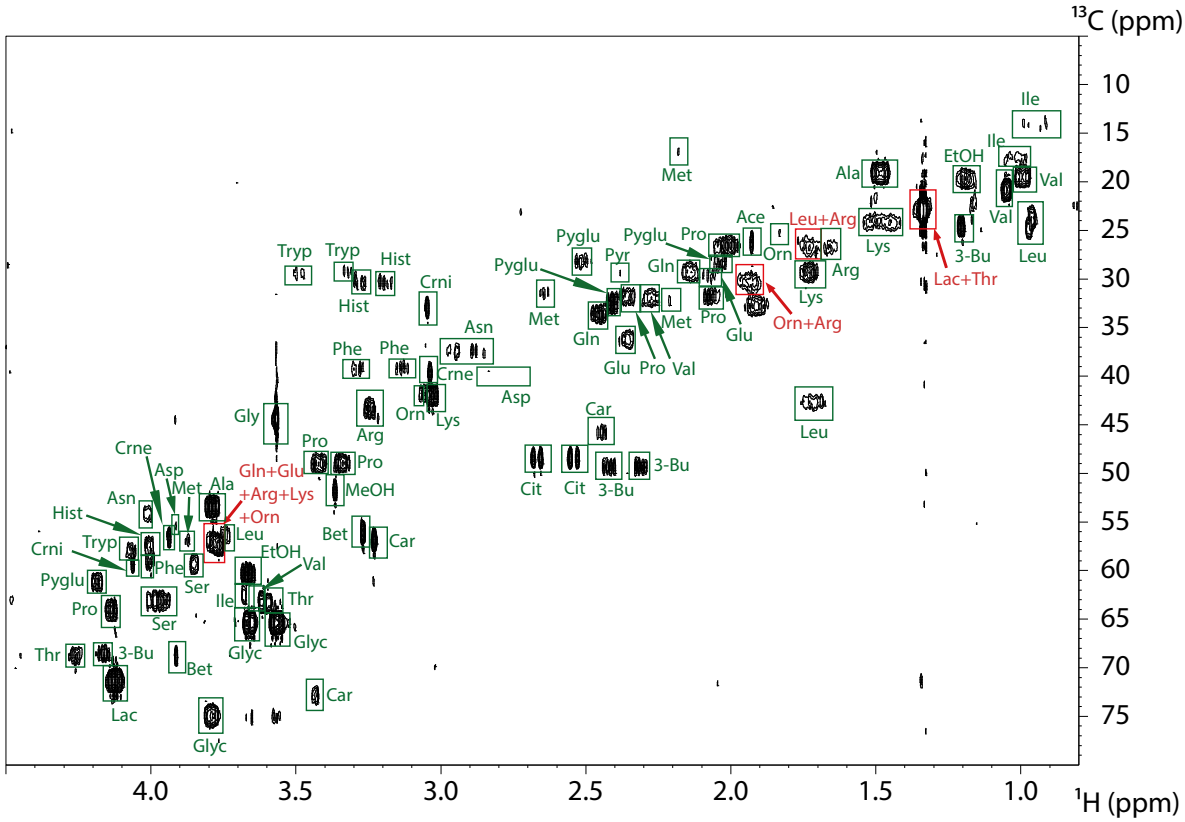


Figure 2.8. ^{1}H-^{13}C HSQC spectrum of one of the synthetic serum with a zoom on the aliphatic portion of the spectrum. The buckets defined for the statistical analysis is shown, with the same legend as for figure 2.6, with the following additions: Ace: acetate, Pyr: pyruvate, Cnri: creatinine, Crne: creatine, MeOH, methanol, Gly, glycine, Bet: betaine.

3.3.3. Statistical analysis

For the statistical analysis, the peaks volumes were extracted by manual bucketing in the integration tool of Topspin, where 96 buckets were defined with the same parameters as stated for DQF-COSY spectra analysis. Then the data were normalized relatively to the total volume and exported to SIMCA. Data were scaled to the UV and analyzed with OPLS-DA. The score plot and loading plot are presented in figure 2.9.

As for all the previously described pulse sequences, the HSQC gives a good separation between the two groups in the score plot. For the loading plot, this time 6 metabolites are mainly responsible for the separation: proline, acetate, glutamine, leucine, alanine and pyroglutamate. These results mean that with HSQC, all the potentials biomarkers are identified in the mixture. The OPLS-DA of HSQC is also the most similar to the OPLS-DA of the concentration. This makes the HSQC spectrum the most faithful representation of the concentration. Similarly to the DQF-COSY, the buckets from one biomarker tend to have the same coefficients in the loading plot, which can make their identification easier. Only two buckets are exceptions. The first bucket contains a peak

from glutamine heavily overlapping with peaks from glutamate, arginine, lysine and ornithine, the other contains a peak from leucine with a rather low S/N and slightly overlaps with the peak from the first bucket.

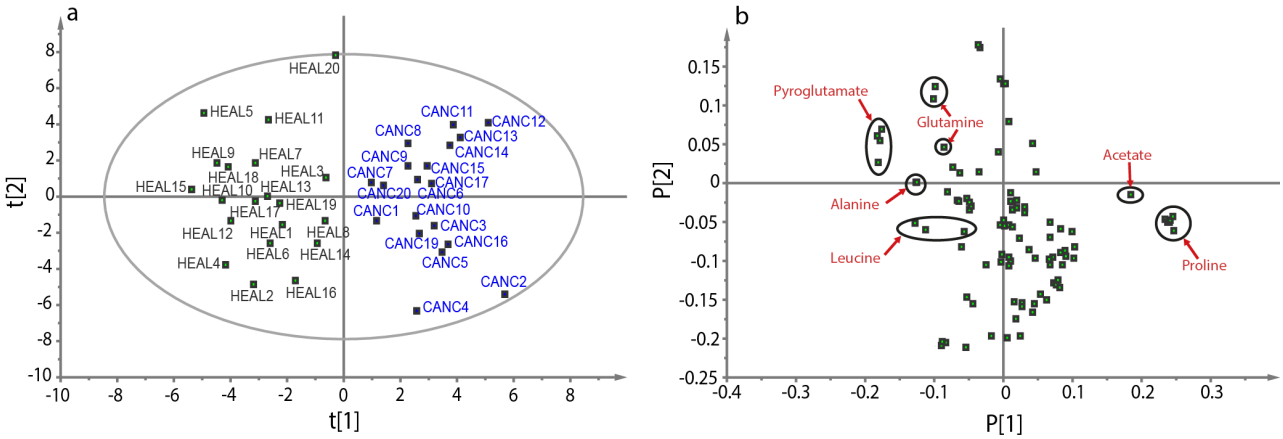


Figure 2.9. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the volume of peaks from the HSQC spectra of 39 synthetic serum samples. In (b), the buckets circled in black contain volumes of peaks from the biomarkers defined in chapter 1.2 plus pyroglutamate (see text in chapter 2.1). The first dimension explains 11.2% of the total variance while the second dimension explains 9.7% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $1.86 \cdot 10^{-5}$.

These results confirm that both sensitivity and resolution are important in order to obtain as much information as possible about the biological changes. For our samples, HSQC gave the best results for biomarker identification, but the low isotopic abundance of ^{13}C restricts its use for samples with high concentrations[140, 261] or isotopic labeling[30, 151]. In our case, with our settings, the limit of detection seemed to be the low hundreds of μM at natural abundance.

In summary, using synthetic samples of known composition and concentration, we showed that 2D NMR is a useful alternative to 1D NMR for the identification of biomarkers. Thanks to the added resolution, more potential biomarkers can be identified in statistical analyses, showing that resolution is also an important parameter in metabolomics studies by NMR. These results are consistent with other results published recently.[5]

Despite this advantage, 2D NMR is characterized by its long acquisition time, limiting his use for metabolomics if a large number of samples have to be analyzed. The next step is to evaluate the use of fast 2D NMR approaches for metabolomics studies to see if we obtain similar results to conventional 2D NMR in a reduced experimental time.

Part 3. Evaluation of fast 2D experiments for metabolomics

Goal: To investigate the performance of fast 2D NMR approaches as an efficient alternative to their conventional counterparts for metabolomics applications. Non-Uniform Sampling and Ultrafast 2D NMR were used for this first study.

1. Calibration of fast 2D experiments for metabolomics analysis

Several approaches exist for the reduction of 2D NMR experimental times, as shown in section 3 of part 1. We decided to focus on three of those approaches. The first two are Non-Uniform Sampling (NUS) and UltraFast (UF) 2D NMR because they have already been used in the case of complex mixture analysis.[33, 34, 38] The last one is the BEST/SOFAST approach, since two recent papers described the use of BEST in the case of small molecule analysis.[262, 263] The other approaches were not tested for several reasons:

- The nature of Hadamard acquisition makes this approach more fitted for targeted analysis[239] than global metabolomics analysis. Since Hadamard acquisition relies on the acquisition of 1D slices, that approach is more powerful if only a fraction of the spectrum needs to be analyzed.

- ASAP was limited to the HMQC at the beginning of the thesis, and IMPACT was limited to the HMBC.

- Folding/aliasing would have increased the risk of overlapping, especially for the homonuclear pulse sequences.

- SMART needs tri-axis gradients for most pulse sequences, which was not available with our probes.

- Linear prediction was stated to be less efficient than NUS to resolve overlapping peaks,[186] so it was not used to reduce the experimental time. Instead, it was used as a tool for an easier bucketing in 2D spectra, similarly to the conventional 2D spectra in part 2.

- Covariance is an algorithm that can be used for conventional or NUS 2D datasets.[192] NUS was chosen for a first study since it was easier to implement for most pulse sequences.

- Radial sampling is only available for nD spectra, with n higher than 2.

These three approaches have been tested with the pulse sequences used in section 3 of part 2. At first, the calibrations for NUS variants of DQF-COSY and HSQC pulse sequences are presented, followed by the Ultrafast variant of COSY. Then the tests for the UF variant of J-resolved spectroscopy are shown and discussed, as well as the early tests for BEST. Finally, the statistical analyses for NUS DQF-COSY, NUS HSQC and UF COSY will be discussed and compared to the statistical analyses of conventional DQF-COSY and conventional HSQC.

1.1. Non-Uniform Sampling (NUS)

NUS is one of the oldest approaches to reduce the experimental time of 2D NMR spectra.[198] Its implementation in vendor-supplied software has become more commonplace, for

both acquisition and processing of NUS data. It is also an approach available for potentially all 2D experiments and rather easily customizable, since the only new parameters to optimize for acquisition are the percentage of NUS, *i.e.* the fraction of t_1 increments recorded during acquisition, the distribution of these points and the choice of reconstruction algorithm. These advantages makes NUS a good candidate for metabolomics, where the optimization of NMR parameters should be as quick as possible.

For this study we chose to use the tools within Topspin for both acquisition and processing. Two pulse sequences have been tested: DQF-COSY and HSQC. J-RES was not included because it did not meet the sparsity condition which makes NUS more efficient.[32] Since the hardware of the spectrometer used in part 2 did not allow the automated acquisition of NUS spectra with the vendor-supplied software, all NUS experiments were acquired with another Bruker Avance I 600 MHz equipped with a 5 mm TCI cryoprobe triple resonance ($^{1}H/^{13}C/^{15}N$).

1.1.1. NUS DQF-COSY

1.1.1.1. Parameters for acquisition and processing

For the NUS DQF-COSY, the spectra were acquired in a similar manner to the conventional DQF-COSY in part 2. In short, the pulse sequence was the “cosydfphpr” with the phase cycling of “cosydfgpph19” for the presaturation pulse. 256 t_1 increments were acquired with 8 scans and an inter-scan delay of 3 s. With NUS, only 30% of t_1 increments were acquired, reducing the experimental time to 34 min and 36 s. The exponential sampling schedule was automatically generated by Topspin at the beginning of the acquisition, with the parameter "T1" set to 1 s.

The processing of NUS DQF-COSY was also similar to conventional DQF-COSY, with the inclusion of Compressed Sensing (CS) to reconstruct the missing points in the indirect FID. After CS, the direct and the indirect FID were weighted with a sine-bell function. 256 t_1 increments were added by linear prediction, to reduce the volume of peaks for an easier bucketing, similarly to the conventional counterpart. Finally the FID was zero-filled to 8192 points in the direct dimension and 1024 in the indirect dimension before FT. The chemical shifts were referenced to the TSP and a baseline correction with a polynomial function of order 2 was applied. The processed spectrum of one of the synthetic samples of serum is shown in figure 3.1 with a comparison to the conventional spectrum.

1.1.1.2. Results and discussion

Visually, the NUS DQF-COSY is characterized by a reduced level of t_1 noise. Thanks to the reduced experimental time, less spectrometer instabilities were experienced, which allowed the reduction of t_1 noise with the preservation of peak intensity, as shown in the case of conventional 2D NMR.[241] Otherwise, the NUS DQF-COSY spectrum is very close to the conventional DQF-COSY spectrum.

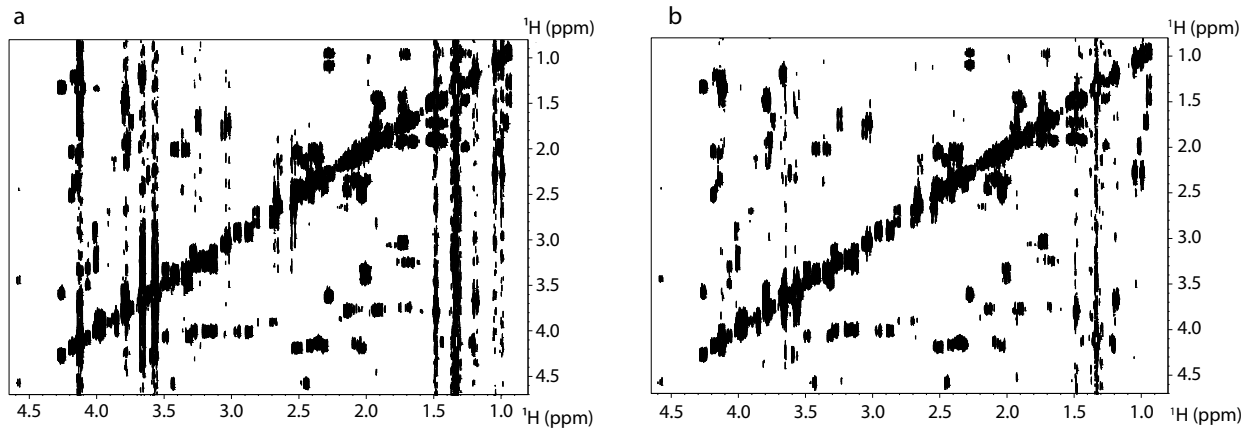


Figure 3.1. Zoom on the aliphatic part of the conventional DQF-COSY spectrum of one of the synthetic serum samples (a), same as figure 2.6, acquired in around 2 h, with its NUS counterpart, acquired with 30% of the t_1 increments (b), acquired in around 35 min. The two spectra are shown with the contour level that allows the visualization of the correlation peak with the lowest S/N.

One advantage of DQF-COSY is that with our samples and parameters, all correlation peaks had a high S/N, which allowed us to reduce more easily the number of t_1 increments recorded during acquisition. Our earlier tests showed that below 30% of NUS, the volume of the peaks with the lowest S/N is not reconstructed correctly anymore. For this reason, the experimental time of the DQF-COSY was not reduced further. Despite this limitation, we managed to reduce the experimental time to less than one hour, which should be short enough for many metabolomics studies. We next used NUS with the HSQC pulse sequence to see if we could reduce the experimental time of HSQC to a level close to NUS DQF-COSY.

1.1.2. NUS HSQC

1.1.2.1. Parameters for acquisition and processing

As with NUS DQF-COSY, the parameters of NUS HSQC were kept similar to those of conventional HSQC. Specifically, the pulse sequence “hsqcgpph” was chosen, with the addition of presaturation during the inter-scan delay. 128 t_1 increments were recorded with 8 scans and an inter-scan delay of 3 s. With NUS, only half of the t_1 increments were acquired with an exponential sampling schedule generated by Topspin at the beginning of the acquisition. The experimental time was 29 min and 15 s.

The processing of NUS HSQC spectra was also kept similar to conventional HSQC. At first, CS was used to reconstruct the missing points in the indirect FID. After CS, the direct and indirect FIDs were weighted with a $\pi/2$ shifted squared sine-bell function. The number of t_1 increments was doubled by linear prediction, to reduce the width of peaks for a better bucketing, similarly to the conventional counterpart. Then the indirect FID was zero-filled to 1024 points before FT. The chemical shifts were referenced to the TSP and a baseline correction with a polynomial function of order 2 was applied. The processed spectrum of one of the synthetic serums is shown in figure 3.2 with a comparison with the conventional spectrum.

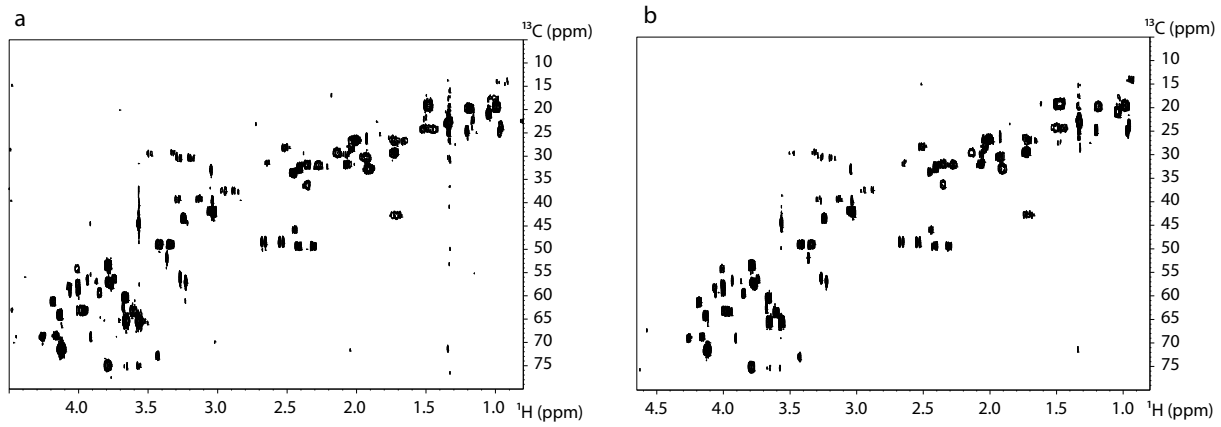


Figure 3.2. Zoom on the aliphatic part of the conventional HSQC spectrum of one of the synthetic serums (a), same as figure 2.6, acquired in around 1 h, with its NUS counterpart, acquired with 50% of the t_1 increments (b), acquired in around 30 min. The two spectra are shown with the contour level close to the noise level.

1.1.2.2. Results and discussion

Once again, the NUS HSQC spectrum is very similar to the conventional HSQC spectrum. All the correlation peaks are still visible, even those close to the noise level. The NUS HSQC spectrum also seems to have a reduced t_1 noise, but since the conventional HSQC spectrum already has little t_1 noise, the expected reduction of t_1 noise with NUS is less obvious than in DQF-COSY. As for the DQF-COSY, lower percentages of NUS were tested, but below 50%, the peaks close to the noise level start to disappear. Therefore, the percentage of NUS was limited to 50%. A more detailed study of sensitivity with the percentage of NUS will be presented in part 5.

Thanks to NUS we managed to obtain DQF-COSY and HSQC spectra with experimental times close to 30 min. We can now see the effects of ultrafast 2D NMR acquisition with these samples.

1.2. UltraFast (UF) 2D NMR

With UF 2D NMR, the acquisition of a 2D spectrum can last a fraction of a second. But with the sensitivity losses inherent to UF 2D NMR, as described in chapter 3.3.2 of the literature review,[230, 264] a single scan was not sufficient with the concentrations in our samples in order to have at least one peak for each metabolite. For this reason, the Multi-Scan Single-Shot (M3S) approach was used to increase sensitivity while keeping reasonable experimental times.[38, 241]

Among the possible UF experiments, UF COSY, UF J-RES and UF HSQC were considered. But early tests have shown that because of the compromise between sensitivity, resolution and spectral width in UF 2D NMR, UF HSQC was not suited for diluted samples. It turned out that even at similar experimental times, it is more practical to acquire a conventional HSQC rather than an ultrafast HSQC, mainly because the spectral width in ^{13}C was too small for be relevant for metabolomics. Thus, UF HSQC was not used in the study. However, the problem of

spectral width could be solved by interleaving,[215] an approach currently evaluated by the EBSI group.[265]

For all ultrafast experiments, the acquisition was made with Bruker Avance III 500 MHz spectrometer equipped with a 5 mm $^{1}H/^{13}C$ dual cryoprobe. This change was motivated by two reasons. The first was a slightly more powerful and stable gradient used in the cryoprobe of the 500 MHz, both parameters being critical in UF experiments. The second was the lower magnetic field, which allowed us to encode a larger region (in ppm) for the UF spectrum. These two factors allowed us to acquire the UF spectrum of the whole aliphatic region of the ^{1}H without compromising between sensitivity and resolution. The aromatic region was not explored because of the lack of sensitivity and the fact that this region is already well separated in the 1D spectrum.

1.2.1. UF COSY

1.2.1.1. Parameters for acquisition and processing

The COSY pulse sequence in UF was preferred to UF DQF-COSY because the choice of spatial encoding already allowed the suppression of the phase difference between diagonal and correlation peaks. Indeed, with this spatial encoding, this pulse sequence is the UF equivalent of constant-time COSY.[241] Like constant-time COSY, diagonal and correlation peaks have the same phase, which makes one of the main advantages of DQF-COSY redundant in the UF variant. The main difference between UF COSY and UF DQF-COSY is a loss of around $\sqrt{2}$ of sensitivity, because of coherence selection. This made UF DQF-COSY impractical for our study.

The M3S UF COSY was developed during my Master internship for quantification in complex mixtures[38], so the parameters were similar to that publication. For the acquisition, the pulse sequence was depicted in figure 1.21. The spatial encoding was performed by a pair of 180° smooth chirp pulses, each one lasting 15 ms with a sweep range of 60 kHz, with the simultaneous application of a 5.33 G/cm bipolar gradient pair (8.3% of the maximum gradient strength available). During the acquisition, 128 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 512 μs and had an amplitude of 52 G/cm (80% of the maximum gradient strength available). 1024 points were acquired for each gradient pair. The number of scans was 176 and the inter-scan delay 5 s, for an experimental time of 14 min and 57 s. This experimental time was optimized to obtain at least one correlation peak for all the metabolites in the mixture.

The processing was done using a home-written routine from the EBSI group,[37] used within the commercial software. Briefly, Inverse Fourier Tranformation (IFT) of the ultrafast dimension was performed, then an optimized Gaussian apodization was applied to eliminate non-idealities at the border of the detection coil and the noise from outside the coil.[232] A FT was applied afterwards to return to the ultrafast dimension. The conventional dimension was weighted with a $\pi/4$ shifted sine-bell function, and then zero-filled to 1024 before FT in this dimension. After FT, the chemical shifts were referenced to the two diagonal peaks from lactate (1.32 and 4.10 ppm) and a baseline correction of order 2 was applied. The processed spectrum of one of the synthetic serum is shown in figure 3.3.

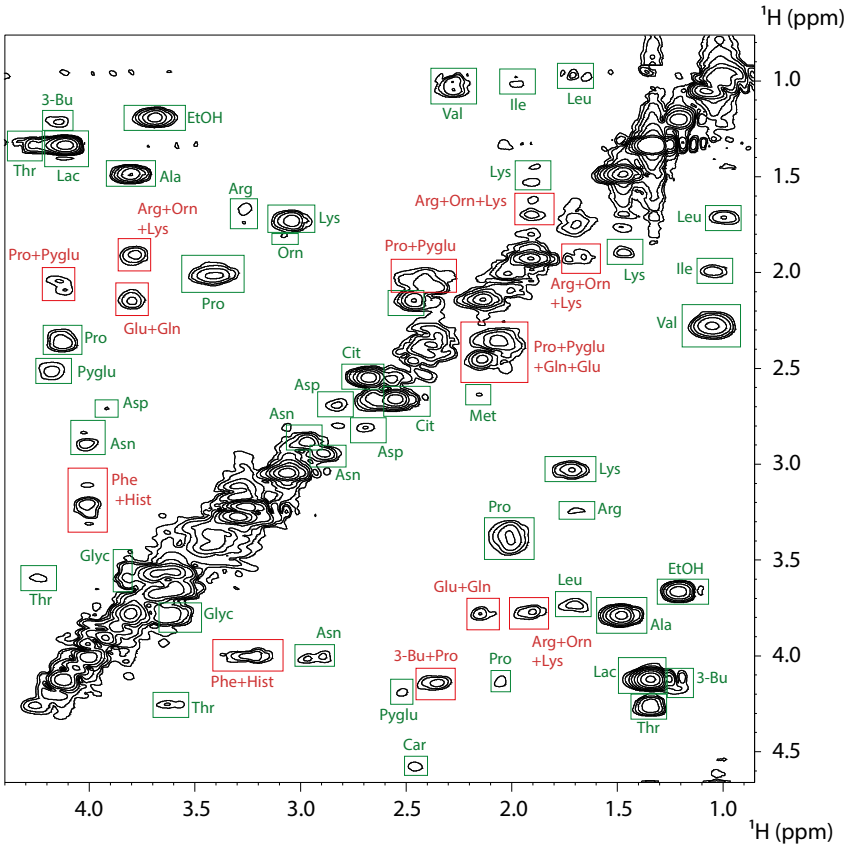


Figure 3.3. Spectrum of the aliphatic part of a synthetic serum acquired with the UF COSY pulse sequence, with an experimental time of around 15 min, with the attribution of the visible correlation peaks. The ultrafast dimension is horizontal while the conventional dimension is vertical. The abbreviations are the same as figure 2.6.

1.2.1.2. Results and discussion

On the spectrum, almost all the correlation peaks are visible, except for some correlation peaks from the most diluted metabolites, like aspartate. The low sensitivity for these peaks is probably caused by the low concentration along with J-modulation. While the resolution of the peaks in the indirect dimension is limited, it allows the separation of most correlation peaks. The remaining overlapping peaks are, for the most part, the same as in conventional DQF-COSY.

The main differences are correlation peaks near the diagonal, mainly from non-equivalent protons in -CH_2- groups. In conventional DQF-COSY, they are, for the most part, well resolved or slightly overlapping with diagonal peaks. In UF COSY, because of the lower resolution, most of these correlation peaks are overlapping heavily with diagonal peaks. This problem will be addressed in part 4.

This analysis shows that, while sensitivity is still limited in UF COSY, it is possible, in 15 min and with a 500 MHz, to see the correlation peaks of species at submillimolar concentrations, which was not obvious at the beginning of the study. This makes the statistical analysis of UF COSY spectra possible, with at least one well-resolved peak for each metabolite having a correlation peak in COSY.

1.2.2. UF J-RES

1.2.2.1. Parameters for acquisition and processing

The UF J-RES pulse sequence has a different way to acquire the data than most UF pulse sequences.[266] Compared to the original way to acquire the UF J-RES spectrum,[218] this allows the chemical shifts and the J-couplings to be already separated during the acquisition. Tilting the resulting spectrum is thus not needed. Compared to UF COSY, UF J-RES has several advantages. Since the spectral width in the conventional dimension is small, long gradient durations can be applied during the acquisition, to obtain a high spectral width in the ultrafast dimension. Since the gradient duration is high, the amplitude of the gradients during acquisition can be lowered, which is beneficial for sensitivity because the receiver bandwidth can be reduced, lowering the noise level in the process. The reduced amplitude during acquisition can also be used to improve the resolution in the indirect dimension by increasing the duration of spatial encoding. In order to limit the diffusion, a multi-echo excitation scheme can be used.[220, 230]

For these reasons, the UF J-RES was tested for metabolomics. The pulse sequence used has been shown in figure 1.19. Spatial encoding was performed with two pairs of smooth chirp pulses, each one lasting 15 ms, with a bandwidth of 60 kHz, with the simultaneous application of two bipolar gradients with an amplitude of 5.33 G/cm (8.2% of the maximum available strength). The acquisition was done with 128 gradients to refocus the dephasing caused by the spatial encoding. Each gradient lasted 6.6 ms with an amplitude of 5.2 G/cm (8% of the maximum available strength). 4096 points were acquired for each gradient. Two successive acquisition gradients were separated with a hard 180° pulse. Within each scan, successive 180° pulses were phase-cycled (y; y; -y; -y) to avoid spurious simulated echoes. The number of scans was 96 and the inter-scan delay 5 s, for an experimental time of 9 min and 36 s.

The processing was done with the same home-written routine as UF COSY. The main difference was the apodization of the conventional dimension with a $\pi/4$ shifted squared sine-bell function. After FT, the chemical shifts were referenced to the two peaks from lactate and a baseline correction of order 2 was applied. The processed spectrum of one of the synthetic serum is shown in figure 3.4.

1.2.2.2. Results and discussion

Compared to UF COSY, the peak lineshapes are narrower, with a linewidth at half-height of around 45 Hz for UF COSY and 25 Hz for UF J-RES. Despite the narrower linewidths, the UF J-RES spectrum is characterized by more overlap of peaks than UF COSY. This is caused by the small spectral width for the J-coupling dimension, which does not allow a good dispersion of peaks, similarly to the conventional J-RES spectra in part 2.

Another problem with UF J-RES is with the reconstruction of some multiplets. To give one example, at 2 ppm, there is a peak that looks like a doublet, while in conventional J-RES, in the same region, several multiplets are present. With the help of TOCSY and HSQC, we could confirm that there should be several multiplets in that region. The reasons for this problem are currently unknown, but since the peaks cited above have strong coupling with other close peaks, one possibility is non-idealities during spatial encoding or acquisition caused by strong coupling.

Another possibility is J-modulation during acquisition. This behavior is currently under study. Because of these problems, UF J-RES was not kept for metabolomics studies and only UF COSY was used for the evaluation of the UF 2D NMR in the case of metabolomics.

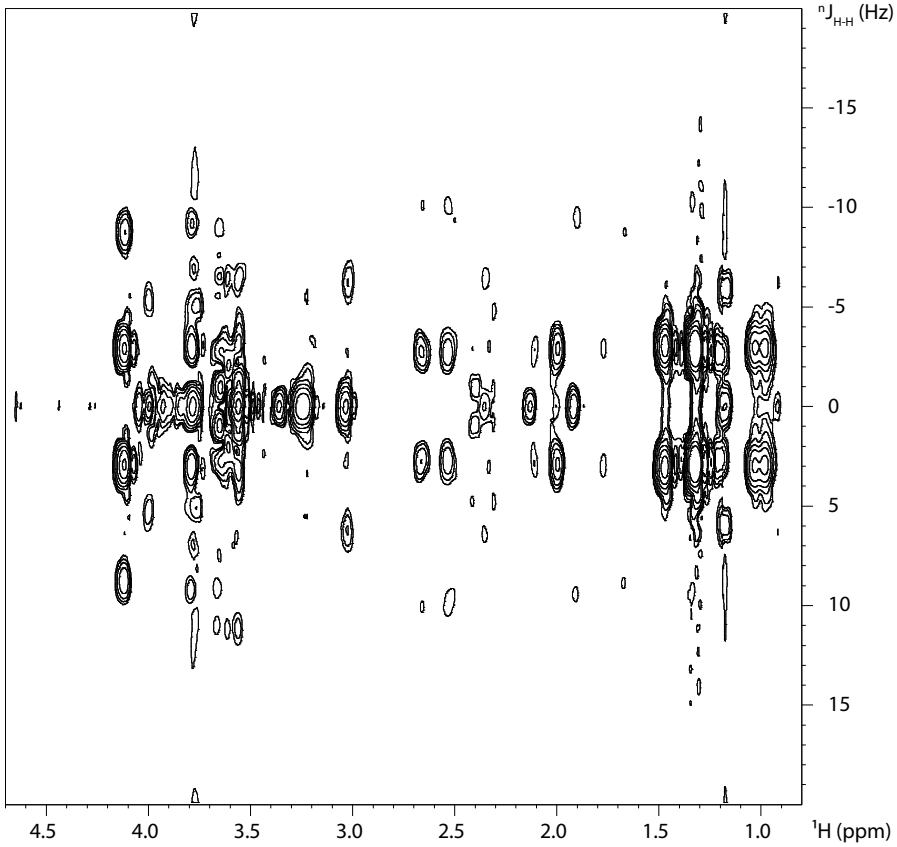


Figure 3.4. Spectrum of the aliphatic part of a synthetic serum acquired with the UF J-RES pulse sequence, with an experimental time of around 10 min. The ultrafast dimension (chemical shifts) is horizontal while the conventional dimension (J-couplings) is vertical.

1.3. SOFAST/BEST: tests with inversion-recovery

Since the SO-FAST/BEST strategy makes it possible to reduce the inter-scan delay by reducing the T_1 of the nuclei observed in the resulting spectrum, it has been decided to study the effectiveness of the strategy for small molecules by directly measuring T_1 with and without selective pulses. As stated before, the SOFAST/BEST approach has been lately used for small molecule analysis.[262, 263] But, as stated in a recent review[166] and explained in section 3.1.1 of part 1, reduction of the T_1 relaxation time by this approach occurs only when cross-relaxation is negative, which happens when the correlation time is sufficiently high. With small molecules in non-viscous solvent like water, the correlation time is not supposed to be high enough to allow a relaxation enhancement via the SOFAST/BEST approach. For the study from Motta and al., the whole cells are studied, with the metabolites still present in the cells.[262] Since cytoplasm is more viscous than deionized water, it is possible for metabolites in cytoplasm to have negative cross-relaxation.

But for the study of Shukla and Dorai,[263] the small molecules were solubilized in methanol or water, with methanol being even less viscous than water. In these conditions, no data about T_1

and T_2 with the use of BEST was indicated. For this reason, we decided to measure the T_1 of a small molecule in deuterated water, to verify if the use of BEST in these conditions allows the reduction of T_1.

1.3.1. Evaluation of the inversion-recuperation with selective pulses for T_1 calculation

This study was conducted on an Avance 600 MHz Bruker spectrometer with a 5 mm TXI cryoprobe that was already used for the NMR experiments in part 3. The idea was to measure directly the change of T_1 with the inversion-recovery (IR) pulse sequence, similarly to the pulse sequence shown in figure 1.4 in the literature review. A variant of this pulse sequence was developed, were all the hard pulses were replaced with selective pulses. The first part of this study was to verify that the T_1 measured with the pulse sequence was not altered by the introduction of the selective pulse. This part was conducted on a sample of serine in D_2O. Since the peaks from serine are within a spectral width of less than 0.2 ppm (120 Hz), it allowed the use of long selective pulses without seeing the effect of BEST on the T_1. Therefore, it was possible to compare the IR with hard or selective pulses without biases.

For the IR with hard pulses, 12288 points were recorded during 1 s of acquisition, with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 30 or 40 s, for reasons explained later. 12 different delays between the 90° and the 180° pulse were acquired, in seconds: 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 4, 8, 15, 30 and 40. Later, the delays 45 seconds and 50 seconds were added, which will be also explained later. The IR with selective pulses was recorded with the same parameters, with the use of a 13 ms Reburp 90° pulse and of a 14 ms 180° Eburp pulse. The results of the T_1 estimation for the two proton groups visible on the spectrum with different parameters are shown in figure 3.5. The T_1 for each peak was estimated with the T_1/T_2 Topspin module.

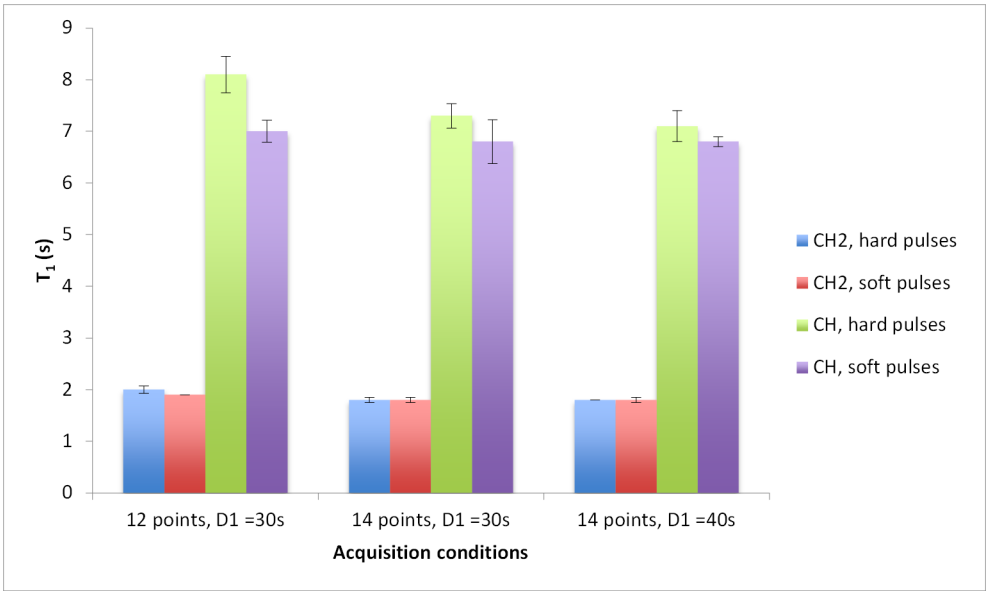


Figure 3.5. Evolution of the T_1 estimation from the protons of serine from inversion-recovery pulse sequence, depending on the number of delays between the 90° and the 180° pulse tested and the inter-scan delay (D1). The T_1 shown are an average of 2 experiments for "12 points and D1=30s", 4 experiments for "14 points and D1=30s" and 3 experiments for "14 points and D1=40s".

From the first IR spectra, it was shown that the T_1 of the proton of the -CH- group was unusually high, around 7-8 seconds. As seen in figure 3.5, the first T_1 analysis showed a difference between the T_1 from this proton between the IR with hard or selective pulses. For this reason, the number of delays between the 90° and the 180° pulses was increased, as well as the inter-scan delay.

With these changes, the estimated T_1 of the proton from the -CH-group for IR with hard or selective pulses are not significantly different. This suggests that, as long as the evolution of the signal is well represented by the choice of delays between the 90° and the 180° pulse, the T_1 will be estimated the same way, with hard or selective pulses.

Since it has been shown that IR with selective pulses gives similar results compared to IR with hard pulses, this tool can be used to verify if the use of BEST can effectively decrease the T_1 of small molecules.

1.3.2. Evaluation of BEST for the reduction of the T_1 of small molecules in water

For this work, the molecule tested had to show peaks in distinctly different regions. Our choice was phenylalanine, since the protons from the -CH_2- and the -CH- groups have more than 450 Hz (around 0.8 ppm) of difference and the aromatic peaks are separated by at least 4 ppm from the other peaks. With these differences in chemical shifts, it is possible to selectively choose each of the groups to see the effect of BEST on each group. In order to see the effect, four experiments were conducted. The first was the IR with hard pulses, where 12288 points were recorded during 1 s of acquisition, with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 30s and the 14 delays between the 90° and the 180° pulse used in the last chapter were acquired.

The three other experiments were IR with soft pulses, each selecting only one of the three groups of protons in phenylalanine, and each having the same parameters as the IR with hard pulses, except for the selective pulses, which were a Reburp pulse for the 180° pulse and a Gaussian pulse for the 90° pulse. The first experiment used a 180° pulse of 13 ms, a 90° pulse of 9 ms and was centered at 7.28 ppm for the aromatic peaks. The second experiment used a 180° pulse of 23 ms, a 90° pulse of 9 ms and was centered at 3.89 ppm for the peaks from the -CH- group. The third experiment used a 180° pulse of 23 ms, a 90° pulse of 5 ms and was centered at 3.11 ppm for the peaks from the -CH_2- group.

For each experiment, the T_1 was also estimated from the T_1/T_2 Topspin toolbox. The 1D ^{1}H spectrum is shown in figure 3.6.a, while the results for T_1 are summarized in figure 3.6.b. From this figure, we can see that for most peaks, the change from hard to soft pulses did not significantly influence the T_1. But for the peaks at 3.9 and 7.24 ppm, the use of selective pulses increased the T_1 significantly. Thus, we can see that the use of BEST in the case of small molecules in water is detrimental to the reduction of the experimental time. This result is probably caused by the fact that phenylalanine has a low correlation time in these conditions, which cause the cross-relaxation to be positive. The T_1 are thus higher instead of lower. From these results, we concluded that the BEST approach was not suited in our study and thus was dropped from the rest of the study.

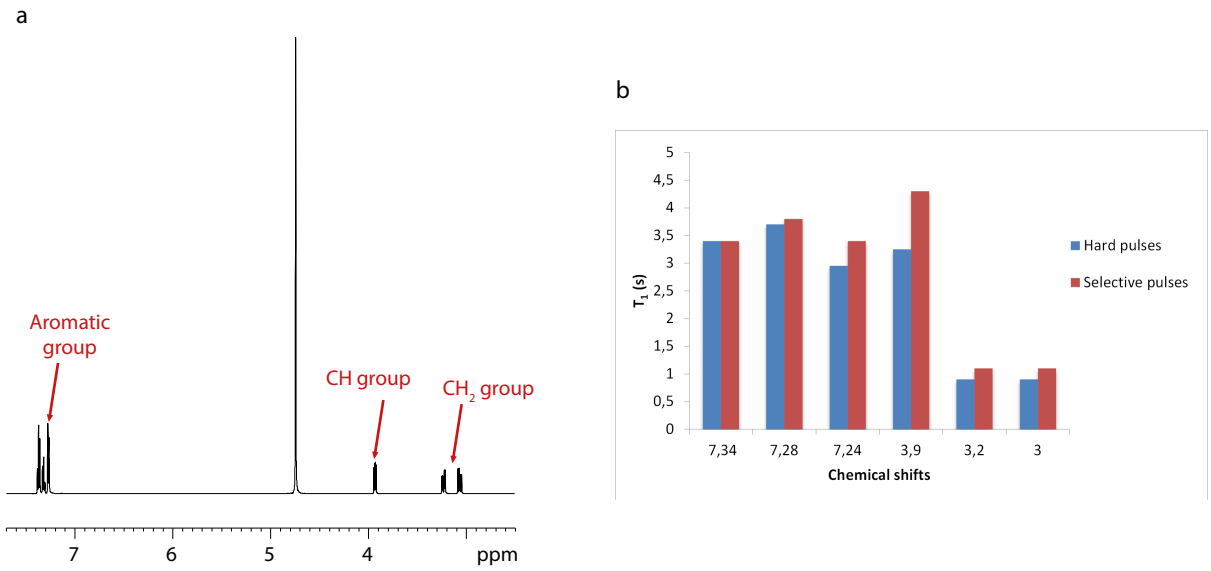


Figure 3.6. 1D spectrum of phenylalanine, with the three different groups of proton from this molecule (a), along with the estimated T_1 from phenylalanine peaks from IR pulse sequence with hard pulses or selective pulses (b).

After these tests, we selected three fast 2D variants of 2D pulse sequences that seemed to work for a global metabolomics study: NUS DQF-COSY, NUS HSQC and UF COSY. With the spectra from each of the 39 samples, statistical analysis could be done for the three pulse sequences to compare their performance to conventional 2D NMR.

2. Statistical analysis and comparison between fast and conventional 2D NMR

With the three fast variants of 2D pulse sequences mentioned above, we can do a statistical analysis and compare it to their conventional counterpart. Similarly to part 2, the samples studied are synthetic serums for a model metabolomics study to separate a control group from a group of patients with a colorectal cancer. Each variant is analyzed separately.

2.1. NUS DQF-COSY

Statistical analysis of the NUS DQF-COSY spectra were conducted the same way as with conventional DQF-COSY. Basically, the bucketing of the correlation peaks was done with the integration tool of Topspin, using the same regions as conventional DQF-COSY (see figure 2.6 to see most of the buckets) and the volumes exported to an Excel file. Then the data were normalized relatively to the total volume and exported to SIMCA. Data were scaled to the UV and analyzed with OPLS-DA. The results are shown in figure 3.7.

We can see from the score plot that the 2 groups (control and diseased) are still well separated. The very low *p*-value calculated from the CV-ANOVA confirms the low chances that the 2 groups are not significantly different. The loading plot showed that the same metabolites as conventional DQF-COSY are mainly responsible for the separation of the 2 groups: pyroglutamate, glutamine, leucine, alanine and proline.

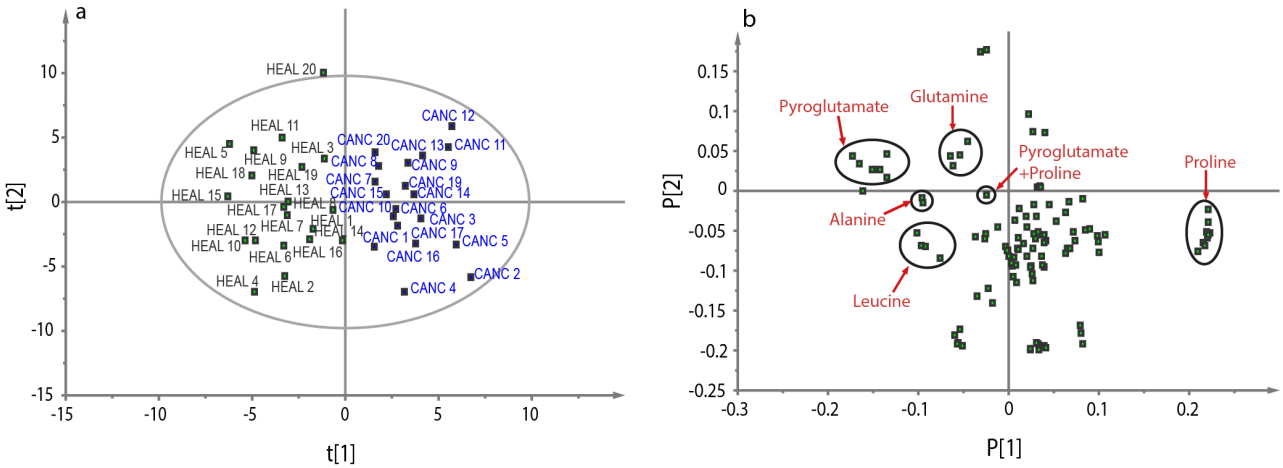


Figure 3.7. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the intensities peaks from the NUS DQF-COSY spectra of 39 synthetic serum samples. In (b), the buckets circled in black contain intensities of peaks from the biomarkers defined in part 2. The first dimension explains 13.3% of the total variance while the second dimension explains 13% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $6.42 \cdot 10^{-7}$.

More interestingly, the coefficients for these biomarkers are similar from the loading plot of the conventional DQF-COSY. For the first dimension, the one important for the group separation, the coefficients are really close to the ones in conventional DQF-COSY, with the exception of glutamine buckets, whose coefficients are slightly lower than for the conventional DQF-COSY. This has probably been caused by further conversion of glutamine into pyroglutamate during the period between the two analyses. Thus, with the NUS variant of DQF-COSY, we manage to obtain a similar analysis compared to conventional DQF-COSY, with only a third of the experimental time needed for conventional DQF-COSY.

2.2. NUS HSQC

Similarly to NUS DQF-COSY, the statistical analysis of NUS HSQC was carried out the same way as for the conventional HSQC. The same buckets as for conventional HSQC were chosen in the Topspin integration tool and extracted in an Excel file. Then the data were normalized relatively to the total volume and exported to SIMCA. Data were scaled to the Unit Variance (UV) and analyzed with OPLS-DA. The results are shown in figure 3.8.

From the score plot, we still have a good separation of the two groups and the p-value from the CV-ANOVA analysis proves that this separation has very little chance to be non-significant. With the loading plot, 6 metabolites are mainly responsible for the separation: proline, acetate, pyroglutamate, glutamine, leucine, and alanine. Once again, the potential biomarkers in NUS HSQC are similar to those from conventional HSQC, and the coefficients in the first dimension are similar between the two statistical analyses.

One interesting note is the behavior of the two buckets highlighted in the statistical analysis of conventional HSQC. The bucket from glutamine with overlapping peaks and the bucket from leucine with low sensitivity and partial overlap are now in other positions in the statistical analysis of NUS HSQC. In contrast, for the other buckets of other biomarkers, the positions are similar in the two analyses. This shows that the position of a bucket with overlapping peaks can become unreliable. The same problem was also seen in the bucket with overlapping peaks from proline and pyroglutamate in DQF-COSY.

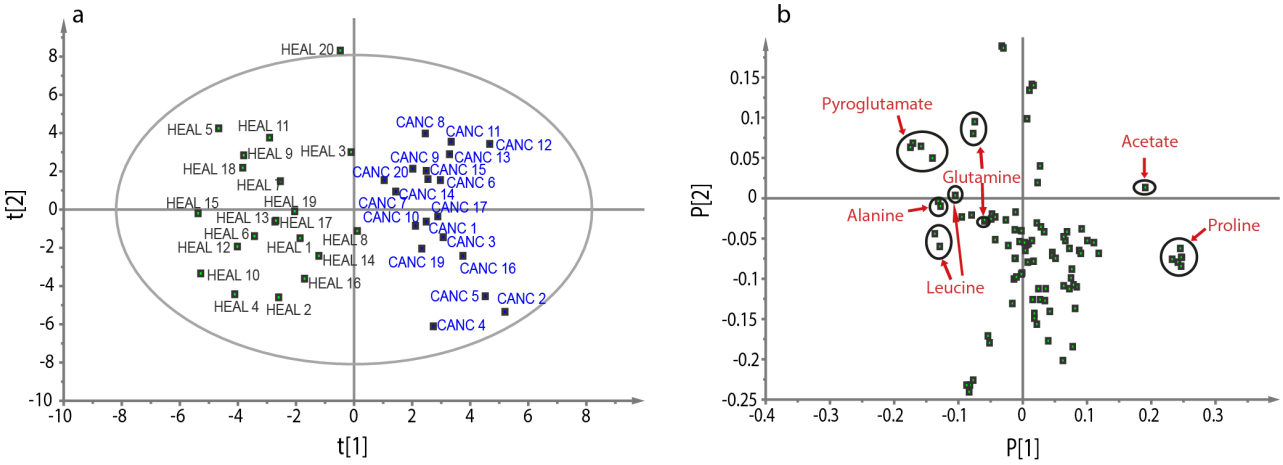


Figure 3.8. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the peak intensities from the NUS HSQC spectra of 39 synthetic serum samples. In (b), the buckets circled in black contain the peak intensities from the biomarkers defined in part 2. The first dimension explains 10.1% of the total variance while the second dimension explains 8.4% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $8.26 \cdot 10^{-5}$.

Overall, we reach the same conclusion with NUS HSQC as with NUS DQF-COSY: using NUS allows obtaining the same information as conventional 2D NMR with less experimental time. This allows analyzing samples with a higher throughput.

2.3. UF COSY

For the UF COSY, the bucketing was similar to the bucketing of conventional DQF-COSY, with the exception that most of the correlation peaks close to the diagonal were not taken into account when it was difficult to distinguish the correlation peak from the diagonal peak. Otherwise, the statistical analysis was similar to the conventional DQF-COSY. The volumes were exported to an Excel file and normalized relatively to the total volume. After the export of the Excel file to SIMCA, the data were scaled to the UV and analyzed with OPLS-DA. The results are shown in figure 3.9.

Once again, a very good separation of the two groups is seen on the score plot. The very low *p*-value from the CV-ANOVA analysis proves that this separation has little to no chance to be non-significant. The loading plot shows that, once again, 5 metabolites are mainly responsible for the separation: proline, glutamine, alanine, leucine and alanine. Similarly to the conventional and NUS DQF-COSY, buckets from a biomarker tend to have the same coefficients, with a few exceptions. Three buckets with proline peaks and one bucket with pyroglutamate peaks have a

significantly lower coefficient in the first dimension compared to the other buckets from the same metabolite. For three of these buckets (two from proline and the one from pyroglutamate), the correlation peak is just above the detection level. The last bucket from proline has a rather low S/N combined with overlap with another peak. These buckets show once again that low sensitivity and overlap can both cause problems for metabolomics studies since they can both raise the standard deviation and thus hinder differences between groups.

For UF COSY, we can conclude that the pulse sequence is also adequate for statistical analysis. While sensitivity is an issue for UF 2D NMR in general, we proved that we could now detect peaks from molecules with concentrations as low as a few hundreds of μM in an experimental time adequate for most metabolomics studies, as showed in a previous study, which makes it possible to detect the major metabolites in many biological samples.[38] One advantage of UF 2D NMR is that resolution is no longer dependent on the experimental time, as long as interleaving is not necessary. The experimental time can be thus adjusted to obtain the desired sensitivity.[241]

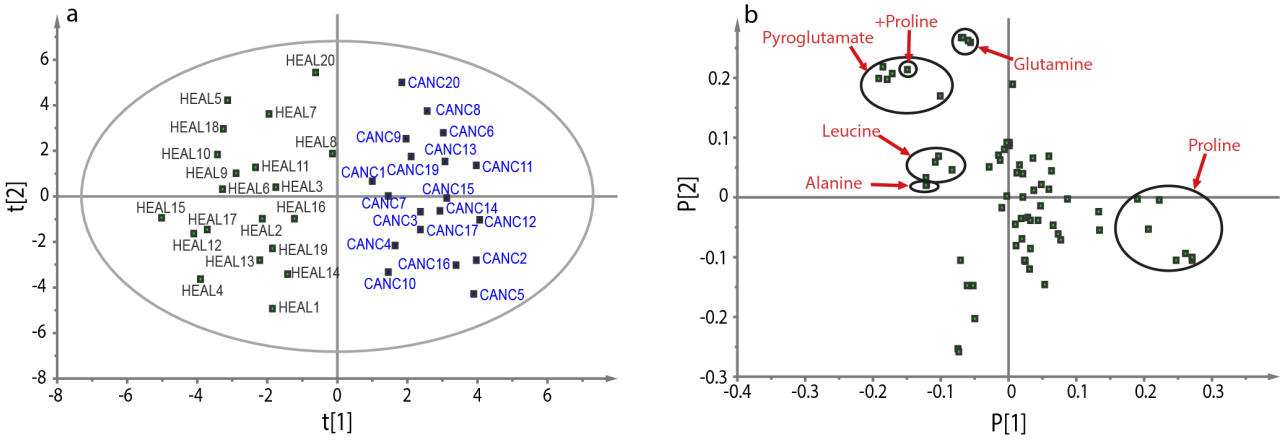


Figure 3.9. Score plot (a) and loading plot (b) of the OPLS-DA analysis of the peak intensities from the UF COSY spectra of 39 synthetic serum samples. In (b), the buckets circled in black contain peak intensities from the biomarkers defined in part 2. The first dimension explains 11.8% of the total variance while the second dimension explains 20.6% of the total variance. The CV-ANOVA analysis showed that the *p*-value was $4.26 \cdot 10^{-4}$.

In conclusion, we manage to show the usefulness of fast 2D NMR methods for metabolomics, thanks to the collection of synthetic serums already used in part 2. We have shown that the NUS DQF-COSY, the NUS HSQC and the UF COSY could be used for the acquisition of the 2D spectra for these samples with shorter acquisition times than conventional DQF-COSY and HSQC. We also showed that these spectra could be used for metabolomics studies and having similar results compared to their conventional counterparts. These findings have been published recently.[267]

From this study, 3 independent other studies were initiated, which will be covered in the next 3 parts:

-As mentioned in section 1.2.1, some correlation peaks in UF COSY were lost because of overlap with diagonal peaks. Therefore, it would be interesting to develop a pulse sequence with sensitivity close to UF COSY and without diagonal peaks. We chose to develop a UF variant of double-quantum spectroscopy and its development and optimization will be covered in part 4.

-In this chapter, NUS was used to reduce the total experiment time. Another option is to keep the total experimental time constant and use NUS to increase sensitivity or resolution, by increasing the number of scans or the number of points in the indirect dimension. This will be explored in part 5.

When compared to the literature, the experimental time of NUS HSQC is still close to experimental times seen in conventional HSQC of complex mixtures,[27] sometimes even higher.[26] One of the main reasons for these changes was the inter-scan delay, which is often set to very short values in 2D experiments, but these short inter-scan delays could be decreasing the experimental time at the cost of sensitivity. Since the T_1 are not mentioned in these studies, it is not possible to know if this low inter-scan delay was really detrimental to the sensitivity. This observation led us to develop a tool to quickly estimate the range to T_1 in a complex mixture to make the choice of the inter-scan delay more adapted to the sample. This study will be covered in part 6.

Part 4. Ultrafast double-quantum spectroscopy for diagonal-free 2D spectra

Goal: The development of an ultrafast variant of double-quantum spectroscopy, in order to obtain a spectrum with sensitivity close to that of UF COSY and no diagonal peaks. Several advancements from the CEISAM laboratory that have been published during this thesis has been used during this study, which will be included in the discussion, along with further collaborations based on this work.

While UF COSY makes it possible to obtain potent results for metabolomics, the presence of diagonal peaks is a hindrance since they can overlap with some correlation peaks. Therefore, we wanted to use a new UltraFast (UF) pulse sequence that would give only correlation peaks, with sensitivity close to that of UF COSY, unlike UF HSQC at natural abundance. One possibility was double-quantum spectroscopy (DQS). The DQS pulse sequence was known since 1980 but, to the best of our knowledge, no UF variant of the DQS had been developed at that point.

Our first step was the development of the pulse sequence, with an explanation of the evolution of double-quantum (DQ) coherences during spatial encoding and the optimization of the parameters in the pulse sequence. Then a variant of the UF DQS was developed to obtain a spectrum close to a COSY spectrum but with no diagonal peaks. Finally the two variants were tested on a mixture of metabolites to test the sensitivity of the pulse sequence and to show the potential of UF DQS for complex mixture analysis.

1. Development and optimization of the DQS pulse sequence

Since the UF version of DQS had not been described before, we started by its development with a simple sample. We chose a sample of sucrose at 100 mM in D_2O, in order to have no concern about sensitivity and to acquire the spectra in a matter of seconds. One advantage of sucrose for the setup of UF 2D NMR experiments is that all the chemical shifts of protons are present in a spectral width of around 2 ppm, between 3.4 and 5.4 ppm. It allowed the single-scan acquisition of the whole spectrum with a Bruker Avance I 600 MHz equipped with a 5 mm TCI cryoprobe triple resonance ($^{1}H/^{13}C/^{15}N$) already used in part C for the acquisition of NUS spectra, without the need to compromise between sensitivity and resolution or to use interleaved acquisitions. A first test was done to verify if the acquisition of UF pulse sequences was possible with this spectrometer, then the UF DQS pulse sequence was developed and compared to the conventional DQS pulse sequence.

1.1. Calibrations of UF experiments for the 600 MHz

The first step was to record the UF COSY spectrum for a first calibration of the spatial encoding with this sample and to verify the performance of the gradients in this spectrometer. Spatial encoding was obtained using the scheme described in Figure 1.23. Two chirp pulses of 15 ms with a sweep range of 10 kHz were used, together with two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). During the

acquisition, 64 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 718 μs and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available). These parameters were calculated from the website developed by the EBSI group.[37] 1048 points were acquired for each gradient pair. Only one scan was acquired, but presaturation of the water peak during 2 s was added to obtain a clean spectrum.

The processing was done in a similar manner to the UF spectra in part C. Briefly; a home-written routine from the EBSI group was used, developed by Benoît Charrier,[37] where apodization of the ultrafast dimension was done with an optimized Gaussian function. For the apodization of the conventional dimension, a sine-bell function was used. After FT, the chemical shifts were referenced to those described in the HMDB database. The processed spectrum is shown in figure 4.1 and compared to the conventional COSY spectrum obtained with the "cosygpprqf" pulse sequence with 256 t_1 increments, 1 scan per increment and a inter-scan delay of 2.5 s.

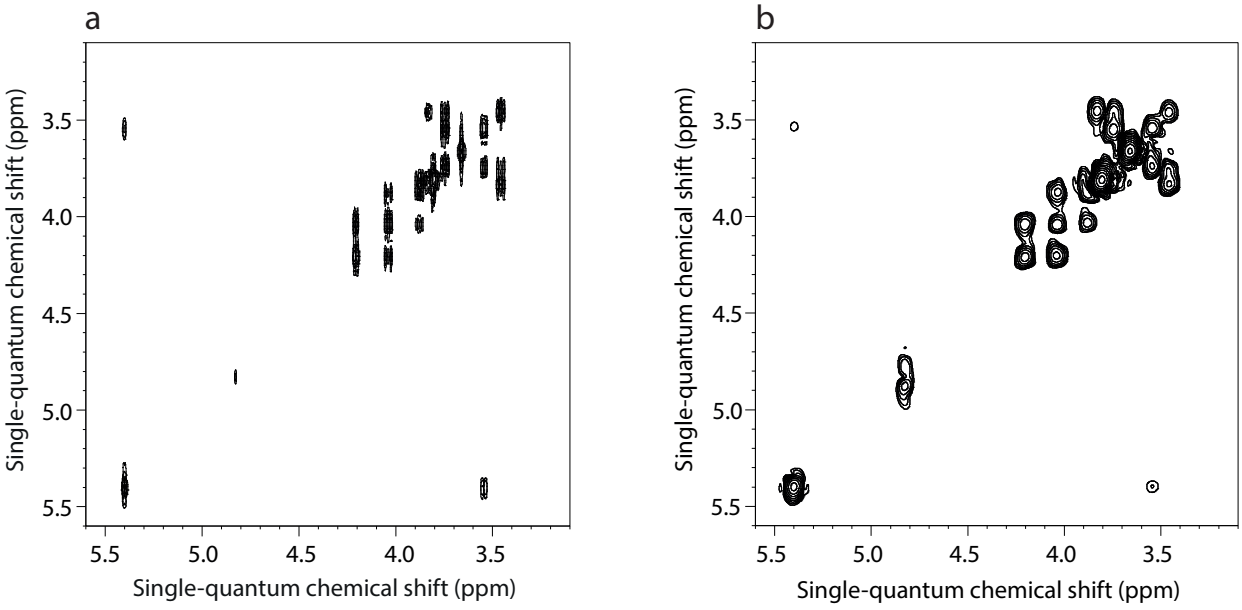


Figure 4.1. Conventional COSY spectrum of sucrose (a), acquired in 12 min and 44 s, along with the UF COSY spectrum (b), acquired in 2 s. The ultrafast dimension is vertical and the conventional dimension is horizontal.

In the conventional COSY spectrum, we can effectively see all the ^{1}H resonances within only 2 ppm for the spectral width. The UF COSY spectrum is also similar to the conventional spectrum, with all the peaks visible in only one scan. The important parameters for the acquisition of the UF COSY (amplitude of the gradients for spatial encoding G_e, amplitude of the gradients for acquisition G_a and length of the gradients for acquisition T_a) were calculated from the desired spectral widths (in Hz) in each dimension. Indeed, as shown during the thesis of a PhD student of the EBSI group, Meerakhan Pathan,[268] it is possible to calculate, prior to the acquisition, all the required parameters specific to UF acquisition by using the targeted spectral widths as an input value.

The duration of the readout gradient, T_a can be linked to the spectral width in the conventional dimension SW_c by the relation:

$$SW_c = \frac{1}{2(T_a+\tau_{gr})} \qquad (4.1)$$

where τ_{gr} is the delay allowed for gradient switching. For G_a, its value is linked to the spectral width in the ultrafast dimension SW_u by the relation:

$$SW_u = \frac{G_a \cdot T_a \cdot L \cdot \gamma}{T_e} \qquad (4.2)$$

where L is the length of the coil, γ is the gyromagnetic ratio and T_e the duration of the spatial encoding (usually set to 30 ms for the best compromise between resolution and diffusion effects).

With these values, it becomes possible to calculate G_e, with the following equation:

$$G_e = \frac{G_a \cdot T_a \cdot \gamma_a \cdot \Delta\upsilon}{2 \cdot SW_u \cdot T_e \cdot \gamma_e} \qquad (4.3)$$

where γ_a is the gyromagnetic ratio of the nucleus during acquisition, γ_e is the gyromagnetic ratio of the nucleus during spatial encoding and $\Delta\upsilon$ is the bandwidth of the chirp pulse.

These equations are used in the website written by B. Charrier, from the EBSI laboratory.[37] From the values of the website, we obtained the desired spectral width with a linewidth at half-height similar to the UF COSY from the 500 MHz, between 40 and 50 Hz. This indicates good performances from the gradients of the 600 MHz. Further proof of the good performance of the gradients is the good alignment of the two spectra obtained by positive and negative gradients, which was done in the spectrum in figure 4.1.b.

For the spectrum itself, we see many correlation peaks close to the diagonal peaks, but little significant overlap visible in the COSY spectrum, for the conventional or UF variant. From these observations, we concluded that we could develop the UF DQS with this system (spectrometer and sample).

1.2. Development of the UF DQS pulse sequence

1.2.1. Behavior of double-quantum coherence during spatial encoding

In the conventional DQS pulse sequence,[269] as shown in figure 1.12, the double quantum coherence is generated during the spin echo and evolves during t_1. Since DQ coherences cannot be directly seen during acquisition, they are encoded in the indirect dimension –in the UF case, the spatial encoding dimension- and reverted into single-quantum (SQ) coherence to observe magnetization during acquisition. But, to the best of our knowledge, the evolution of DQ coherence under spatial encoding had not been described. Spatial encoding has already been applied to generate multiple-quantum (MQ) spectra,[270] but in this case, the spatial dimension is encoded onto the MQ transition frequencies rather than the evolution time, and the resulting spectrum is an one-dimensional image where losses in intensity shows the MQ transition frequencies. Zero-quantum coherence has also been encoded in several studies, which allow a good resolution in

inhomogeneous fields since zero-quantum coherences are not affected by these inhomogeneous fields.[271, 272]

For our first test, the hypothesis was that during spatial encoding, the DQ coherence acquires a phase shift proportional to the position in the NMR tube and the sum of the two frequencies, as shown in the equation below:

$$\phi = C(\Omega_1 + \Omega_2)z$$

where ϕ is the phase shift acquired during the spatial encoding, C is a constant depending on the gradient strength and duration during spatial encoding, Ω_1 and Ω_2 are the chemical shifts of the two peaks that are involved in the DQ coherence and z the position in the NMR tube.

If this equation is correct, then during the acquisition, the magnetization will be refocused at a time proportional to the sum of the two frequencies Ω_1 and Ω_2. In other words, the position of the peak in the k-space will be proportional to Ω_1 and Ω_2, meaning that the ultrafast dimension in UF DQS will be the equivalent of the indirect dimension in the conventional DQS. This hypothesis is tested below.

1.2.2. Parameters for acquisition and processing

The conventional DQS spectrum was acquired with the pulse sequence described in figure 4.2.a. The last pulse before acquisition was a 120° pulse instead of a 90° pulse to maximize the intensity of the correlation peaks, as described elsewhere.[107] The delay τ in the spin echo was fixed to a value of 25 ms, for an average $^3J_{H\text{-}H}$ of 10 Hz. Coherence selection was done by gradients of 1 ms, to select the echo pathway for the magnetization that was in the DQ coherence before the last pulse and became SQ coherence after the last pulse. The acquisition lasted 0.34 s with 4096 points acquired. 512 t_1 increments were recorded with two scans per increment and an inter-scan delay of 2.5 s for relaxation and presaturation of the water peak. The experimental time was 50 min and 41 s.

Before FT, the FID was weighted with a sine-bell function in both dimensions, and then the FID was zero-filled to 8192 points in the direct dimension. After FT, the chemical shifts were referenced to those of the HMDB in the direct dimension and to the sum of chemical shifts in the indirect dimension.

For the UF DQS, the pulse sequence is described in figure 4.2.c. Spatial encoding was done similarly to the UF COSY, with two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The delay τ in the spin echo was also fixed to a value of 25 ms, for an average $^3J_{H\text{-}H}$ of 10 Hz. The final hard pulse before acquisition was a 120° pulse, to maximize the intensity of the correlation peaks, and was flanked by the same gradients as conventional DQS for coherence selection. During the acquisition, 64 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 718 μs and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available), similarly to the UF COSY. Only one scan was acquired, but the water peak was presaturated during 2 s.

The processing was done using the same home-written routine as for UF COSY, with the same parameters. After FT, the chemical shifts were referenced to those in the conventional DQS spectrum. The conventional DQS spectrum is shown in figure 4.2.b, while the UF DQS spectrum is shown in figure 4.2.d.

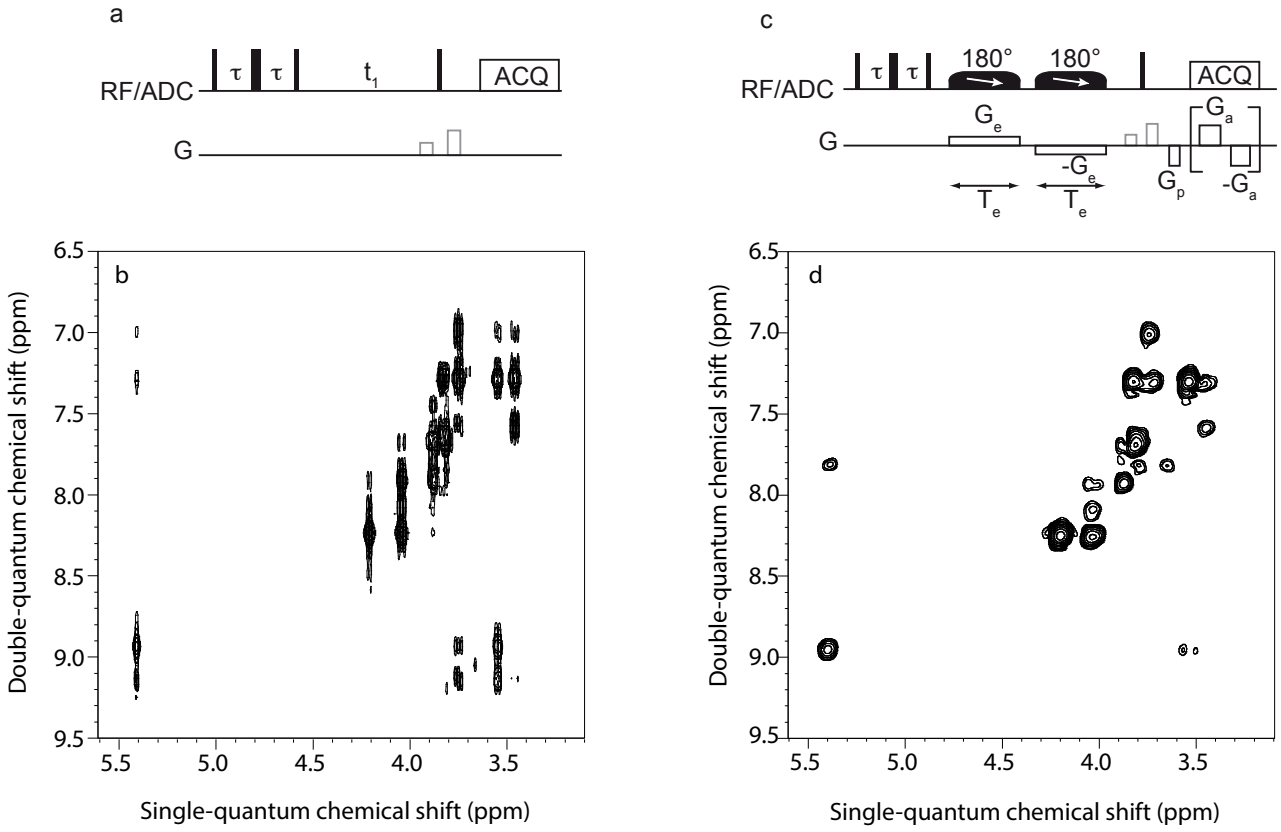


Figure 4.2. The conventional DQS pulse sequence (a) with the DQS spectrum of sucrose acquired in 51 min (b) and the UF DQS pulse sequence (c) along with the UF DQS spectrum of sucrose acquired in 2 s (d). For the UF DQS spectrum, the ultrafast dimension is vertical, while the conventional dimension is horizontal, to make the comparison between conventional and UF DQS spectra easier.

1.2.3. Results and discussion

On the conventional DQS spectrum, we can see that the spectral width necessary to map all the frequencies in the indirect dimension is less than 2.5 ppm, along with the 2 ppm of spectral width in the direct dimension. It is possible with this spectrometer to reach this spectral width with the UF pulse sequences without compromise between resolution and sensitivity, justifying the parameter choices for the UF DQS.

As for the UF DQS spectrum, we can see the similarities with the conventional DQS spectrum. Most of the visible peaks in the conventional DQS spectrum are visible in the DQS spectrum, at the same position. This validates the hypothesis formulated in section 1.2.1 about the behavior of DQ coherence during spatial encoding. The few invisible peaks are relay (also called remote) peaks which, while a possible source of structural information,[273] have usually a lower intensity than correlation peaks. The sensitivity is further reduced by the use of a 120° pulse for the

last hard pulse, which improves the intensity of correlation peaks at the cost of the reduction of the intensity of the relay peaks.

Since the spatial encoding works similarly for DQ coherences compared to SQ, it means that this pulse sequence can be set up the same way as the other pulse sequences like UF COSY or UF HSQC. A good first guess can be obtained from a UF COSY spectrum, by adding the two coordinates from each correlation peak. From that spectrum, the spectral width can be deduced and with the use of the formula explained in section 1.1, and can be set up with the website from the EBSI,[37]. One last note is the presence of one peak in the UF DQS spectrum (at around 5.4 ppm in the conventional dimension and 7.9 ppm in the ultrafast dimension) that is not present in the conventional DQS spectrum. This peak is probably an artifact, but since the use of longer gradients for coherence selection had no visible impact on this peak intensity, it seems that the peak follows the same coherence pathway as the correlation peaks.

In conclusion, the UF DQS pulse sequence works, despite the presence of an unknown artifact in the UF DQS spectrum. But the DQS pulse sequence is, to our knowledge, rarely used and no database exists for peak positions in the spectrum. It could be possible to create such a database, but one other possibility could be to change the appearance of the spectrum so it looks like a spectrum from much a more commonly used pulse sequence. This possibility is discussed in the next chapter.

2. COSY-like spectrum without diagonal peaks from the DQS pulse sequence

It is possible to obtain, from a DQS pulse sequence, a spectrum that looks like a COSY spectrum without diagonal peaks, but with the addition of the relay peaks. There are two ways to obtain this kind of spectra. The first is a processing operation applied to the spectrum. The second is from the DQSSY pulse sequence, as depicted in section 2.2.4 of part 1. Both possibilities will be presented in this order for this chapter.

2.1. Processing approach to obtain a COSY-like spectrum from the DQS pulse sequence

In the DQS spectrum, each correlation peak has a chemical shift Ω_1 in the single-quantum frequency dimension and an apparent chemical shift $\Omega_2+\Omega_1$ in the double-quantum frequency dimension, Ω_2 being the chemical shift of the spin I_1 sharing a J-coupling with the spin I_2. We can also say that the position of the peak is $(\Omega_2; \Omega_2+\Omega_1)$ in the DQS spectrum. We can tilt the spectra, similarly to homonuclear J-RES in section 2.2.1, and for the DQS spectrum, the resulting transformation is shown below:

$$(\Omega_i'; \Omega_j') = (\Omega_i; \Omega_j - \Omega_i)$$

By changing Ω_1 by Ω_i and Ω_2 by $\Omega_i+\Omega_j$, we can see that the new coordinates of a correlation peak after shearing are $(\Omega_i; \Omega_j)$, which is the same as the coordinates from a correlation peak in a COSY spectrum and a DQSSY spectrum. This transformation can be done in Topspin, by changing the value of the parameter ALPHA along F_1, then using the command "ptilt". For our

experiments, a value of 0.5 allowed us to obtain the sheared spectrum with correct positions for the correlation peaks. After shearing, the chemical shifts were referenced to those of the COSY spectrum. The results for both the conventional DQS spectrum and the UF DQS spectrum are presented in figure 4.3.

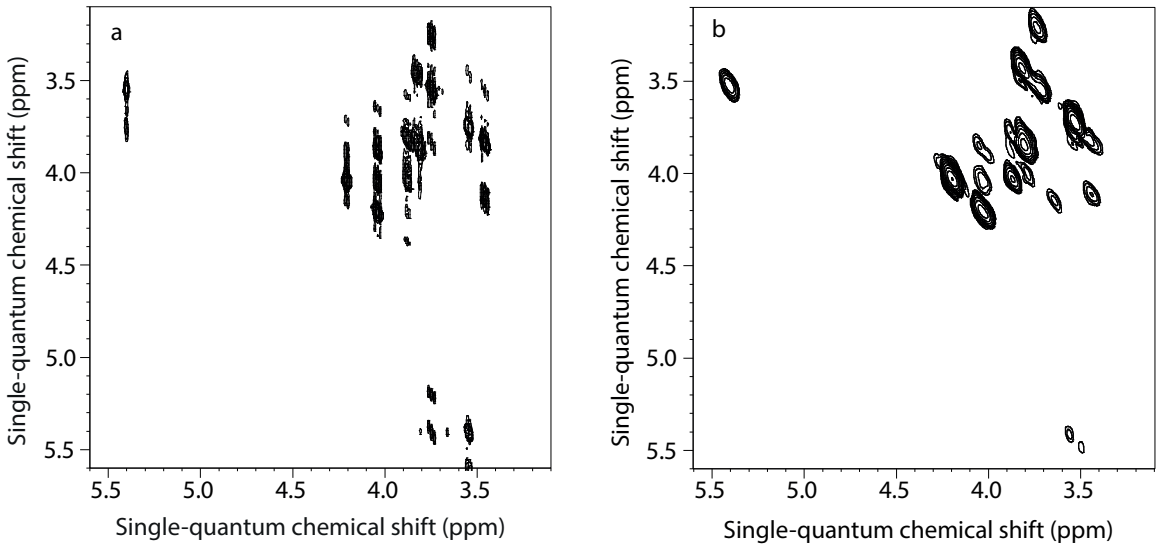


Figure 4.3. Conventional DQS spectrum from figure 4.2.b after shearing (a) and UF DQS spectrum from figure 4.2.d after shearing (b).

For the sheared conventional DQS spectrum, the correlation peaks are all in the same position as the correlations peaks from conventional COSY. The same conclusion can be drawn for the UF DQS spectrum. This way, we effectively obtained the equivalent of a COSY spectrum without the diagonal peaks. But two main differences remain; the first is the lineshape of the peaks, which is distorted because of the shearing process; the second is the presence of the relay peaks. The first problem could be resolved with the other method to obtain the COSY-like spectrum from a DQS pulse sequence

2.2. DQS pulse sequence variant to obtain a COSY-like spectrum: symmetrized DQS pulse sequence

2.2.1. Description of the conventional and UF variants of the symmetrized DQS pulse sequences

As explained in part 2.2.4 of part 1, the DQSSY pulse sequence can also be used to obtain a COSY-like spectrum. For the UF variant of the DQSSY, we tried to obtain the same simplification as in the conventional DQSSY pulse sequence by replacing the second t_1 evolution time by a second spatial encoding. With the same reasoning as for conventional DQSSY, magnetization during the second spatial encoding would be dephased by the following term:

$$\phi_2 = -C\Omega_1 z$$

This would give a net dephasing after the two spatial encodings of:

$$\phi_2 = C\Omega_2 z$$

Which means that, during acquisition, the position of the peak in the k-space would be dependent of Ω_2 only, while magnetization is evolving at the frequency ν_1 during acquisition. The correlation peaks in the UF DQSSY spectrum should be at a position (Ω_1; Ω_2) in the conventional dimension and in the ultrafast dimension, respectively.

To obtain this result, the constant C must remain the same for the two spatial encodings. This means that the gradients must act the same way in the two spatial encodings (same strength, same shape, same duration). This will be verified by the comparison between the spectrum from conventional DQSSY and UF DQSSY.

2.2.2. Parameters for acquisition and processing

For the conventional DQSSY pulse sequence, in figure 4.4.a, the same parameters as for the conventional DQS were chosen. Briefly, 512 t_1 increments were acquired with 2 scans per increment and the inter-scan delay was 2.5 s. The delay τ was also fixed at 25 ms, and the acquisition lasted 0.34 s with 4096 points recorded during acquisition.

The processing was also similar to the conventional DQS spectrum. The FID was weighted with a sine-bell function in each dimension, and then zero-filled to 8192 points in the direct dimension. After FT, the spectrum was referenced to the peaks in the COSY spectrum. The processed spectrum is shown in figure 4.4.b.

For the UF DQSSY pulse sequence, in figure 4.4.c, the parameters are also similar to the UF DQS pulse sequence. Spatial encoding was done with two series of two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two pairs of two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The first two chirp pulses were placed before the last hard pulse, of 120°, and the last two after the last hard pulse. The delay τ in the spin echo was also fixed to a value of 25 ms, for an average $^3J_{H\text{-}H}$ of 10 Hz. During the acquisition, 64 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 718 μs and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available), similarly to the UF COSY. Only one scan was acquired, but the water peak was presaturated during 2 s.

The processing was also similar to the UF DQS spectrum, with the same home-written routine, with the same parameters as UF COSY. After FT, the chemical shifts were referenced to those in the conventional DQSSY spectrum. The processed UF DQSSY spectrum is shown in figure 4.4.d.

2.2.3. Results and discussion

For the conventional DQSSY spectrum, the result is very similar to the conventional DQS spectrum after shearing, in figure 4.3.a. Both the correlation and the relay peaks are at the same position, showing the similarities between the two procedures. The UF DQSSY spectrum is also very similar to the UF DQS spectrum after shearing. The main difference is the better sensitivity of some relay peaks. The analysis of the spectrum shows that peaks are still present in the zone of the diagonal peaks, at least two at around 3.9 ppm and the other at around 4 ppm. These peaks are

present in the four spectra, both conventional and Ultrafast, which excludes a possible artifact from the DQSSY pulse sequence or the UF variants. The analysis of the 1D ^{1}H spectrum in the HMDB database shows that the peak at 3.9 ppm was expected, since two groups of coupled proton have chemical shifts in this area, but the last one has only one group of protons at this chemical shift. Further analysis showed that it is a relay peak since this peak has a chemical shift of around 8.1 ppm, which is the sum of the chemical shift of two protons (at 3.89 ppm and 4.21 ppm) sharing a coupling with the peak at 4.1 ppm. Therefore, we confirmed the complete absence of diagonal peaks in the DQSSY spectra.

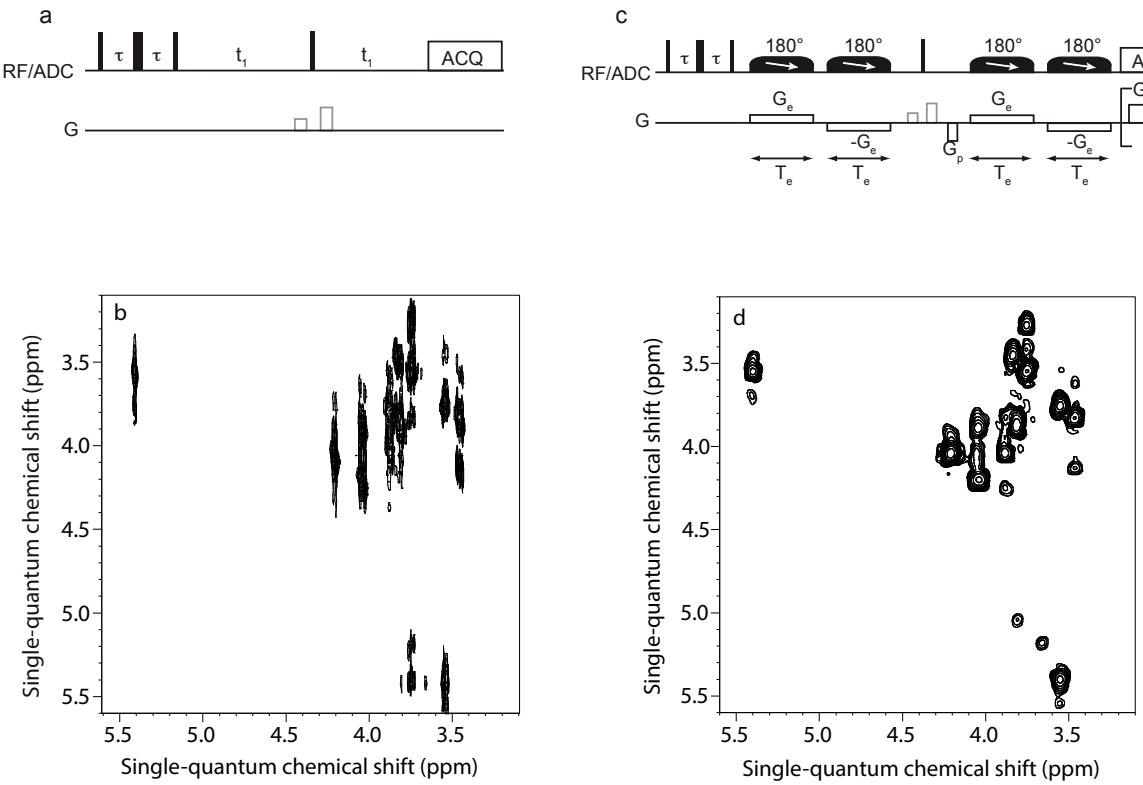


Figure 4.4. The conventional DQSSY pulse sequence (a) with the DQSSY spectrum of sucrose acquired in 51 min (b) and the UF DQSSY pulse sequence (c) along with the UF DQSSY spectrum of sucrose acquired in 2 s (d). For the UF DQSSY spectrum, the ultrafast dimension is vertical, while the conventional dimension is horizontal, to make the comparison between conventional and UF DQSSY spectra easier.

From this first analysis, we showed that the two approaches to obtain COSY-like spectra from UF DQS both work and give similar results. Both approaches have pros and cons. The processing approach comes from the DQS spectrum, and the knowledge of the frequencies in the DQ dimension makes the assignment of relay peaks easier, as shown above. However, since the lineshapes are distorted, it makes more difficult the definition of peaks for a potential quantification. On the contrary, the DQSSY approach gives a spectrum with cleaner lineshapes, but the elucidation of relay peaks is more difficult. At first glance, shearing from the DQS spectrum seems more like a tool for elucidation, to quickly identify correlation peaks while facilitating the elucidation of the remote peaks. Meanwhile, the DQSSY approach seems to be more adequate to identify the correlation peaks from a COSY database and quantify thanks to the clean lineshapes.

One last point to discuss is that an alternative approach, UF-DISSECT (for DIagonal-Suppressed Spin-Echo Correlation specTroscopy), was published almost at the same time as the present work, to obtain COSY-like spectra without diagonals.[274] It would be interesting to compare these different approaches, in terms of sensitivity, repeatability and robustness. While the UF DQS pulse sequence worked correctly for the sucrose sample, we were more interested in the analysis of complex mixtures for metabolomics, with a larger range of chemical shifts and usually a strong water peak. The adaptation of the UF DQS pulse sequence for these kinds of samples will be described in the next chapter.

3. Application of the UF DQS pulse sequence for the analysis of a complex mixture

For this analysis, we decided to use a sample of 7 metabolites in H_2O/D_2O (90/10) with enough sensitivity to obtain all the correlation peaks in one scan. The 7 metabolites are: lactate (18 mM), alanine (13 mM), asparagine (20 mM), citrate (13 mM), serine (17 mM), proline (24 mM) and glycerol (27 mM). This sample was obtained by the dilution of the stock samples described in chapter 1.1 of part 2.

The spectral width to represent all the peaks without folding are around 3 ppm in both dimensions (and possibly more for the DQS spectrum), which is not possible to obtain without sacrificing sensitivity or resolution for the UF experiments on the 600 MHz used above. For this reason, we chose to use interleaving approach, which will be described in a first part, and then the results for UF DQS and UF DQSSY will be presented, with a comparison of sensitivity for the two pulse sequences with the UF COSY.

3.1. Interleaved UF pulse sequences

The interleaved UF spectroscopy is an approach proposed since the beginning of UF 2D NMR,[215] to address the problem of gradient strength in UF experiments and inspired by interleaved Echo Planar Imaging (EPI).[275] The idea of this approach is to acquire at least two scans with different delays between the end of the mixing period and the beginning of acquisition, as shown in figure 4.5.a for the UF COSY pulse sequence. This change in the timing of acquisition changes the trajectory in the k-space, as seen in figure 4.5.b, creating two different datasets. By combining the two FIDs in an interleaved fashion, it is possible to obtain a new FID whose dwell time in the conventional dimension is halved, which decreases the effective spectral width in the conventional dimension after FT.

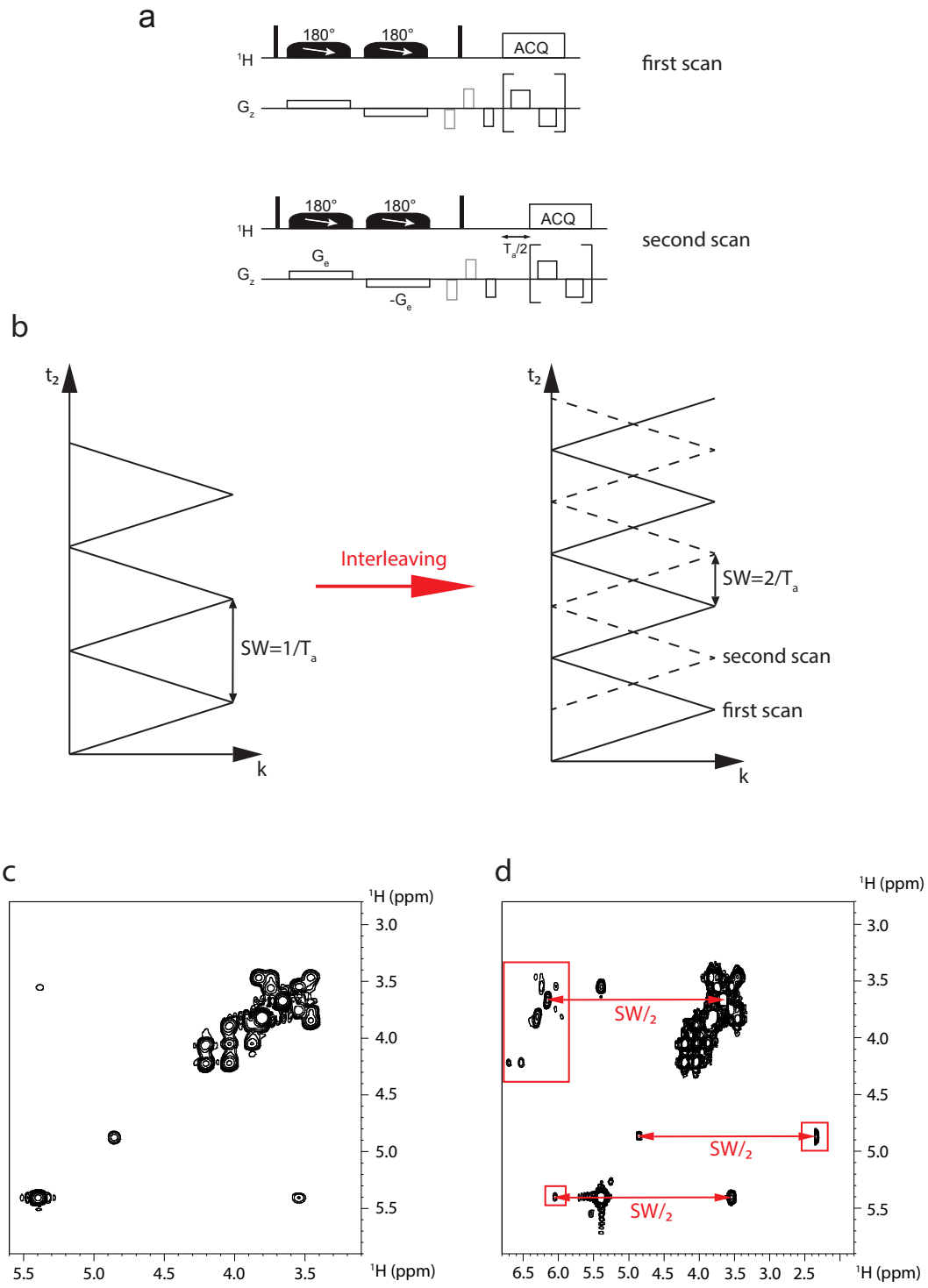


Figure 4.5. Principle of interleaving in UF COSY (a), with the example of a two-fold interleaving. The first scan is recorded as usual for the UF COSY pulse sequence, but the second scan has a delay before acquisition. This delay changes the trajectory in the k-space (b), creating two datasets that, when combined, form a new FID with the dwell time in the conventional dimension reduced by half. This allows doubling the spectral width in the conventional dimension, as seen in the difference between non-interleaved (c) and interleaved (d) UF COSY spectra of sucrose. But because of differences of phase between the two FID, the interleaved FID is often plagued by artifacts, which are shifted by half the spectral width in the conventional dimension. In figure 4.5.d, these artifacts are framed in red.

In figure 4.5 d, the spectral width is increased by two, but it is possible to increase even further the spectral width. More importantly, this allows more flexibility for the acquisition of UF spectra, since it can also be used to increase the spectral width in the ultrafast dimension without decreasing the spectral width in the conventional dimension.[265] It is also possible to increase resolution in the ultrafast dimension without decreasing spectral width in either dimension and, as mentioned above, to decrease the strength of gradients during acquisition while keeping the same spectral width.

Unfortunately, the interleaved spectrum often contains artifacts. These artifacts mainly come from unwanted differences in phase or amplitude between the FID, which makes the interleaved FID slightly different to an FID with a shorter dwell time. These artifacts are often referred as ghost peaks, a term from Echo Planar Imaging (EPI), where the artifact is a shifted replica of the main image. In a UF spectrum, they have the same chemical shift as genuine peaks in the ultrafast dimension and are shifted by a fraction of the spectral width in the conventional dimension.

These artifacts, which are well known in EPI,[275] are difficult to suppress in UF 2D NMR during acquisition since it would require a robustness in the gradient coil power delivery and timing that cannot be obtained at the moment. This may explain why interleaving is rarely used in UF 2D NMR.[236, 276] However, recent studies from the EBSI group, as part of a master thesis from Laetitia Rouger, have shown the possibility to optimize the interleaved UF pulse sequences to reduce these artifacts,[265] by the introduction of a purge gradient before each scan to suppress the residual transversal magnetization or using higher inter-scan delays to reduce partial saturation. While this optimization makes it possible to reduce the artifacts, it does not suppress them entirely; therefore post-processing strategies should also be used, like symmetrization in UF COSY spectra.

For UF DQS spectra, it is especially important to increase the spectral width in the spectrum since the potential range of chemical shifts is the double compared to UF COSY in the ultrafast dimension. At the very least, the spectral width will be higher in UF DQS than UF COSY. This point is especially important with the magnetic field we used (600 MHz). With this magnetic field, the commercially available gradients do not allow the acquisition of spectra with spectral width high enough to see the aliphatic region for protons in UF COSY, at least not without penalties in resolution or sensitivity

However, one interesting possibility to note is that in UF COSY spectra, the most intense peaks are usually the diagonal peaks. Therefore, one possible advantage of interleaved UF DQS spectra could be the reduction of ghost peak intensities compared to interleaved UF COSY. For this reason, we wanted to compare interleaved UF DQS to interleaved UF COSY. At first, interleaved UF COSY was tested to verify if the interleaved UF approach was correctly implemented in our pulse sequences. Then interleaved UF DQS was tested, along with interleaved UF DQSSY.

3.2. Acquisition of UF DQS and UF DQSSY spectra in water and comparison with UF COSY

3.2.1. UF COSY and implementation of interleaved spectra

For the UF COSY acquisition, the parameters were similar to the UF COSY of sucrose, in section 1.1. Spatial encoding was obtained using two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). During the acquisition, 32 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 1.436 ms and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available). These parameters were calculated from the website developed by the EBSI laboratory[37], using the same calculations introduced in section 1.1. 1048 points were acquired for each gradient pair. Interleaving was used to double the spectral width in both dimension, which resulted in the acquisition of 4 datasets, with 1 scan recorded by data set. The inter-scan delay was 5 s and 4 dummy scans were added to avoid as much as possible the apparition of ghost peaks. The total experimental time was 41 s.

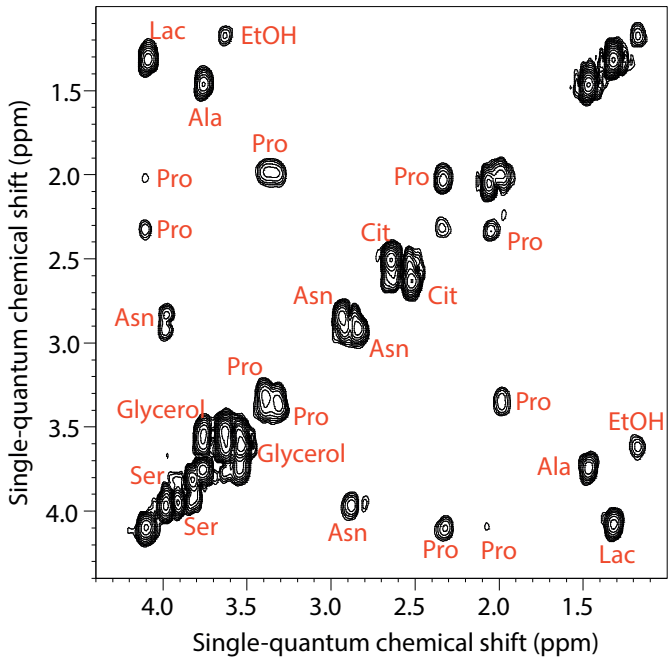


Figure 4.6. UF COSY spectrum of the mixture of 7 metabolites. Each correlation peak is labeled similarly to figure 2.6 and 2.8, except for glycerol. Pro: proline, Cit: citrate, EtOH: ethanol, Ala: alanine, Asn: asparagine, Ser: serine, Lac: lactate.

The processing of the UF COSY spectra was also very similar to the processing of the UF COSY spectrum of sucrose. A home-written routine was used, with an additional step to combine the four datasets into one FID. The apodization of the ultrafast dimension was done with an optimized Gaussian function. For the apodization of the conventional dimension, a sine-bell function was used. After FT, the chemical shifts were referenced to the two diagonal peaks of lactate. The processed spectrum is shown in figure 4.6.

In this spectrum, we can see that we manage to obtain a UF spectrum with a spectral width of at least 3.5 ppm in both dimensions while keeping a good resolution to avoid peak overlap for most correlation peaks. All the visible peaks are either diagonal or correlation peaks, and no ghost peaks are seen, which may be caused by the rather low concentration of the metabolites, which cause some correlation peaks to be close to the noise level. Therefore, it will be difficult to show the potential of UF DQS to reduce the risk to see ghost peaks, since they are not even seen in UF COSY.

For the sample itself, we can see that ethanol is still present in the mixture, but it still has no overlap with any correlation peak. Similarly to the UF spectrum in figure 3.3, some correlation peaks are close to the diagonal, and part of them are overlapping with diagonal peaks, either partially (citrate, asparagine), or completely (serine, glycerol). This justifies the use of the UF DQS pulse sequence. As expected, all the correlation peaks are seen with only 1 scan per data set.

3.2.2. UF DQS

3.2.2.1. Parameters for acquisition and processing

The conventional DQS spectrum was acquired with the pulse sequence described in figure 4.2.a. The last pulse before acquisition was a 120° pulse to maximize the intensity of the correlation peaks, as described elsewhere.[107] The delay τ in the spin echo was fixed to a value of 25 ms, for an average $^3J_{H\text{-}H}$ of 10 Hz. Coherence selection was done by gradients of 1 ms, to select the echo pathway for the magnetization that was in the DQ coherence before the last pulse and became SQ coherence after the last pulse. The acquisition lasted 0.34 s with 4096 points acquired. 1024 t_1 increments were acquired with four scans per increment and an inter-scan delay of 2.5 s for relaxation and presaturation of the water peak. The experimental time was 3 h 23 min and 4 s.

Before FT, the FID was weighted with a sine-bell function in both dimensions, and then the FID was zero-filled to 8192 points in the direct dimension. After FT, the chemical shifts were referenced to the two peaks from lactate in the direct dimension and the sum of the chemical shifts of lactate in the indirect dimension.

For the UF DQS, the parameters for acquisition were similar to the parameters in the UF DQS spectrum of figure 4.2.b. Spatial encoding was done with two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The delay τ in the spin echo was also fixed to a value of 25 ms, for an average $^3J_{H\text{-}H}$ of 10 Hz. The final hard pulse before acquisition was a 120° pulse, to maximize the intensity of the correlation peaks, and was flanked by the same gradients as conventional DQS for coherence selection. During the acquisition, 32 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 1.436 ms and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available), similarly to the UF COSY. Interleaving was used to double the spectral width in both dimensions, which resulted in the acquisition of 4 datasets, with 1 scan recorded per dataset. The

inter-scan delay was 5 s and 4 dummy scans were added to avoid as much as possible the apparition of ghost peaks. The acquisition time was 41 s.

The processing was done using the same home-written routine as UF COSY, with the same parameters and the step to combine the 4 datasets into one FID. After FT, the chemical shifts were referenced to those in the conventional DQS spectrum. The conventional DQS spectrum is shown in figure 4.7.a, while the UF DQS spectrum is shown in figure 4.7.b.

3.2.2.2. Results and discussion

In the conventional DQS spectrum, we can see the good separation of all the correlation peaks, with no overlap between peaks from different metabolites. While not shown in the figure, several artifacts exist for the proline peaks between 2 ppm and 2.3 ppm in the direct dimension. This suggest an effect of strong coupling, similar to certain spectra of INADEQUATE already published.[269]

Despite the rather high concentration of the 7 metabolites, some correlation peaks are still missing in the UF DQS spectrum. These missing peaks concern mainly proline, which has many peaks with strong coupling or loss of sensitivity via J-modulation. Despite this shortcoming, the UF DQS spectrum has a high degree of fidelity with the conventional DQS spectrum. The lineshape is also similar to peaks in UF COSY spectra, another sign that the spatial encoding of double-quantum coherence is very similar to single quantum coherence. We can now verify if the UF DQSSY have similar characteristics.

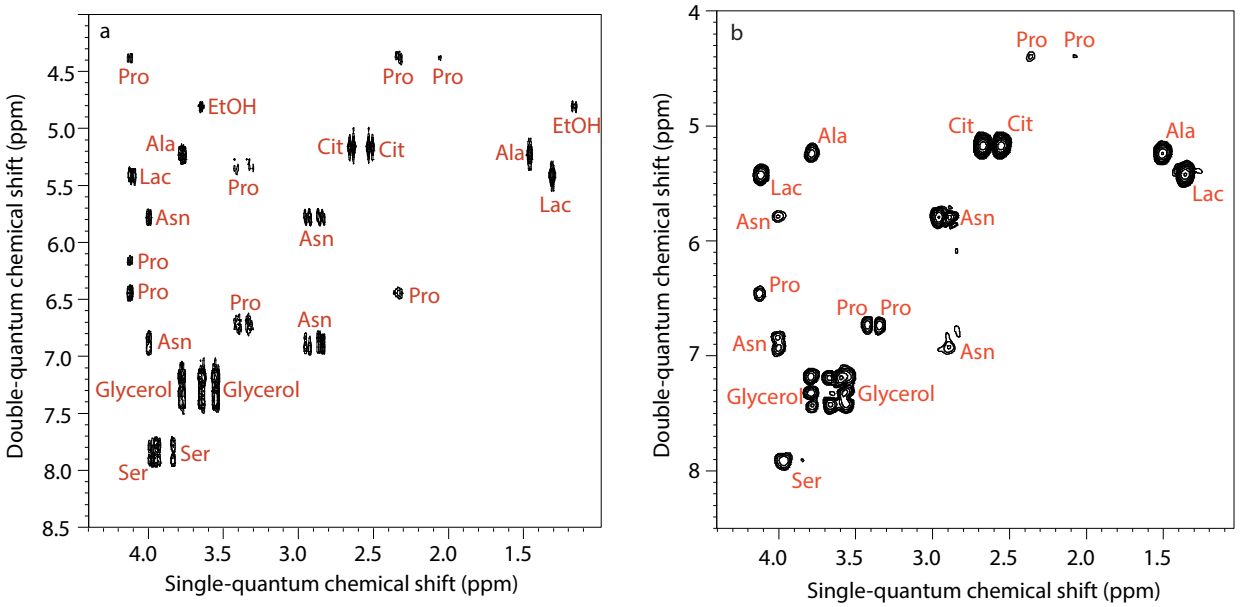


Figure 4.7. The conventional DQS spectrum of the mixture of 7 metabolites acquired in 3 h and 23 min (a) and the UF DQS spectrum of sucrose acquired in 41 s (b). For the UF DQS spectrum, the ultrafast dimension is vertical, while the conventional dimension is horizontal, to make the comparison between conventional and UF DQS spectra easier.

3.2.3. UF DQSSY

3.2.3.1. Parameters for acquisition and processing

For the conventional DQSSY pulse sequence, the same parameters as for the conventional DQS pulse sequence were chosen. 1024 t_1 increments were acquired with 4 scans per increment and the inter-scan delay was 2.5 s. The delay τ was also fixed at 25 ms, and the acquisition lasted 0.34 s with 4096 points recorded during acquisition. The acquisition time was 3 h 31 min and 48 s.

The processing was also similar to the conventional DQS spectrum. The FID was weighted with a sine-bell function in each dimension, and then zero-filled to 8192 points in the direct dimension. After FT, the spectrum was referenced to the peaks in the COSY spectrum. The processed spectrum is shown in figure 4.8.a

For the UF DQSSY pulse sequence, the parameters are also similar to the UF DQS pulse sequence. Spatial encoding was done with two series of two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two series of two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The delay τ in the spin echo was also fixed to a value of 25 Hz, for an average $^3J_{H\text{-}H}$ of 10 Hz. During the acquisition, 32 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 1.436 ms and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available). Interleaving was used to double the spectral width in both dimensions, which resulted in the acquisition of 4 datasets, with 1 scan recorded by data set. The inter-scan delay was 5 s and 4 dummy scans were added to avoid as much as possible the apparition of ghost peaks. The acquisition time was 41 s.

The processing was also similar to the UF DQS spectrum, with the same home-written routine, with the same parameters as UF COSY. After FT, the chemical shifts were referenced to those in the conventional DQSSY spectrum. The processed UF DQSSY spectrum is shown in figure 4.8.b

3.2.3.2. Results and discussion

The conventional DQSSY spectrum of the mixture of metabolites is very close to a COSY spectrum without diagonal peaks. Only few relay peaks exist, like from asparagine (4.1 ppm in the direct dimension, 1.7 ppm in the indirect dimension) or proline (between 2 ppm and 2.3 ppm in the direct dimension, like in the DQS spectrum). Once again, the UF DQSSY spectrum is very similar to its conventional counterpart. Like with the UF DQS, some correlations peaks from proline are missing with an acquisition time of 41 s.

However the correlation peaks from glycerol at 3.8 ppm (in the direct dimension) are also missing, leaving only the relay peak, while the correlation peaks are present in the UF DQS spectrum. The lineshapes of some peaks are also distorted, for example the correlation peaks of citrate or serine, which are peaks with strong coupling. One possibility is that, because of the presence of a second spatial encoding during acquisition, potential effects of strong coupling are enhanced in UF DQSSY spectra compared to UF DQS. We can verify if it also impacts the

sensitivity of other correlation peaks, and if the problem is more pronounced in the UF DQSSY spectrum.

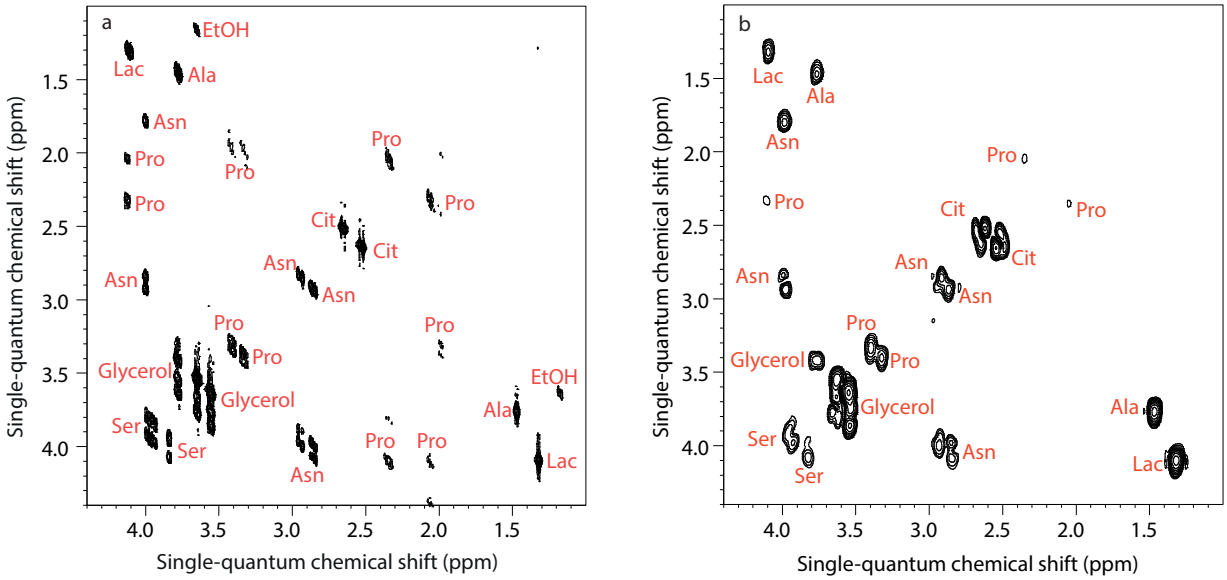


Figure 4.8. The conventional DQSSY spectrum of the mixture of 7 metabolites acquired in 3 h and 32 min (a) and the UF DQS spectrum of sucrose acquired in 41 s (b). For the UF DQS spectrum, the ultrafast dimension is vertical, while the conventional dimension is horizontal, to make the comparison between conventional and UF DQSSY spectra easier.

To compare the sensitivity of UF DQS and UF DQSY with UF COSY, we measured the S/N_{RMS} of some correlation peaks not affected by the lineshape distortion explained above. The S/N_{RMS} was calculated with the S/N toolbox within Topspin and the results are presented in table II.

Table II. S/N_{RMS} of correlation peaks in the UF COSY, UF DQS and UF DQSSY spectra of the mixture of 7 metabolites, measured from 1D slices of rows from the 2D spectrum. The peaks are identified by their position in the spectra, with their chemical shift in the conventional and ultrafast dimensions, respectively. The metabolite from which the peak comes from is also given. The ratio between the S/N_{RMS} of UF COSY and UF DQS is also given. For the UF DQS spectrum, the chemical shifts given are those after shearing the UF DQS spectrum.

Sequence	1,32; 4,1; Lac	3,37; 3,32; Pro	3,55; 3,76; Gly	3,75; 1,47; Ala	2,84; 3,98; Asp
UF COSY	60	39	67	39	17
UF DQS	32	15	22	52	13
UF DQSSY	21	15	27	52	18
COSY/DQS	**1.9**	**2.6**	**3.0**	**0.8**	**1.3**

From these results, we can see that the sensitivity of the correlation peaks are globally lower in the UF DQS and UF DQSSY spectra compared to the UF COSY spectrum. The biggest drop in sensitivity concerns peaks from proline and glycerol, where the strong coupling is the most prominent. However, sensitivity is still close to the UF COSY, contrary to UF HSQC, and some correlation peaks have a higher S/N_{RMS} in the UF DQS and UF DQSSY spectra compared to the UF COSY spectrum, with the example of the alanine peak in the table.

These results show that the behavior of spins during spatial encoding is probably not completely known and that strong couplings probably have an effect on the intensity of peaks when double-quantum coherence is present during spatial encoding. The potential of UF DQS for quantification should also be explored, notably in terms of repeatability and linearity of the peak volume. Two PhD from the EBSI group students will pursue this work, Boris Gouilleux and Laetitia Rouger, the latter having already worked on interleaved UF pulse sequences.

In conclusion, we developed a new pulse sequence, the UF DQS, and its variant, the UF DQSSY. These pulse sequences allow us to obtain spectra without diagonal peaks. The identification of correlation peaks can be done from a COSY database and the sensitivity is close to UF COSY, despite drops in the case of peaks with strong coupling. The pulse sequence is also compatible with the interleaving method of acquiring the data and samples with highly protonated water.

All these qualities makes the UF DQS pulse sequence a credible alternative to UF COSY, especially in the case of spectra where several correlation peaks are close to the diagonal, with the example of tomato extracts, where several sugars are present in the sample.[242]

The next study will be about the time-equivalent NUS, as explained in part 3.

Part 5. Time-equivalent non-uniform sampling for complex-mixture analysis

Goal: To evaluate the capability of time-equivalent non-uniform sampling (NUS) to enhance the information in 2D NMR compared to conventional 2D NMR (also named in this chapter Uniformly Sampled (US) 2D NMR). Two approaches are considered: both the increase of the number of scans and the increase of the number of t_1 increments are tested on small molecules to evaluate the sensitivity of the two approaches.

The question of the sensitivity of NUS in nD NMR is a topic studied since the beginning of its use as a tool to reduce experimental times.[277-279] Today, NUS is mainly considered as a tool to improve sensitivity in the case of biomacromolecules.[280-283] Since small molecules have different dynamics, it is necessary to do a new evaluation with mixtures of small molecules. At first, the case of sensitivity-enhanced NUS (increase of the number of scans) is considered, followed by resolution-enhanced NUS (increase of the number of t_1 increments).

1. Sensitivity-enhanced NUS for small molecules

In macromolecules, the sensitivity enhancements published previously[280-283] are possible mainly because of the short T_2* of their nuclear spins. With standard parameters for the ^1H-^{15}N HSQC of macromolecules, the acquisition time can be past 1.5 T_2* and additional resolution is obtained usually at the cost of sensitivity with US.[280] With exponential-weighted non-uniform sampling, most of the acquired t_1 increments are at the beginning of the FID, where the signal is the most intense. Therefore, increasing the number of scans can provide a significant increase in sensitivity. But in small molecules, the T_2* are usually much longer than those of macromolecules. Therefore, the FID is barely sampled with the standard parameters for 2D spectra, especially with ^1H-^{13}C HSQC, and thus the contribution of all t_1 increments to sensitivity is similar.

In this section, the first attempts to use sensitivity-enhanced NUS are discussed, followed by an analysis of the volume integrity when NUS is used, and finally the analysis of the impact of the number of t_1 increments in sensitivity.

1.1. Analysis of sensitivity in sensitivity-enhanced NUS

1.1.1. Parameters for acquisition and processing

This study was performed with a sample of valine at 1.26 mM in 600 μL of H_2O/D_2O (90/10), with 0.5 mM of TSP for chemical shift reference. The pulse sequence chosen was "hsqcetgpsp", an HSQC variant using 180° adiabatic pulses for the INEPT and retro-INEPT blocks in the heteronuclear channel (^{13}C in this case). The US HSQC spectrum was acquired with 256 t_1 increments, 8 scans per increment and an inter-scan delay of 2.45 s. During acquisition, 2048 points were acquired for an acquisition time of 0.155 s. The delay for the INEPT transfer was fixed to 1.72 ms for a mean value of $^1J_{C\text{-}H}$ of 145 Hz). Adiabatic refocusing was performed with

chirp pulses of 500 μs with a bandwidth of 60 kHz and decoupling during acquisition was done with the GARP supercycle. The total experimental time was 1 h 30 min and 14 s.

Before FT, the FID was weighted with a $\pi/2$ shifted squared sine-bell function in both dimensions, then zero-filling was applied to 4096 points in the direct dimension and 512 points in the indirect dimension. After FT, the chemical shifts were referenced to the TSP and an automatic baseline correction with a polynomial function of order 2 was applied in both dimensions.

For the time-equivalent NUS spectra, the same parameters were applied, with different values of the percentage of NUS of 66%, 50%, 33%, 25%, 20% and 12.5%. The time savings given by recording less t_1 increments was used to record more scans, respectively 12, 16, 24, 32, 40 and 64. The experimental time remained at 1 h 30 min and 14 s and the same processing was applied.

For each spectrum, the Signal-to-Root-Mean-Square-Noise Ratio (SNR_{RMS}) was calculated for 2 of the 4 correlation peaks from valine, one the CH_3 peaks at 1.04 ppm in the proton dimension and the CH peak at 2.28 ppm. The S/N toolbox from Topspin was used to calculate the S/N_{RMS} in 1D slices. The 1D spectrum of valine is shown in figure 5.1.a to show the position for each peak in the proton dimension, while the results for the SNR_{RMS} are presented in figure 5.1.b

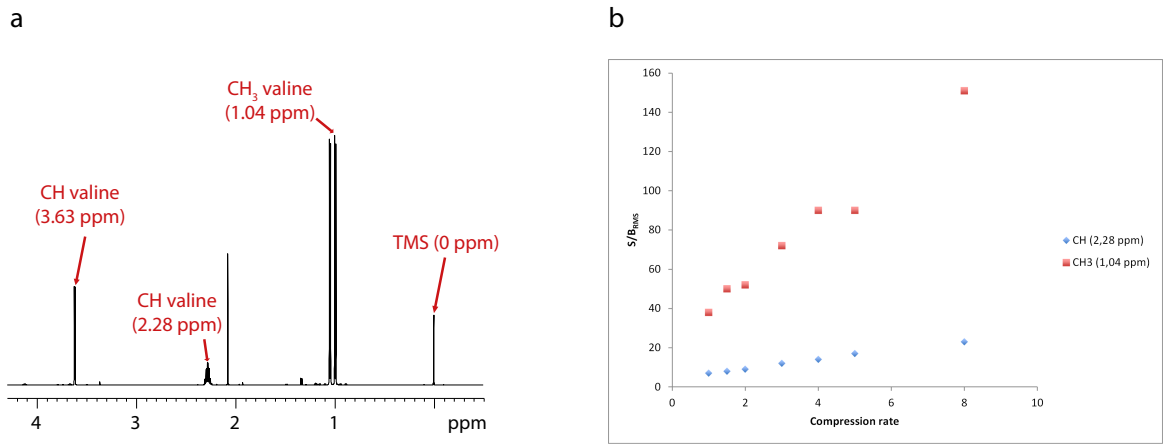


Figure 5.1. 1D spectra from valine (a), along with the evolution of the S/N_{RMS} from two correlation peaks of valine with the compression rate (b). All the spectra were recorded with the same experimental time .The compression rate is defined as the ratio between the number of t_1 increments after the FID reconstruction to the number of t_1 increments actually acquired during the NUS experiment. By definition the US spectrum have a compression rate of 1.

1.1.2. Results and discussion

The position of each valine peak in the HSQC spectrum can be seen in figure 2.8. The two peaks used to evaluate the S/N_{RMS} were the two extreme cases for sensitivity: The CH_3 peak has the best sensitivity while the CH peak has the worst. In order to see the effect of NUS on peaks with low sensitivity, the concentration of valine has been calculated to obtain a S/N_{RMS} between 5 and 10 for the CH peak with the acquisition conditions described above. Since the best S/N value per unit of time for a given peak is when the total relaxation value (acquisition time plus inter-scan delay) is 1.3 times the T_1 of this peak and the highest T_1 value for the valine peaks was two seconds for the

CH peak at 2.28 ppm on the proton dimension, the total acquisition time was fixed at 2.6 s. With an acquisition time of 0.15 s, the inter-scan delay was set at 2.45 s.

In figure 5.1.b, the S/N_{RMS} of the two peaks raise with the compression rate, reaching up to three times the S/N_{RMS} of the US HSQC at a compression rate of 8. However, as shown in figure 5.2, this increase in the S/N_{RMS} does not correlate to a higher ratio between the intensity of the peak and the highest intensity of the noise level. Therefore, it does not seem possible to compare sensitivity between US and NUS spectra from the S/N_{RMS}. This figure also shows that the lineshapes of the peaks become slightly distorted with the use of NUS. Because of this result, we wanted to verify if this lineshape distortion in NUS could cause problems in the measurement of the volume.

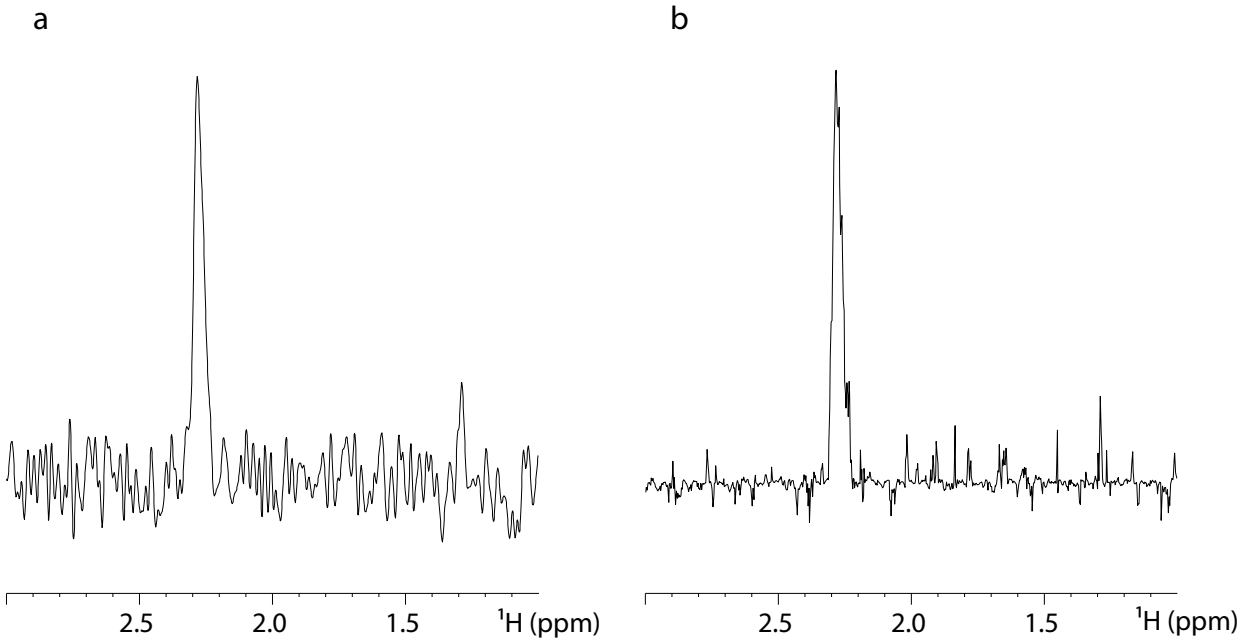


Figure 5.2. Horizontal cross-section of the CH peak of valine (2.28 ppm in the proton dimension) with US acquisition (a) and time-equivalent NUS at 12.5% (b)

1.2. Accuracy of the volume reconstitution with the percentage of NUS

In order to correctly verify the accuracy of volume measurements, it is necessary to obtain a spectrum with the same number of scans and maximum evolution time in the indirect dimension. If the volume is correctly reproduced for the algorithm, the volume should be the same for US and NUS spectra.

The acquisition parameters were the same as for the analysis of sensitivity, except for the number of scans, which remained the same for all experiments. The processing parameters were also identical. The volume from three valine peaks and the peak from TSP are extracted from the Integration toolbox from Topspin. The evolution of the volume depending on the acquired t_1 increments (linked to the percentage of NUS) is shown in figure 5.3.

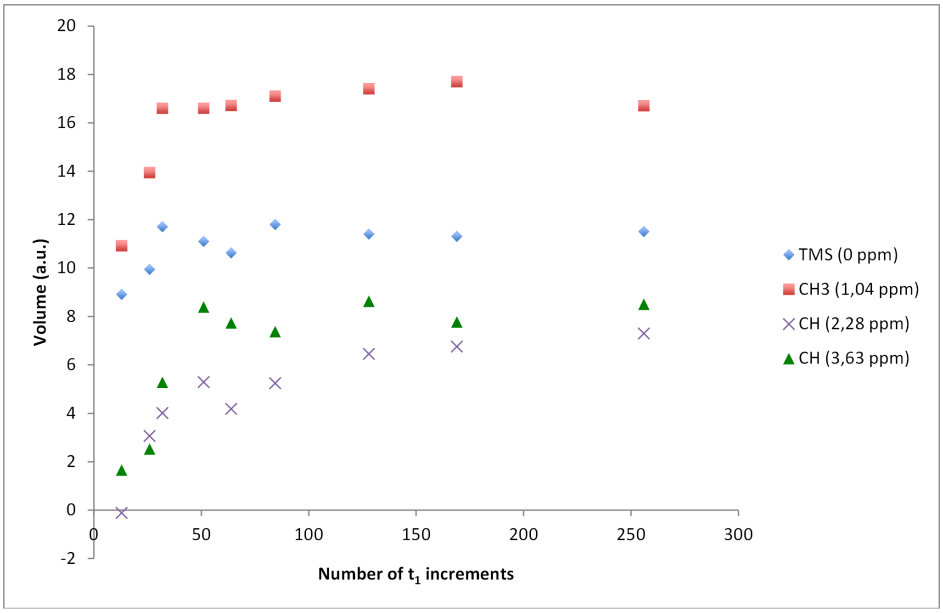


Figure 5.3. Evolution of peak volumes with the percentage of NUS, represented by the number of t_1 increments recorded during acquisition. After the FID reconstruction, 256 t_1 increments constitute the indirect FID. By definition, the spectrum from which 256 t_1 increments were recorded is the US spectrum.

In this figure, we can observe that for low intensity peaks, like the CH peak at 2.28 ppm in the proton dimension, clear volume reduction occurs below 128 t_1 increments recorded (50% of NUS). For the other peaks (S/N_{RMS} > 20 in the US spectrum), the volume reduction occurs at or below 32 t_1 increments recorded (12.5% of NUS). This analysis shows that there are probably two distinct contributions of NUS to the loss of the peak volume: one linked to the difficulty of reconstructing weak peaks and the other linked to the formation of artifacts in the indirect dimension, decreasing the volume of all correlation peaks.

Since even peaks with high S/N_{RMS} can be unfaithfully reproduced by NUS past a certain percentage of NUS, this sets a lower limit for the time reduction allowed by NUS. But one question remains: is that lower limit caused by the percentage of NUS itself or by the low number of t_1 increments recorded during acquisition? As the figure shows, the volume decreases for peaks with high S/N_{RMS} when only 20 t_1 increments are recorded 10% of NUS). Since we are using echo-antiecho acquisition, the real number of points in the indirect FID is ten, which starts to be too low for a correct representation of the peak in the indirect dimension, even in conventional 2D NMR.[164]

Based on the literature review, our first hypothesis was that the most limiting factor was the number of t_1 increments. The basis for this hypothesis was in the reports of NUS for 3D spectra or even spectra with higher dimensionality, where a percentage of NUS even lower than we used can be applied without apparent sensitivity issues.[32, 284, 285] Since the acquisition of US 3D spectra or 4D spectra requires much more increments than 2D spectra, it could explain the lower percentages of NUS available for these experiments.

This hypothesis, along with another point explained below, was the main reason why our focus shifted from sensitivity-enhanced NUS to resolution-enhanced NUS during this study. Since the number of recorded t_1 increments during acquisition remains the same, whatever the percentage of NUS, this problem should be avoided.

1.3. Role of the number of t_1 increments for sensitivity in HSQC

During our optimization of the parameters for sensitivity-enhanced NUS, one of the parameters assessed was the number of t_1 increments in the US HSQC. From this study, we also checked the S/N_{RMS} of two US HSQC spectra with two different numbers of t_1 increments. The results for the S/N_{RMS} from one of the CH_3 peak are shown on figure 5.4.

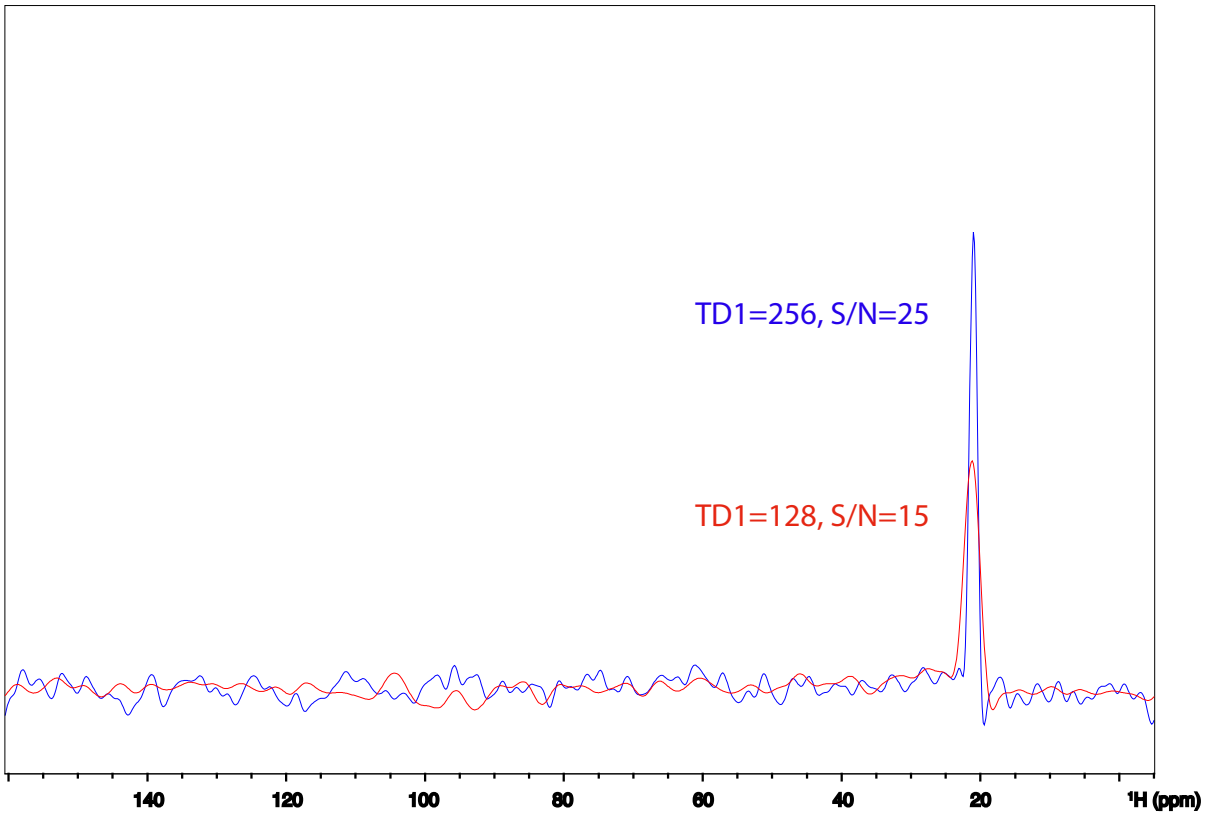


Figure 5.4. Vertical cross-sections of the CH_3 peak from valine (1.04 ppm in the proton dimension) from two HSQC with the same parameters, except for the number of t_1 increments, increased from 128 (red) to 256 (blue). The S/N_{RMS} is indicated for the two spectra.

We can see from this figure that increasing the number of t_1 increments, along with the increase in resolution in the indirect dimension, also increases the S/N_{RMS}, with a factor seemingly similar to the increase in the number to scans, which is $\sqrt{2}$ in this case. This is consistent with an earlier study, where this argument was used to show that aliasing cannot increase the sensitivity per unit of time.[286] This property of the number of increments to increase resolution and sensitivity simultaneously does not seem to be exploited in some studies, especially in metabolomics studies with HSQC spectra of samples with low concentration, where the number of scans in the HSQC spectra is clearly superior to the required number for phase cycling, while the number of t_1 increments remains low.[25, 29, 113, 144]

This point led us to the hypothesis that we could use resolution-enhanced NUS to obtain lower linewidths in the indirect dimension without losing sensitivity. If the evolution of sensitivity with the numbers of t_1 increments was similar to the one of the number of scans, then the same conclusions from figure 5.2 would apply, but with a reduced linewidth.

2. Resolution-enhanced NUS for small molecules

In 2D experiments, in order to reduce as much as possible the experimental times, the resolution in the indirect dimension is often sacrificed by reducing as much as possible the number of t_1 increments. The idea behind resolution-enhanced NUS is to increase the resolution in the indirect dimension with no increase in the experimental time. To verify if this approach is possible with small molecules without losses in sensitivity, the experimental condition for the study will first be fixed (pulse sequences, samples sampling schedule and reconstruction algorithm). Then the results for two pulse sequences, HSQC and TOCSY, will be presented, for the volumes, the signal, the noise and the S/N.

2.1. Optimization of experimental conditions

2.1.1. Pulse sequences

The idea of resolution-enhanced NUS is to increase the maximum evolution time in the indirect dimension after reconstruction without increasing the number of t_1 increments recorded during acquisition. For this study, we wanted to verify the impact of resolution-enhanced NUS for two pulse sequences: HSQC and TOCSY, which have different spectral width in the indirect dimension, sparsity, multiplicity of correlation peaks and sensitivity. They are also interesting for 2D NMR metabolomics studies.[18, 27, 30] In order to be useful in metabolomics, however, a good suppression of the water residual peak is necessary. This property is especially important for resolution-enhanced NUS, since the delay between the first pulse and the acquisition (t_{1max}) will be increased. For example, with the pulse sequence "hsqcetgpsp" is used, as shown in figure 5.5, increasing the acquisition time in the indirect dimension can be detrimental for water suppression.

The increased intensity in the last increments comes solely from the water peak and therefore the water residual peak becomes too high in the resulting spectrum for a good reconstruction of the correlation peaks. For this reason, we changed the pulse sequence for the HSQC spectrum to obtain a better suppression of the water peak. Surprisingly, the best results did not come from one of the water suppression scheme presented in figure 1.3, but from the sensitivity-enhanced HSQC in figure 1.12.

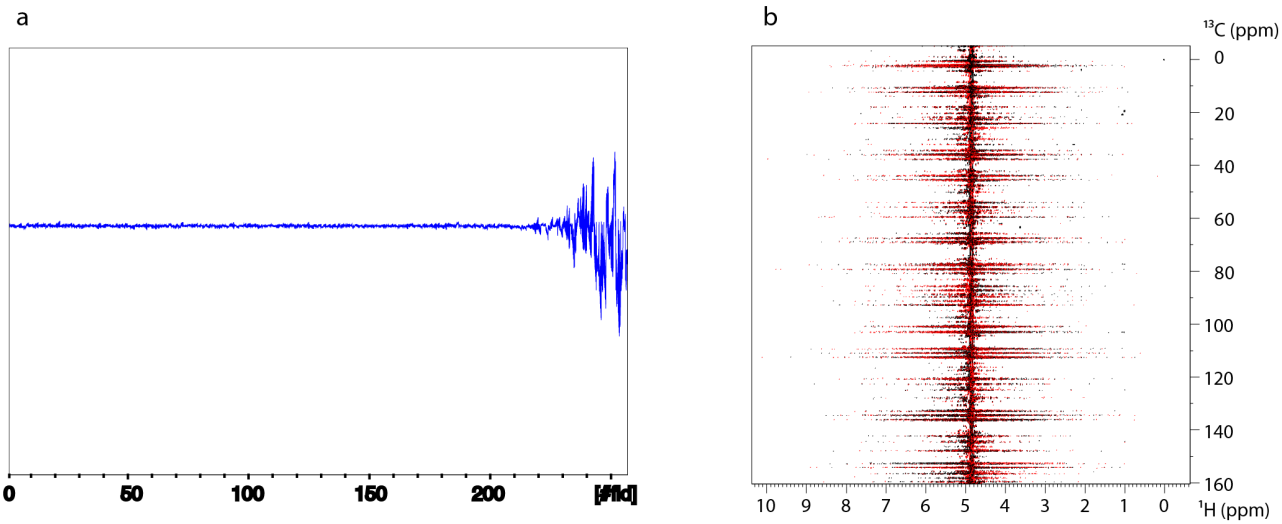


Figure 5.5. FID from a resolution-enhanced NUS HSQC with the pulse sequence "hsqcetgpsp" for the valine sample used in section 1.1 (a). 256 t_1 increments were recorded during acquisition, with the final number of t_1 increments after reconstruction of 2048 points. Gradients were already used at their maximum strength, and their length could not be increased. The t_1 delay for the final increment was 40.9 ms. The resulting spectrum after the use of Compressed Sensing (CS) for reconstruction of the indirect FID is shown in (b).

As explained in section 2.2.5 of part 1, for small molecules, sensitivity-enhanced (SE) HSQC allows an increase in the S/N_{RMS} up to a factor of $\sqrt{2}$, but at the price of artifacts.[287] But the block for SE HSQC has also been recognized as a module for water suppression,[288] which could be incorporated in 3D experiments for proteins. The pulse sequence chosen for this study was "hsqcetgpprsisp2.2", where all the 180° in the heteronuclear channel are replaced by adiabatic pulses. Presaturation was also present to decrease even further the residual water peak.

Another possibility was a variant of that sequence, "hsqcetgpprsisp2.3". This pulse sequence allows reducing substantially the intensity of the artifacts from SE HSQC, but we found out that the water peak was poorly suppressed with that variant. So we chose to use the pulse sequence with a better suppression of the water peak, even at the expense of the artifacts. Another reason is the fact that the artifacts in question are in anti-phase in the proton dimension, while the correlation peaks are in-phase, making the two kinds of peaks easily distinguishable.

For this study, TOCSY was preferred to DQF-COSY because TOCSY spectra are considered very useful for metabolomics.[159] This test also makes it possible to test the limits of NUS for spectra that have less sparsity than those usually used to test the effectiveness of NUS acquisition and reconstruction. In order to obtain a spectrum without anti-phase peaks from zero-quantum coherence, the Z-TOCSY variant with z-filter was employed[104] (figure 1.10). For water suppression, an excitation sculpting module was added after the spin-lock (figure 1.3). This choice has been made after testing several other water suppression schemes, including presaturation, WET, WATERGATE, and the combination of presaturation and WET. The excitation sculpting gave the best water suppression, especially for the suppression of "faraway" water, which once again may have been enhanced by the phase cycling of the TOCSY pulse sequence.

2.1.2. Samples

We chose to work on model mixtures for which the intensity of the less sensitive peaks was very close to the noise level ($S/N_{RMS} < 10$). Since the sensitivity of the two pulse sequences is different, two samples were used. These two samples were dilutions of one of the samples used in part 2 and 3 for the metabolomics studies from 1D and 2D NMR spectra. For TOCSY, the sample was diluted 16 times, while for HSQC, the sample was diluted 4 times.

In order to be even closer to the metabolic profile of serum samples, glucose was added in each sample. Two stock samples of 125 mM and 250 mM of glucose were prepared with the solution of H_2O/D_2O (90/10) in phosphate buffer (100 mM) with 0.5 mM of TSP and 4 mM of NaN3 described in section 1.1 of part 2. Then 25 μL of the 125 mM glucose sample were added in the sample for TOCSY, for a final concentration of 5 mM. Finally, 50 μL of the 250 mM glucose sample were added in the sample for HSQC, for a final concentration of 20 mM. The dilution of the other metabolites was taken into account and the final concentration for each metabolite in the two samples is shown in table III.

Glucose aside, for the HSQC samples, the concentrations range from high mM (lactate) to tens of μM, while for the TOCSY sample, the concentrations range from low mM to low μM. The pyroglutamate concentration is shown in place of the glutamine since the glutamine peaks are barely visible in the HSQC and TOCSY spectra, suggesting that the conversion is almost complete.

2.1.3. Sampling schedules

In most NUS studies, exponential sampling is the most used sampling schedule. One of the main reasons is that the probability of a given t_1 increments to be recorded during acquisition can be matched to the decay of the FID (itself caused by T_2*), and it has been argued that this matching is the cause for sensitivity improvements by NUS. However, it has also been shown that one of the main reasons for the creation of sampling artifacts in NUS is the presence of large holes in the sampling schedule, something exponential sampling is prone to do, especially at high compression rates. Alternate sampling schedules, such as Poisson-gap sampling,[206] have already been shown to give more faithful reconstructions of the US spectrum than exponential sampling.[32] For this reason, we wanted to verify if this comparison is still true with spectra of complex mixtures.

To test it, we recorded two HSQC spectra from the complex mixture described in section 2.1.3 and in table III. The two spectra were acquired with 2048 t_1 increments in the indirect dimension and 13% of NUS. Eight scans were acquired per t_1 increment and the inter-scan delay was 3 s for a total acquisition time of 1 h 50 min and 25 s. One of the sampling schedules was generated with a Poisson gap sampling function described in another publication,[206] while the other was generated with an exponential sampling used in Topspin 3.0, with an exponential decay of 1 s. Compared to the value of the maximum evolution time along the indirect dimension (40 ms at best), this exponential sampling is similar to a random sampling. The two FIDs were reconstructed with the CS algorithm after weighting the FID with a sine-bell function. After FT, chemical shifts were referenced with the TSP peak and an automatic baseline correction of order 2 was applied. The processed spectra are shown in figure 5.6.

Table III. Concentration of each metabolite in the samples for HSQC and TOCSY

Metabolite	Concentration of metabolites in the sample for HSQC (μM)	Concentration of metabolites in the sample for TOCSY (μM)
3-hydroxybutyrate	510	130
Acetate	250	60
Alanine	2630	660
Arginine	550	140
Asparagine	320	80
Aspartate	70	20
Betaine	160	40
Carnitine	190	50
Citrate	670	170
Creatine	160	40
Creatinine	300	80
Glucose	20000	5000
Glutamate	420	100
Glycerol	2450	610
Glycine	2080	520
Histidine	420	100
Isoleucine	300	80
Lactate	8840	2210
Leucine	440	110
Lysine	1150	290
Methanol	420	100
Methionine	120	30
Ornithine	250	60
Phenylalanine	320	80
Proline	1130	280
Pyroglutamate	1340	340
Pyruvate	50	10
Serine	760	190
Threonine	550	140
Tryptophan	320	80
Valine	1410	350

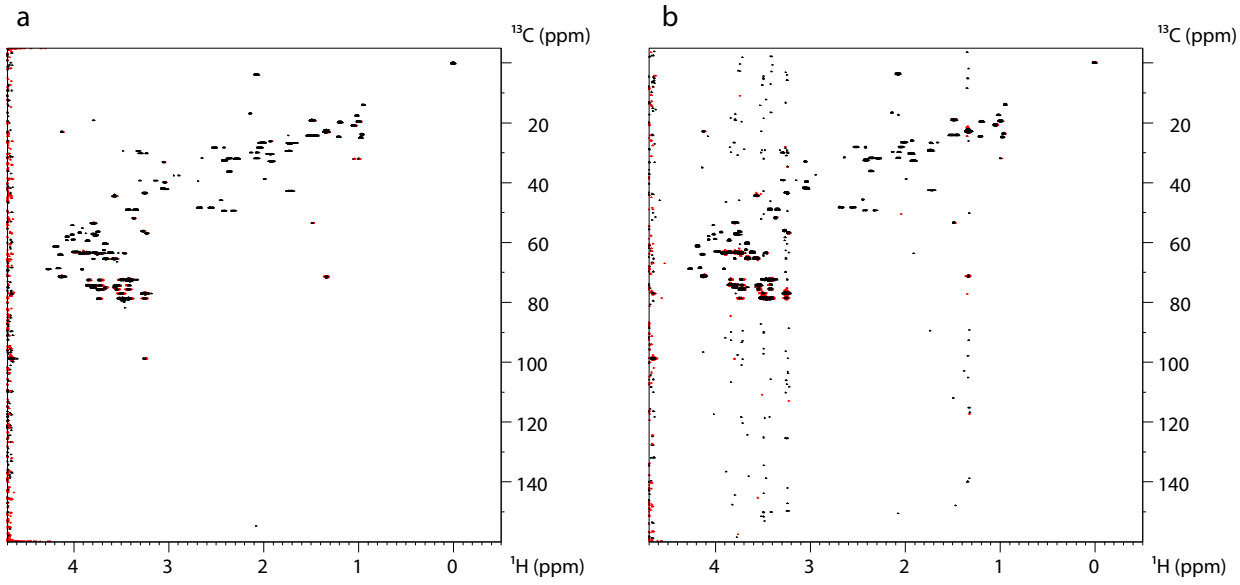


Figure 5.6. HSQC spectrum of the sample detailed in table III, with 1024 t_1 increments for the indirect dimension and 13% of NUS (266 t_1 increments recorded during acquisition) with Poisson gap sampling (a) or exponential sampling (b). The indirect FID was reconstructed with Compressed Sensing (CS). The two spectra are shown with the contour level close to the noise level.

In this figure, we can see the presence of ridges along the indirect dimension in the spectrum acquired with an exponential sampling, which are not present in the spectrum acquired with Poisson gap sampling. These ridges are essentially seen for the most intense peaks, from lactate and glucose, suggesting that their intensities are proportional to the peak intensity, not unlike t_1 noise. They were also present in the comparison between Poisson gap and exponential sampling in another study.[32] This confirms that Poisson gap sampling seems to lead to a more faithful reconstruction of the spectrum; therefore this sampling schedule was used during all the rest of the study.

2.1.4. Reconstruction algorithm

Within Topspin, two reconstruction algorithms are proposed: Multi-Dimensional Decomposition (MDD) and Compressed Sensing (CS). We decided to test the two algorithms for the reconstruction of the same data, which was the NUS HSQC spectrum recorded with a Poisson gap sampling schedule, as already described in the last section, and the processed spectrum is shown in figure 5.7 for the two reconstruction algorithms.

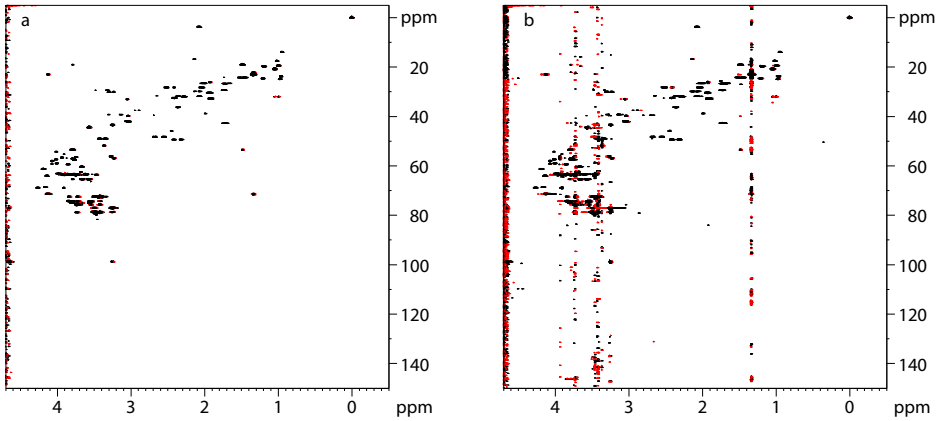


Figure 5.7. HSQC spectrum of the sample detailed in table III, with the same data as figure 5.6.a, recorded with CS (a) or MDD (b). The two spectra are shown with the contour level close to the noise level.

In this figure, the F1 ridges already seen in figure 5.6.b are also present, even more distinctly in the spectrum reconstructed with MDD. For this reason, the CS algorithm was kept for the remaining of the study.

2.2. Analysis of sensitivity in resolution-enhanced NUS

2.2.1. Parameters for acquisition and processing

The important parameters for acquisition used in both SE HSQC and z-filtered (zf) TOCSY are summarized in table IV.

Table IV. Main experimental parameters for acquisition of HSQC and TOCSY spectra

Parameters	Sequence	D1 (s)	SW2 (ppm)	SW1 (ppm)	o1p (ppm)	DS	NS	AQ (s)	Sampling schedule
HSQC	Sensitivity enhanced with presaturation	3	11	165.6	4.702	16	8	0.155	Poisson gap
TOCSY	DIPSI with excitation sculpting and z-filter	3	11	11	4.694	16	2	0.155	Poisson gap

D1: delay for relaxation, SW: spectral width, o1p: offset of the transmitter frequency, DS: number of dummy scans, NS: number of scans per t_1 increment, AQ: acquisition time in the direct dimension

Along with these parameters, others specific to SE HSQC or zf-TOCSY have to be mentioned. For SE HSQC, the INEPT delays were fixed to 1.72 ms, corresponding to an average

$^1J_{C-H}$ coupling constant of 145 Hz. Decoupling of $^1J_{C-H}$ during acquisition was done with the GARP supercycle and 2096 points were recorded during acquisition. The total experimental time was 1 h 50 min 25 s.

For zf TOCSY, a Decoupling In the Presence of Scalar Interactions 2 (DIPSI2) motif of 80 ms was used for the spin lock. A z-filter was added to suppress zero-quantum coherences, with a chirp pulse of 50 ms before the spin-lock and another chirp pulse of 30 ms after the spin-lock. Both chirp pulses had a 20 kHz bandwidth and were applied together with 2.3 G/cm gradients. Like SE HSQC, 2096 points were recorded during acquisition. The total experimental time was 58 min and 39 s.

To evaluate the impact of resolution-enhanced NUS for the two pulse sequences, different percentages of NUS were evaluated. These percentages are seen in table V, along with the corresponding compression rates and number of t_1 increments in the indirect dimension after the CS reconstruction for both pulse sequences. For each experiment, the sampling schedule was constructed using a program to generate Poisson gap sampling schedules develop recently,[289] allowing us to choose a sampling schedule with little to no chances to generate sampling artifacts or unfaithfully reconstruct the peaks volumes[290]. The sampling schedule with the best score from the program was chosen for the experiment and was placed as a variable counter list in the Topspin folder.

Table V. Compression rates used for NUS spectra and respective % of NUS (% of points acquired on the spectrometer), TD_1 and maximum evolution time for each spectrum.

Compression rate	1	2	3	4	5	8	16	32	50	100
% NUS	100	50	33	25	20	12.5	6.25	3.125	2	1
TD_1 HSQC	256	512	776	1024	1280	2048	4096	8192	12800	25600
TD_1 TOCSY	512	1024	1552	2048	2560	4096	n/a	n/a	n/a	n/a
Maximum evolution time TOCSY (ms)	38.7	77.4	117	155	194	298	n/a	n/a	n/a	n/a
Maximum evolution time HSQC (ms)	5.12	10.2	15.5	20.5	25.6	39.4	85.4	171	256	512

For SE HSQC, after acquisition, the indirect dimension was reconstructed with the CS algorithm, except for the US spectrum. Then the FID were weighted with a $\pi/2$ shifted squared sine-bell function in both dimensions and the direct FID was zero-filled to 4096 points, while the

indirect FID was zero-filled to 16384 points for all experiments, to evaluate fairly the volumes when the compression rate is changed. After FT, the chemical shifts were referenced to the peak of TSP and an automatic baseline correction with a polynomial function of order 2 was applied.

For zf TOCSY, the processing was similar to SE HSQC. After acquisition, the indirect dimension was reconstructed with the CS algorithm, except for the US spectrum. Then the FID were weighted with a $\pi/2$ shifted squared sine-bell function and zero-filled to 4096 points in both dimensions, to evaluate fairly the volumes when the compression rate is changed. After FT, the chemical shifts were referenced to the peak of TSP and an automatic baseline correction with a polynomial function of order 2 was applied.

2.2.2. Analysis of the HSQC and zf TOCSY spectra

2.2.2.1. Reconstruction fidelity and effect of resolution-enhanced NUS

The processed US spectra are shown in figure 5.8.a (HSQC) and figure 5.8.b (TOCSY). All the expected correlation peaks are present in both spectra. The water residual peak, while seemingly high on the 2D contour plot, has in reality a low contribution to the total intensity, as shown by the projection alongside the direct dimension. In these figures the location of the correlation peaks used for the sensitivity analysis are shown.

In the HSQC spectrum, the lactate peak is the most intense peak of the spectrum, the phenylalanine peak has a rather low intensity but the S/N_{RMS} is still enough to easily detect the peak in the spectrum. The isoleucine peak is one of the least intense peaks and is close to the noise level with this experimental time.

For the TOCSY spectrum, the lactate correlation peak is also the most intense correlation peak. The leucine and histidine peaks have a S/N_{RMS} high enough for a good repeatability, but the histidine peak has a simple multiplicity in the indirect dimension, while the leucine peak has a very complex multiplicity in the indirect dimension. The aspartate peak is one of the least intense correlation peaks with a simple multiplicity in the indirect dimension. The zoom in the aliphatic part, in figures 5.8.c for HSQC and 5.8.d for TOCSY allows seeing better the contrast in the resolution with the resolution-enhanced NUS, in figures 5.8.e for HSQC and 5.8.f for TOCSY.

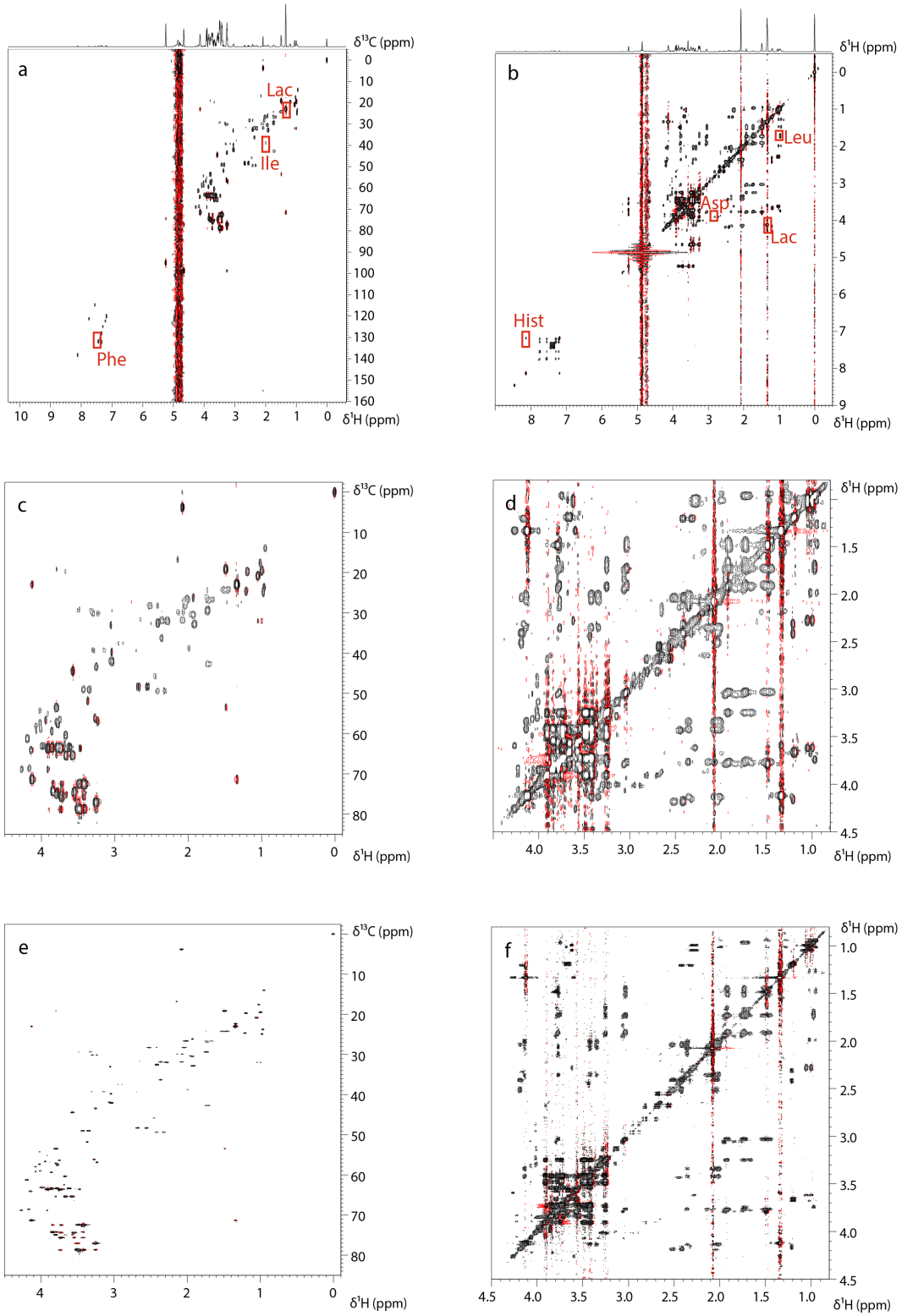


Figure 5.8. 2D NMR spectra of the mixture of metabolites described in table III a) Uniformly sampled HSQC, recorded in 1 h 50 min with 256 t_1 increments and b) Uniformly sampled TOCSY, recorded in 59 min with 512 t_1 increments. c) and d) show zooms on the aliphatic part of a) and b), respectively. e) and f) Non-uniformly sampled versions of c) and d), respectively, with a compression rate of 8 and a number of t_1 increments set to reach the same overall experimental time. Relevant experimental parameters are detailed in table 1. Abbreviations: Lac: Lactate; Ile: Isoleucine; Phe: Phenylalanine; Leu: Leucine; Asp: Aspartate; Hist: Histidine.

For SE-HSQC, we clearly see the gain in resolution, even with a compression rate of 8. This gain of resolution can allow the complete separation of peaks that are overlapped with usual acquisition parameters (see figure 2.8 for example). A better representation is shown in figure 5.9.a and 5.9.b for the lactate peak, whose resolution becomes sufficient to allow a complete separation between the lactate and threonine correlation peak. To show the ability of NUS to separate the peaks in HSQC, figure 5.10 demonstrates that it is possible to separate all the peaks from the region of the HSQC spectrum with the most overlap, where five correlation peaks are present.

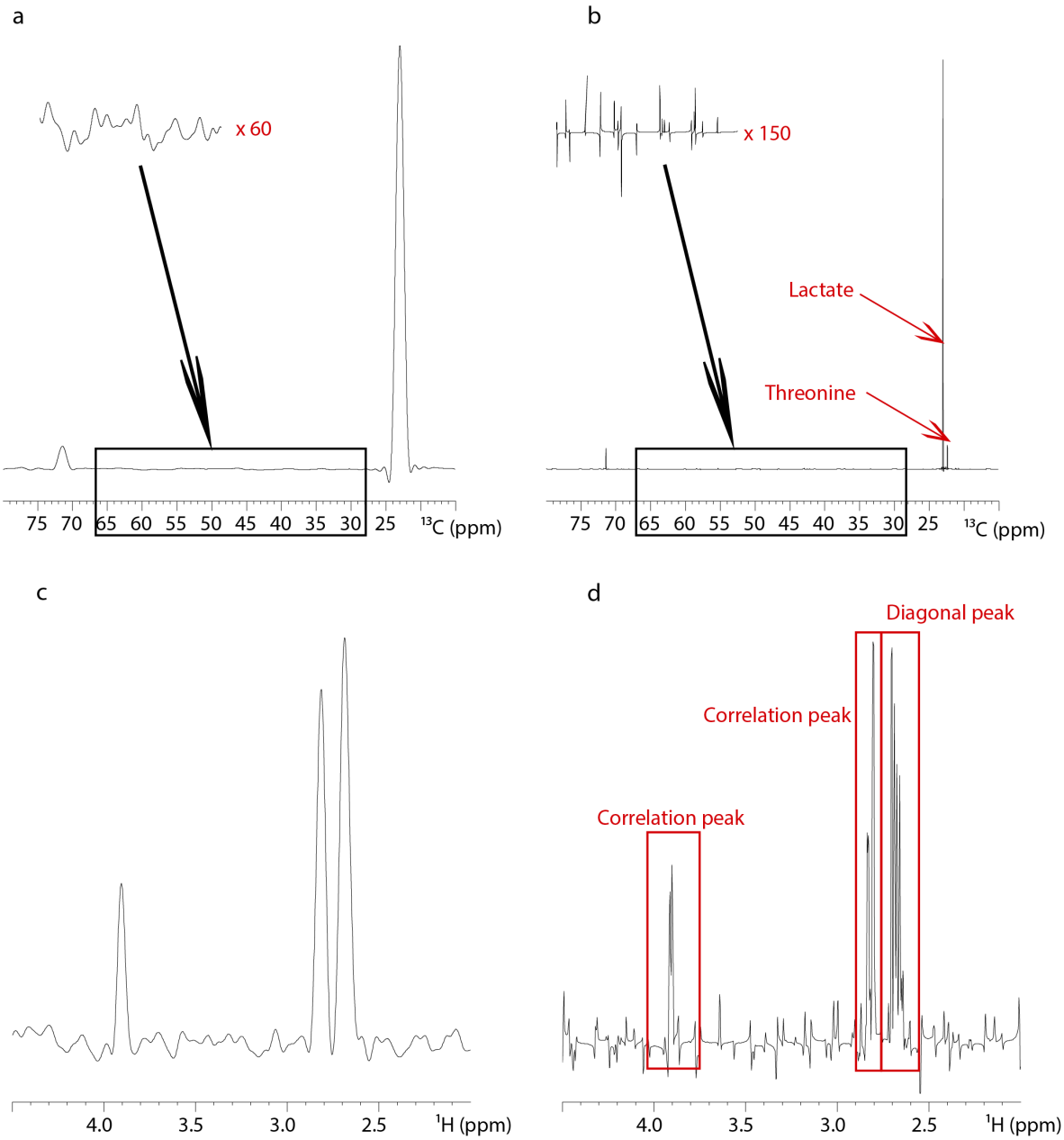


Figure 5.9. a and b: Slices along the indirect dimension of the lactate correlation peak in HSQC with a compression rate of 1 (a) or 32 (b), with a zoom on the noise level in both cases. The chemical shift along the direct dimension is 1.336 ppm. c and d: Slices along the indirect dimension of the aspartate correlation peak in HSQC with a compression rate of 1 (c) or 32 (d). The chemical shift along the direct dimension is 2.71 ppm.

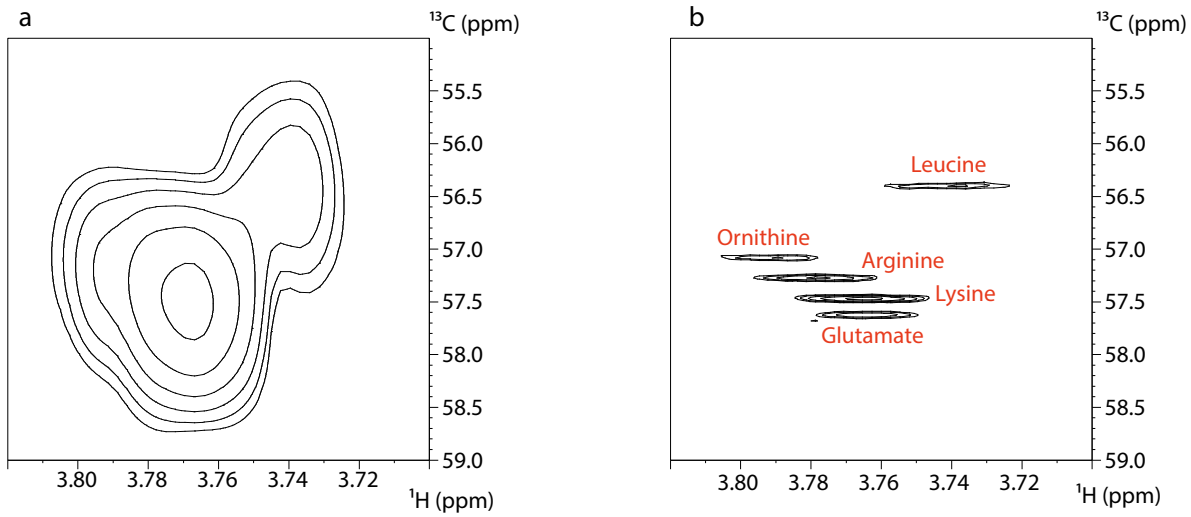


Figure 5.10. Contour plot of the HSQC spectrum, zoomed in the region with the highest level of overlap between peaks, with a compression rate of 1 (a) and 32 (b).

For TOCSY, the gain in effective resolution is less important, because of the peak splitting in multiplets in the indirect dimension. Nevertheless, a better separation of correlation peaks is obtained, and the multiplets in the indirect dimension become distinguishable, as shown in figures 5.9.c and 5.9.d. Since the linewidths in the indirect dimension were close to the linewidths in 1D ^{1}H spectra with a compression rate of 8, we did not use higher compression rates with this pulse sequence.

2.2.2.2. Analysis of volumes

With these changes in linewidths, we wanted to verify if the peak volumes were still correctly reconstructed in the NUS spectra. Since the number of scans was the same and the spectra had the same number of points in the indirect dimension, we expected that the volumes would not be changed by modifying the compression rate. So the volumes from each correlation peaks framed in red in figure 5.8.a and 5.8.b were extracted with the Topspin integration toolbox. The results are shown in figure 5.11.

Figure 5.11.a shows that for lactate, the peak volume remains constant with the compression rate. The volume for the phenylalanine peak remains close to the initial volume, with variations of less than 5% of the volume in the US spectrum, except at compression rates above 32, where the volume starts to regularly decrease. As for the isoleucine peak, the volume evolves considerably when the compression rate changes, which may be caused by its very low intensity in the US spectrum. Globally, it seems that the volume is not significantly modified by the use of NUS in this study; at least until the compression rate is above 32.

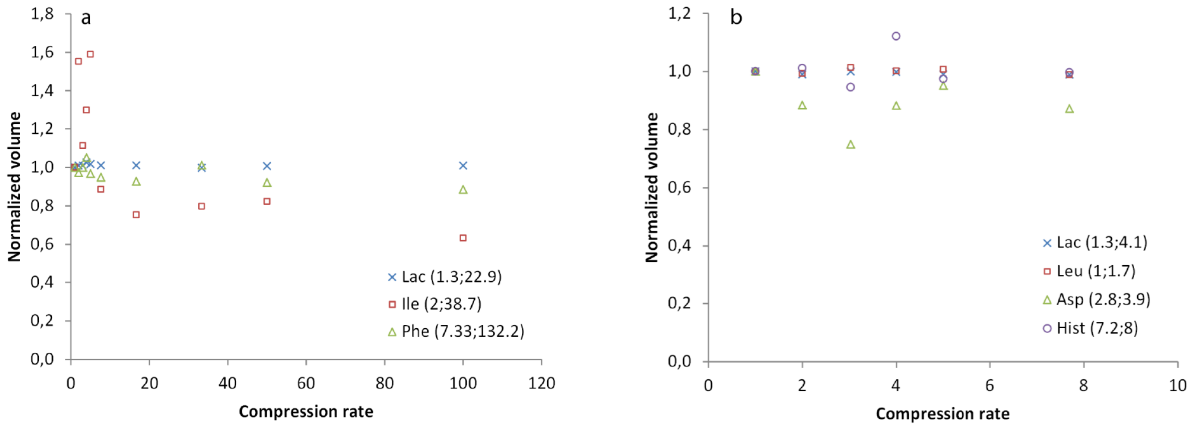


Figure 5.11. Evolution of 2D peak volumes in NUS spectra as a function of the compression rate for a) HSQC and b) TOCSY. The corresponding peaks are shown in figure 5.8.a and 5.8.b. All the experiments were recorded with identical experimental times (1 h 50 min for HSQC and 59 min for TOCSY) by adapting the number of t_1 increments. Normalized volumes were obtained by dividing the volume for a given compression rate by the volume in the US spectrum (compression rate of 1). For each metabolite, the numbers in parenthesis indicate the coordinates of the corresponding peak in (F_2; F_1).

For TOCSY, the lactate and leucine peak volumes are not modified by the changes in compression rate. The histidine and aspartate peak volumes seem to slightly change when the compression rate increases, but since these changes are not consistent when the compression rate increases, this observation may be linked to repeatability instead of reconstruction issues. In order to confirm these results, an analysis of the repeatability of peak volumes was performed with both pulse sequences, by recording the US spectrum 5 times along with 5 NUS spectra with a compression rate of 32 for the HSQC and a compression rate of 8 for the TOCSY. The volumes of the isoleucine and phenylalanine peaks were measured for the HSQC spectra, along with the volume of the histidine peak, the aspartate peak and the leucine peak for TOCSY, by using the integration tool within Topspin. For each series of spectra, the mean value and the standard deviation for all peaks were calculated. The results are shown in figure 5.12.

For the 2 HSQC peaks evaluated, in figure 5.12.a, we can see a reduction of the mean volume for each peak, with the inclusion of the standard deviation, we can conclude that this reduction is either non-significant (phenylalanine, at around 5% of coefficient of variation (CV)) or only slightly different (isoleucine, between 10 and 15% of CV). One interesting fact is that the standard deviation is seemingly reduced with resolution-enhanced NUS compared to US. At the very least, the repeatability is similar between the two approaches. A similar conclusion can be drawn for the peaks of the TOCSY spectra, in figure 5.12.b. We can now focus on the noise analysis within each approach.

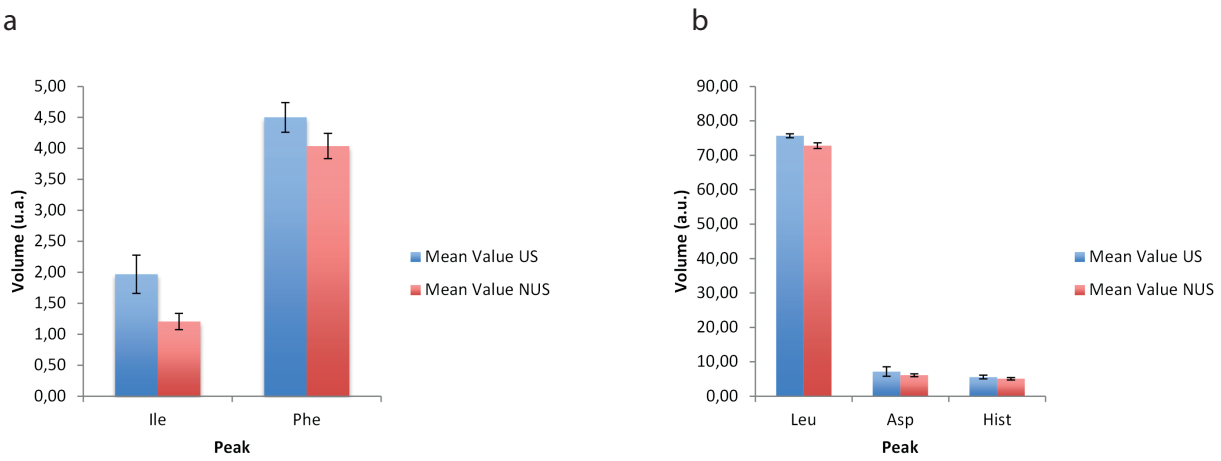


Figure 5.12. Mean value of peaks volumes for several peaks with 5 US spectra and 5 NUS spectra for HSQC (a) and TOCSY (b). For the NUS spectra, the compression rate was 32 for HSQC spectra and 8 for the TOCSY spectra. The error bars represent the standard deviation for each peak. The identity of each correlation peak is shown in figure 5.8.

2.2.2.3. Analysis of noise

From the analysis of figure 5.10, it can be seen that the noise behaves differently between NUS and US spectra, which is consistent with the results in figure 5.2. This behavior seems then to depend on the percentage of NUS alone. To confirm the hypothesis, a study on the behavior of noise in NUS spectra was conducted, on a sample with only D_2O inside. The HSQC spectra were acquired, similarly to the spectra in section 2.2.1, with all the powers for pulses in ^{1}H and ^{13}C channels reduced to 0 W, in order to avoid obtaining any magnetization in the transverse plane. The resulted spectra had no correlation peaks, only noise. It was then possible to see the behavior of noise alone with the compression rate.

The spectra were then transferred to MATLAB, where the intensity from all points of the matrix representing the spectrum were analyzed, and the intensities were grouped in a histogram shown in figure 5.13 for six different compression rates. In this figure, we can see that noise has a Gaussian profile for the US spectrum, as expected. But for the NUS spectra, the noise profile is no longer Gaussian and differs more and more from the Gaussian profile when the compression rate increases. The noise has more probability to have an intensity close or equal to zero, but the few zones non-equal to zero have a higher intensity. This analysis of noise brings similar results to the analysis of noise in sensitivity-enhanced NUS.[283] Because of this change in the noise behavior, the S/N_{RMS} appears no longer reliable. To find a new way to measure sensitivity, a different parameter from noise have to be used for the calculation of S/N.

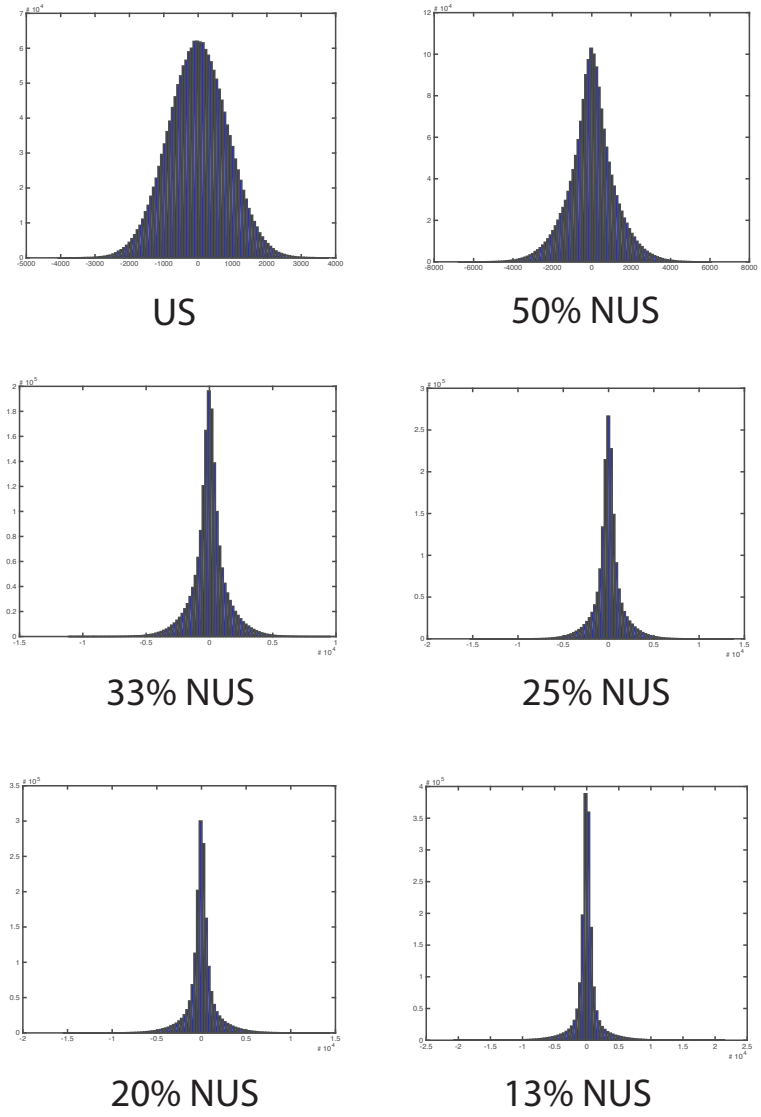


Figure 5.13. Evolution of the noise profile with the percentage of NUS from HSQC spectra with no peaks. The spectra were acquired in a D_2O sample with all powers from pulses in ^{1}H and ^{13}C reduced to 0 W. The corresponding compression rates are summarized in table V.

2.2.3. Analysis of peak intensity and S/N in resolution-enhanced NUS

2.2.3.1. Parameters for analysis

In order to analyze intensities and noise in the spectra of the complex mixtures, all the HSQC and TOCSY spectra were transferred into MATLAB. Within MATLAB, sections of the matrix were defined, representing each correlation peak in HSQC and TOCSY, as well as a region with only noise in each spectrum. For the peak regions, the maximum intensity was taken, while for the noise regions, the RMS and the maximum intensity were extracted. From these parameters, two different S/N were calculated:

-The S/N_{RMS}, defined as:

$$S/N_{RMS} = \frac{\text{Signal intensity}}{\text{Noise RMS}}$$

-The signal-to-maximum intensity of noise (S/N_{max}), defined as:

$$S/N_{max} = \frac{\text{Signal intensity}}{\text{Maximum intensity of noise}}$$

As a side note, the S/N_{RMS} calculated in this section is different from the S/N_{RMS} calculated by the Topspin S/N toolbox. Within Topspin, the S/N_{RMS} is defined as:

$$S/N_{RMS}\ (\text{Topspin}) = \frac{\text{Signal intensity}}{2 * \text{Noise RMS}}$$

where Noise RMS is corrected depending on the noise profile. So the S/N_{RMS} calculated in the next sections will be roughly twice the S/N_{RMS} that would have been calculated within Topspin. For each parameter, a normalized value was used by dividing each value by the value in the US spectrum (compression rate of 1). For the two pulse sequences, the noise, the intensity, the S/N_{RMS} and the S/N_{max} will be discussed.

2.2.3.2. HSQC

The results for peaks from HSQC are summarized in figure 5.14. In figure 5.14.a, with the evolution of noise with the compression rate, we have a further proof than the noise has a different profile in NUS spectra compared to US spectra, since the maximum intensity of noise evolves linearly with the compression rate (R^2>0.99 for the linear regression), while the RMS noise increases as $X^{0.35}$, where X is the compression rate. This evolution is not expected from an acquisition viewpoint, since the number of scans and t_1 increments acquired were the same for each spectrum. This remark shows that we still do not know exactly the behavior of noise in NUS spectra and that further improvements in the quality of NUS spectra will probably need a better understanding of this behavior.

For the intensity, in figure 5.14.b, the evolution is linear with the compression rate for the lactate peak. This evolution is also the case for the phenylalanine and isoleucine peaks for compression rates up until 32. For compression rates above 32, the intensity increase is less important.

This leads to the evolution of S/N. The S/N_{RMS}, in figure 5.14.c, evolves as $X^{0.561}$, where X is the compression rate. However, the S/N_{max}, in figure 5.14.d, does not seem to evolve with the compression rate, and remains near constant for all spectra. This analysis is close to what we see in the spectra, in both resolution-enhanced and sensitivity-enhanced NUS (see figure 5.2). With the fact that noise is no longer Gaussian in NUS spectra, this suggests that the S/N_{max} describes more accurately the evolution of sensitivity in NUS spectra than S/N_{RMS}. In the end, even the isoleucine peak was visible until a compression rate of 50, despite being barely above the noise level, even in the US spectrum.

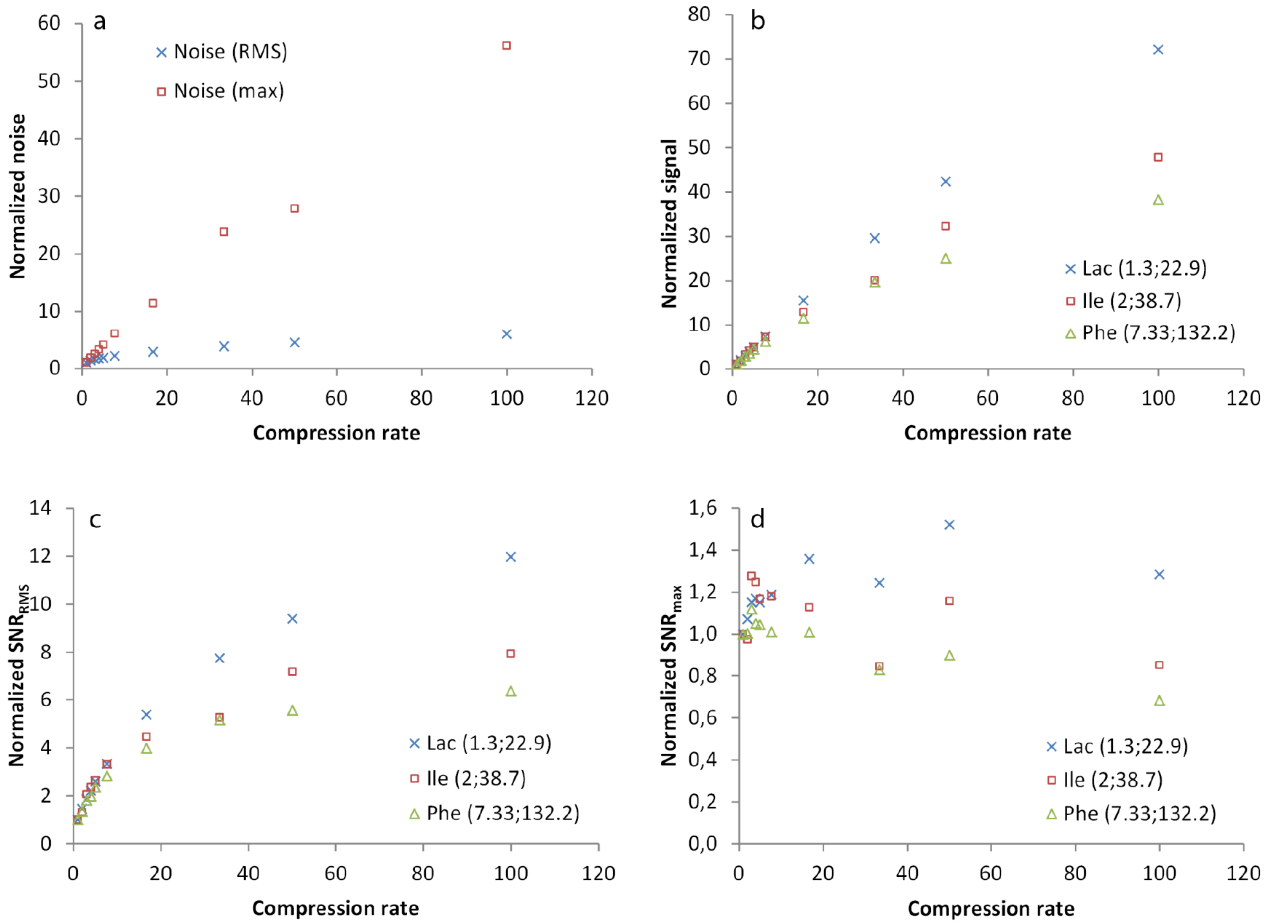


Figure 5.14. Evolution of a) signal, b) noise, c) SNR_{RMS} and d) SNR_{max} for HSQC peaks as a function of the compression rate. All the experiments were recorded with identical times by adapting the number of t_1 increments. Normalized signal, SNR_{RMS} and SNR_{max} were obtained by dividing the parameter for a given compression rate by the same parameter in the US spectrum. For each metabolite, the numbers in parenthesis indicate the coordinates of the corresponding peak in (F_2; F_1)

To summarize, with resolution-enhanced NUS, we can obtain HSQC spectra with higher resolution without significant losses in sensitivity for compression rates up to 32. The compression rates obtained were much higher than in sensitivity-enhanced NUS, showing the importance of the number of t_1 increments recorded during acquisition.

2.2.3.3. zf TOCSY

The results from zf TOCSY are summarized in figure 5.14. We can see from the evolution of noise in figure 5.15.a that the noise behavior in NUS TOCSY is similar to that in NUS HSQC, confirming the visual impression from figure 5.9. But for the evolution of peak intensities, in figure 5.15.b, we can see a peak-dependent behavior. For example, the histidine correlation peak has a linear evolution similar to the peaks in HSQC, but for the leucine peak, the intensity remains nearly constant for all spectra. This behavior of intensity is probably linked to the possibility to see the multiplet in the indirect dimension for a given correlation peak. This may explain the peak-dependent behavior, since the histidine correlation peak has a J_{H-H} too small to be resolved, even in 1D ^{1}H spectrum, while the leucine correlation peak has a complex pattern.

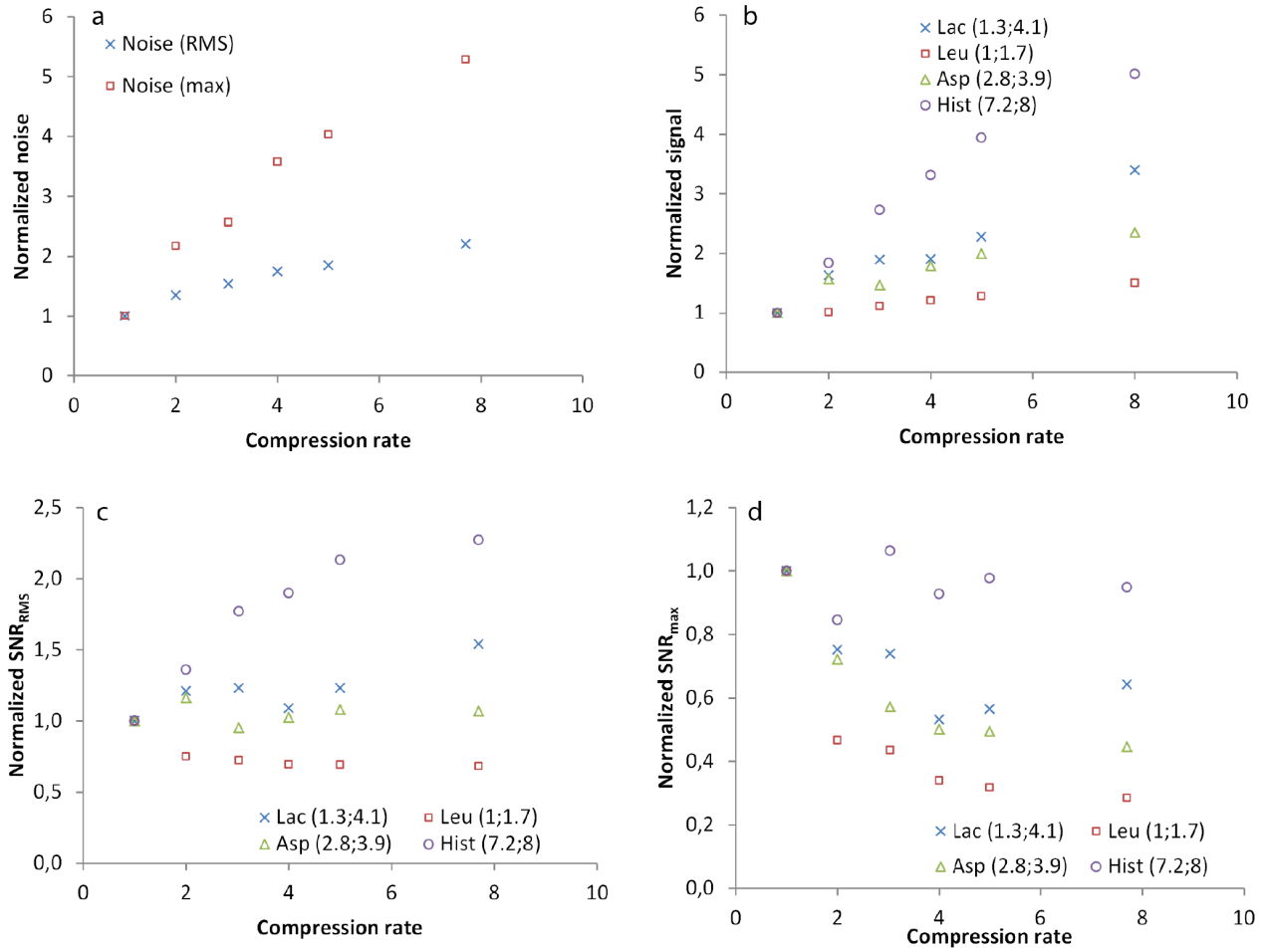


Figure 5.15. Evolution of a) signal, b) noise, c) SNR_{RMS} and d) SNR_{max} for TOCSY peaks as a function of the compression rate. All the experiments were recorded with identical times by adapting the number of t_1 increments. Normalized signal, SNR_{RMS} and SNR_{max} were obtained by dividing the parameter for a given compression rate by the same parameter in the US spectrum. For each metabolite, the numbers in parenthesis indicate the coordinates of the corresponding peak in (F_2; F_1)

This change in the intensity behavior impacts the S/N behavior with the evolution of the compression rate, as it also becomes peak-dependent. In figure 5.15.c, the S/N_{RMS} increases for the histidine correlation peak similarly to the peaks in HSQC, while it decreases for the leucine peak. The slight increase in intensity is not enough to compensate the increase in the noise level. For the aspartate and lactate peaks, the behavior changes with the compression rate, probably reflecting the progressive splitting of the multiplets. But in figure 5.15.d, the S/N_{max} decreases with the compression rate for most correlation peaks. The only exception is the histidine peak, probably because there is no peak splitting in the indirect dimension.

To summarize, while the noise behavior in resolution-enhanced NUS TOCSY is similar to NUS HSQC, the peak splitting in the indirect dimension causes decreases in sensitivity depending on the complexity of the multiplet pattern. On the other hand, seeing the multiplet pattern could be an additional source of information for the identification of the correlation peak.

In conclusion, while sensitivity-enhanced NUS does not provide a significantly better sensitivity for small molecule analysis, resolution-enhanced NUS provides a possibility to increase the resolution in the indirect dimension in HSQC without increasing the experimental time or decreasing sensitivity. For TOCSY, while sensitivity decreases, multiplets become resolved in the indirect dimension, giving another parameter to increase the probability of identification of a metabolite.

This increase in resolution could be compared to those obtained by the ASAP HSQC pulse sequence, which has been developed and published during our study.[176] It should be noted that the comparison should explore the possibility to integrate variants of the HSQC pulse sequence, like SE HSQC, as used in this study, or pure-shift HSQC, as developed recently,[291] which could give further improvements in sensitivity and resolution in the proton dimension. Even more interestingly, the two approaches could be combined, to eventually obtain the same resolution in a lower experimental time.

The approach can also benefit from further improvements in the reconstruction algorithm, like the Low-Rank reconstruction presented recently,[292] which could give even more faithful peak reconstruction than CS, given the results presented in this publication. The use of direct covariance may be also tested for the acquisition of spectra with very high resolution but a low value of the maximum evolution time. An analysis of noise behavior could also be performed to verify if that behavior is changed depending on the algorithm used for reconstruction.

For TOCSY, another possible study could be the ^{1}H-^{1}H constant-time TOCSY,[253] which would resolve the problem of peak splitting. But for high acquisition times in the indirect dimension, sensitivity could be reduced by relaxation, and the peak intensity will become dependent on the maximum acquisition time in the indirect dimension. It is also possible to use the optimization of NUS parameters (for acquisition and processing) from this chapter to improve the results from previous studies, including the one from part 3.[181, 267]

The next part will be about the T_1 and T_2 calculation in complex mixtures, as explained in part 3.

Part 6. Estimation of T_1 and T_2 in complex mixtures

Goal: Adapting and evaluating software and pulse sequences for the estimation of T_1 and T_2 of a range of peaks in a complex mixture in highly protonated water, in order to set up accordingly important parameters for 1D and 2D spectra, like the inter-scan delay.

1. The T_1 toolbox

Measurements of T_1 are scarcely used for complex mixture analysis, despite their importance to set up the inter-scan delay for optimal signal-to noise ratio per unit of time. One of the possible reasons is the lack of tools for an automated measurement of T_1 values for each chemical shift, which may itself be due to the high variability of T_1 values, depending on the magnetic field, the pH, or the chemical interactions. The few T_1 pseudo-2D plots published, as far as we know, have been limited to the field of solid-state NMR,[293, 294] probably because the T_1 in these systems are molecule-dependent, rather than peak-dependent.[293] However, another coefficient is often measured by NMR: the diffusion coefficient, with the Diffusion Ordered Spectroscopy (DOSY) pulse sequence, which is also molecule-dependent rather than peak-dependent. Several similarities exist between the analysis of DOSY spectra and T_1 calculation:

- The analysis involves the acquisition of several spectra whose peak intensity decrease is linked to the parameter measured.
- Estimating the parameter in question involves the fitting of experimental curves to an exponential decay. A common tool for this estimation is the Inverse Laplace Transform (ILT).
- Obtaining accurate and reproducible results requires clean spectra, since T_1 and T_2 measurement are sensitive to modifications in baseline or phase errors between different values.

For these reasons, we decided to adapt a toolbox initially used for the processing of DOSY spectra, to allow T_1 measurements. An added advantage would be the representation of T_1 in a pseudo-2D plot, in order to see directly the range of T_1 in a complex mixture. We chose a DOSY Toolbox written within MATLAB[295] to easily change the mathematical formula for the calculation of T_1 parameters. The needed changes in mathematical formula will be described first, and then the results for a pure compound and a complex mixture will be detailed.

1.1. The DOSY Toolbox

Before introducing the T_1 Toolbox, it is necessary to show how the DOSY Toolbox works. When opening the software, the GUI shown in figure 6.1 can be used to process data from commercial software. After selecting the data generated from a DOSY pulse sequence, this GUI can be used to do the processing that could also be performed in commercial software like Topspin. The phase correction is done in a similar manner to Topspin; zero-filling and window functions can be applied, along with calibrating the chemical shifts from a reference peak of known chemical shift. It is also possible to do an automated baseline correction with a polynomial function.

When a phased spectrum with flat baseline is obtained, it is possible to process the data in the diffusion dimension by ILT, with the possibility to choose only a part of the spectrum for this processing. This processing is done point per point in the 1D spectrum, so it is possible to choose an intensity limit for processing (therefore avoiding to fit in a noise region). Different strategies can be chosen for the processing, either monovariate or multivariate,[295] but our focus was the direct fitting for ILT.

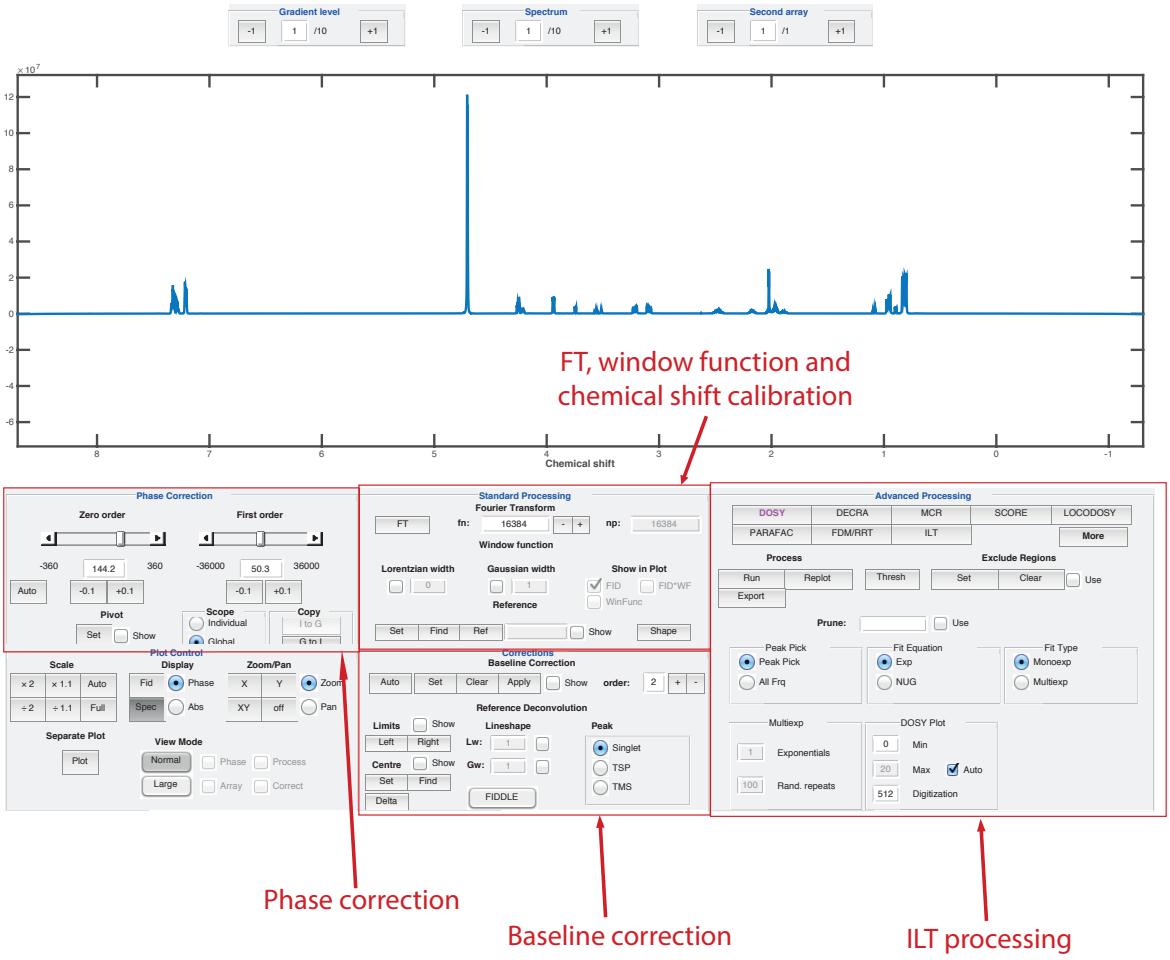


Figure 6.1. GUI of the DOSY Toolbox for the processing of 1D sub-spectra from DOSY data.

With this processing, the diffusion curve is fitted with an exponential decay, which is described by the Stejskal-Tanner equation[296]:

$$M(g, D) = M_0 e^{-D\gamma^2\delta^2 g^2 \Delta'} \qquad (6.1)$$

where M_0 is the equilibrium magnetization, D the diffusion coefficient, γ the gyromagnetic ratio, δ the gradient delay, g the gradient strength and Δ' the delay during which the magnetization is defocused. The wanted parameter is D, which is linked mainly to the weight and the shape of the molecule, and the 1D sub-spectra are obtained by varying the gradient strength.

In order to fit correctly the experimental data, the equation (6.1) is necessary, along with its derivative along D, which is:

$$\frac{dM}{dD} = -\gamma^2\delta^2 g^2 \Delta' \, x \, M_0 e^{-D\gamma^2\delta^2 g^2 \Delta'} \quad (6.2)$$

Once the data fitted, the diffusion coefficient has to be extracted from the equation of the fitted exponential. From equation 6.1, we can see that:

$$\frac{M(g, D)}{M_0} = e^{-D\gamma^2\delta^2 g^2 \Delta'}$$

By using the natural logarithm, we obtain:

$$\ln \frac{M(g, D)}{M_0} = -D\gamma^2\delta^2 g^2 \Delta'$$

It becomes therefore possible to isolate D with the equation:

$$D = \frac{-\ln \frac{M(g,D)}{M_0}}{\gamma^2\delta^2 g^2 \Delta'} \quad (6.3)$$

Once the coefficients extracted from each fit, a new GUI opens, containing the pseudo-2D spectrum, as seen in figure 6.2.

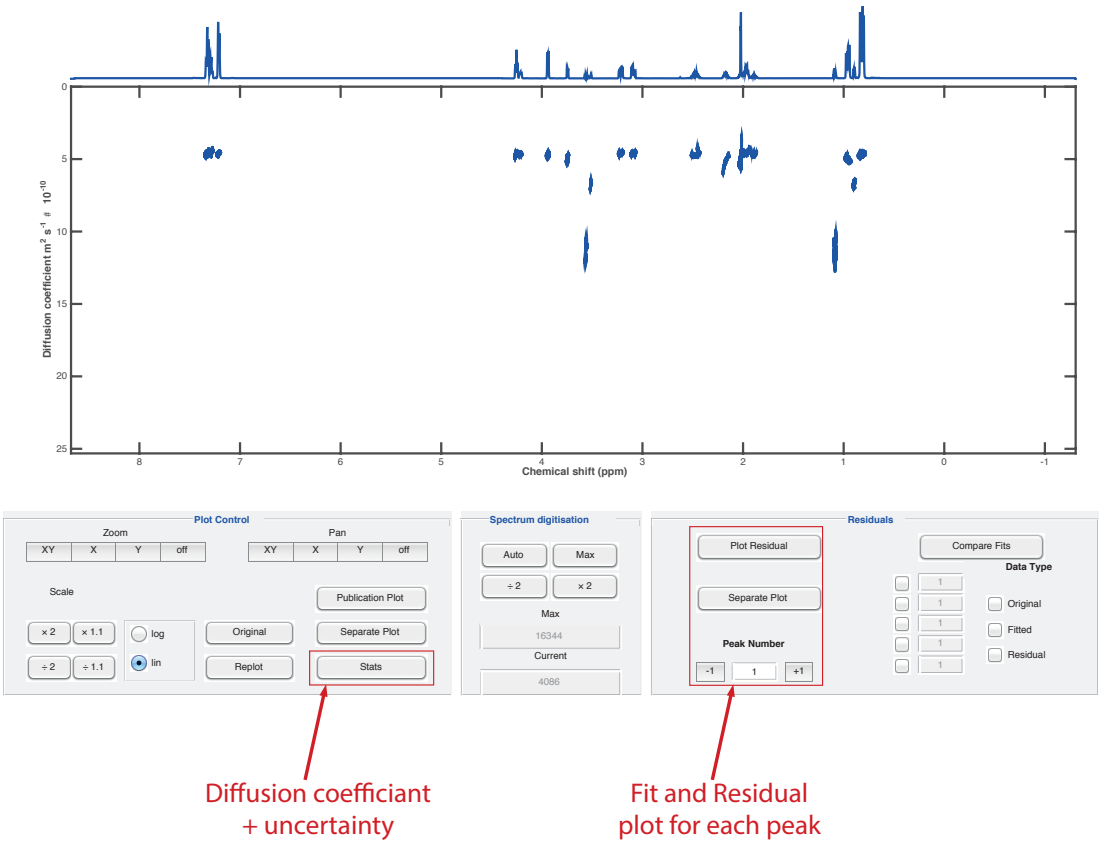


Figure 6.2. GUI of the DOSY Toolbox for the display of the pseudo-2D plot.

Since D is molecule-dependent, each row in the diffusion dimension works as a 1D spectrum from one molecule, provided that no overlap between peaks is present in the 1D spectrum of the mixture. For each peak, the value of the diffusion coefficient can be retrieved from the 'Stats' button in the second GUI, along with the uncertainty from the T_1 measurement. In the case of overlap, the apparent diffusion coefficient becomes a mean value of the different diffusion coefficient for each peak.[297] In this case, fitting with multi-exponentials is a possibility within the DOSY Toolbox.

For each peak, it is possible to verify the accuracy of the fit by looking at the residual plot, as seen in figure 6.3.

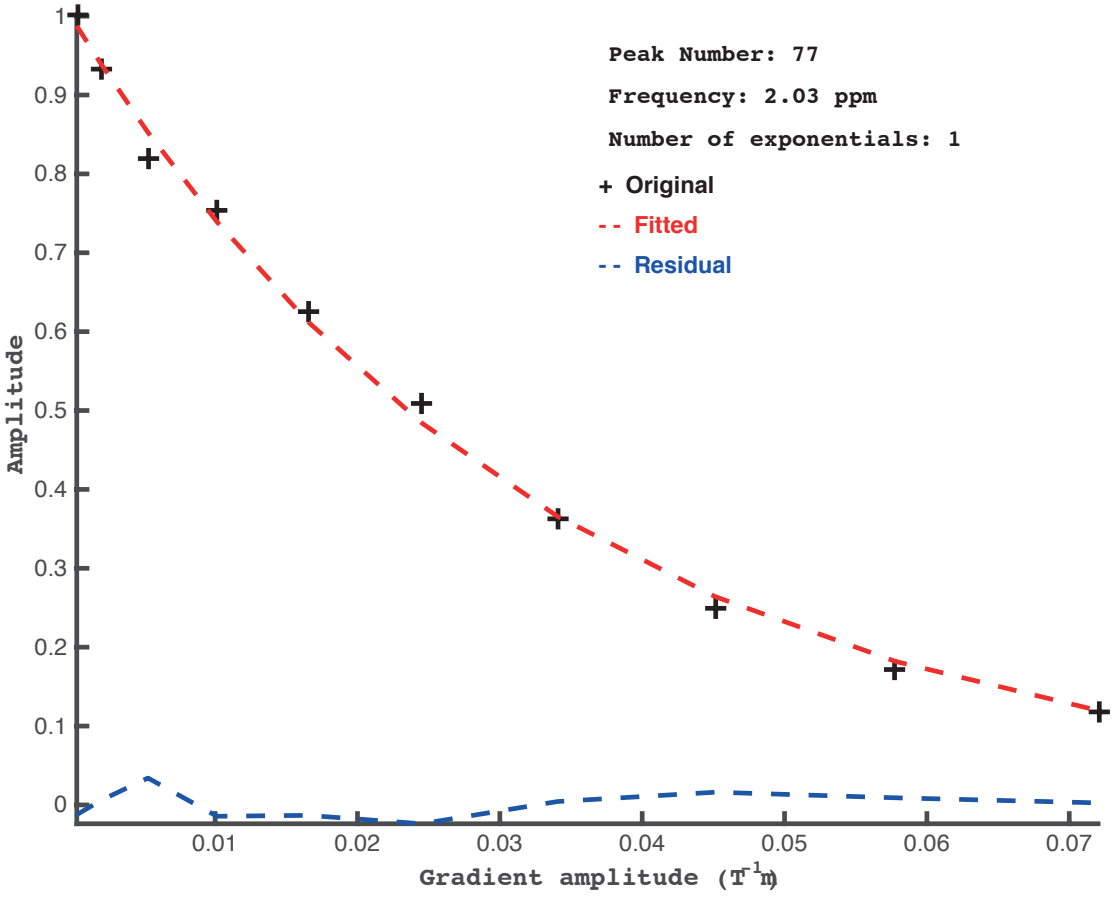


Figure 6.3. Experimental data for the diffusion coefficient from one peak, along with the fitted exponential and the residual plot

In this plot the experimental points from the diffusion dimension are shown, along with the fitted exponential from the Stejskal-Tanner equation and the residual plot. Along with the accuracy of the fit, it allows to verify any possible outlier that may decreases the accuracy. If it is the case, it would be possible to do another fit with the possibility to exclude one or several points from the fit.

1.2. Equations for T_1 measurements

To use the DOSY Toolbox for T_1 estimation, the changes needed are the parameter conversion from gradient intensity to the variable relaxation delay, the exponential function and its derivative in order to fit the experimental curve and the conversion from exponential decay to the parameter T_1 to draw the pseudo-2D plot.

For the relaxation delays, the exponential function, derived from Bloch equations,[298] is:

$$M(t, T_1) = M_0\left(1 - 2e^{-\frac{t}{T_1}}\right) \qquad (6.4)$$

where M_0 is the equilibrium magnetization. The derivative of the researched parameter is:

$$\frac{dM}{dT_1} = 2M_0\frac{t}{T_1^2}e^{-\frac{t}{T_1}} \qquad (6.5)$$

Finally, the parameter T_1 can be extracted, by following a similar reasoning for equation 6.3 from equation 6.1 with the equation:

$$T_1 = -\frac{t}{\ln\left(\frac{1-\frac{M(t,T_1)}{M_0}}{2}\right)} \qquad (6.6)$$

These changes, along with cosmetic changes in the Graphical User Interface (GUI) to adapt to the new parameter evaluated, were integrated into the DOSY Toolbox. The modified DOSY Toolbox will be called T_1 Toolbox in the next sections.

1.3. First test: phenylalanine sample

In order to test the capacity to measure the T_1 from IR data, a first test was carried out with a sample with no overlapping peaks, in order to compare the estimated T_1 from the T_1 Toolbox to the estimated T_1 from the T_1/T_2 calculation module from Topspin. The same data that have been recorded for Figure 3.6 have been used for the test of the T_1 Toolbox. For this spectrum, 12288 points were recorded during 1 s of acquisition, with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 50 s, in order to let all the protons relax completely between peaks, as seen in the next figure. Twelve different delays between the 90° and the 180° pulses were used, in seconds: 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 4, 8, 15, 30, 40, 45 and 50.

The pseudo-2D plot is presented in figure 6.4.a, along with an example of a fitted curve and a residual plot. The residual plot, in figure 6.4.b, shows that the exponential fit is very close to the experimental data, giving a residual plot very close to zero.

To verify if the T_1 estimation is correctly extracted from the exponential decay, the values from the pseudo-2D plot are compared to the values given by the T_1 module in Topspin. The results are shown in figure 6.5 and show very similar estimation between the two programs. Although the T_1 values are slightly lower with the T_1 Toolbox, the difference is always less than 0.2 s. However, the uncertainty in the measurement of T_1 is not presented for reasons shown in the next section.

Nevertheless, we can conclude that the modified T_1 Toolbox seems to give accurate T_1 estimations from the IR data, at least for simple samples. This Toolbox can now be tested for a complex mixture.

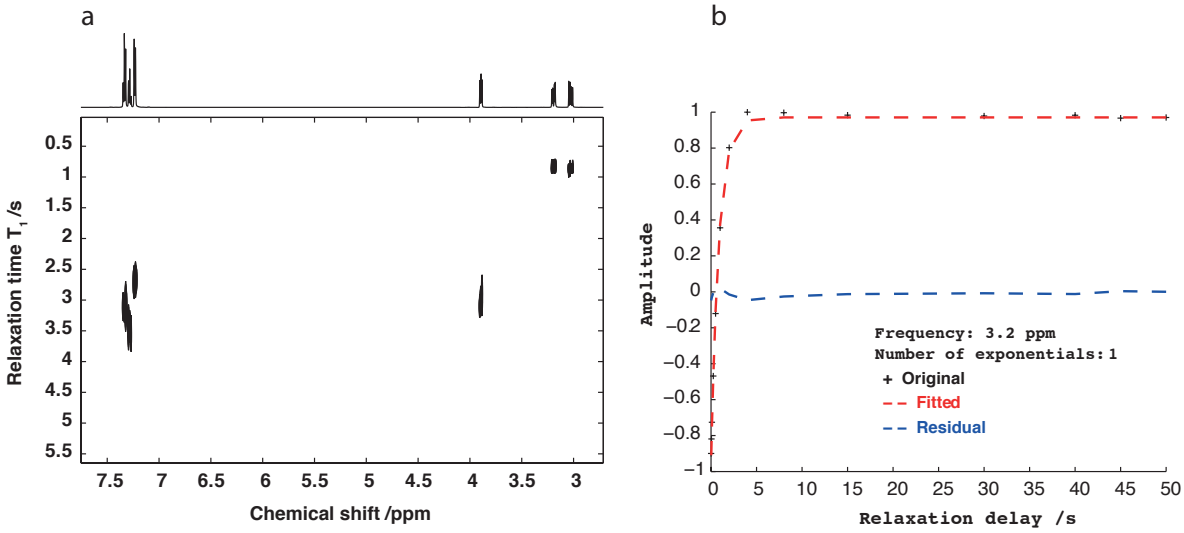


Figure 6.4. Pseudo-2D plot of the inversion recovery data with the phenylalanine sample in D_2O (a), with the fit and the residual plot (a) from one peak from the CH_2 group at 3.2 ppm (b).

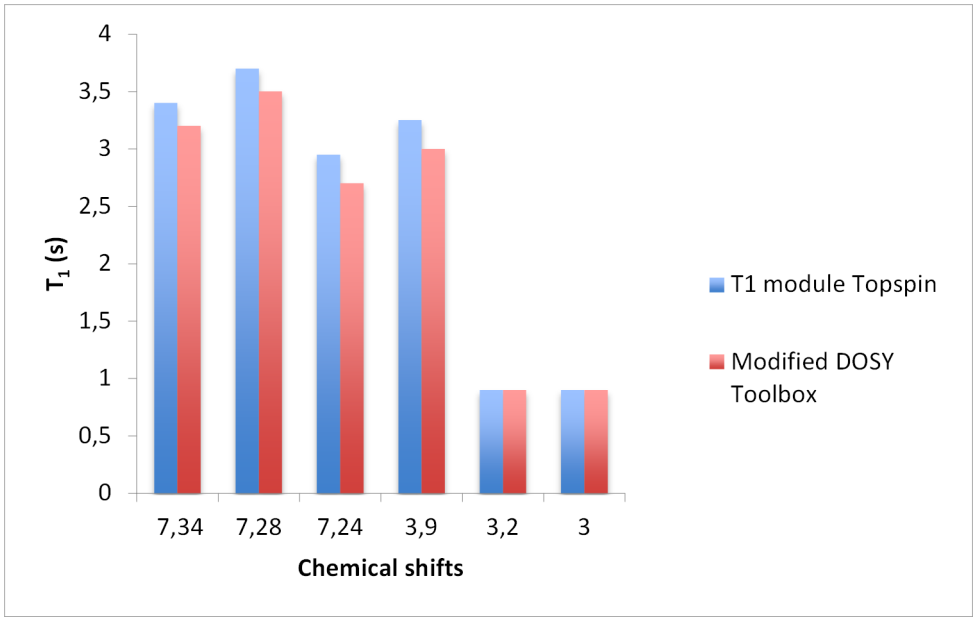


Figure 6.5. Histogram of estimated T_1 from the T_1 module of Topspin and the pseudo-2D plot from the T_1 Toolbox.

1.4. The T_1 Toolbox for T_1 estimation of a complex mixture

For this test, the IR pulse sequence was used for a complex mixture similar to the samples used in part 2 and 3, but more concentrated (roughly 5 times more concentrated) and in D_2O instead of a buffered H_2O/D_2O (90/10) solvent. The parameters for the IR pulse sequence were the same as in section 1.3.1 of part 3. Briefly, 12288 points were recorded during 1 s of acquisition,

with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 50 s, in order to let all the protons relax completely between peaks, as seen in the next figure. Twelve different delays between the 90° and the 180° pulses were used, in seconds: 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 4, 8, 15, 30, 40, 45 and 50.

The analysis in the T_1 Toolbox was similar to the IR data of the phenylalanine sample. First, the 1D spectra were phase corrected, referenced to the -CH_3 peak of lactate (1.33 ppm) and automatically baseline corrected. Then the regions with peaks from metabolites were selected for ILT with the equations from section 1.1. The pseudo-2D plot is shown in figure 6.6, along with a residual plot from the alanine peak at 1.46 ppm and another from a glutamine peak at 2.44 ppm.

The pseudo-2D plot shows two interesting trends: there is a rather high dynamic range of T_1 in the mixture, from a few hundreds of ms to around 10 s, and the higher the T_1, the higher the uncertainty in the T_1 value, which is shown by a larger peak in the indirect dimension. For a low T_1 from peaks with high SNR, for example the CH_3 peak from alanine at 1.46 ppm, the exponential decay fit very well the experimental data. However, for peaks with high T_1 and/or low SNR, like the glutamine peak at 2.44 ppm in figure 6.6.c, the fit is much less accurate.

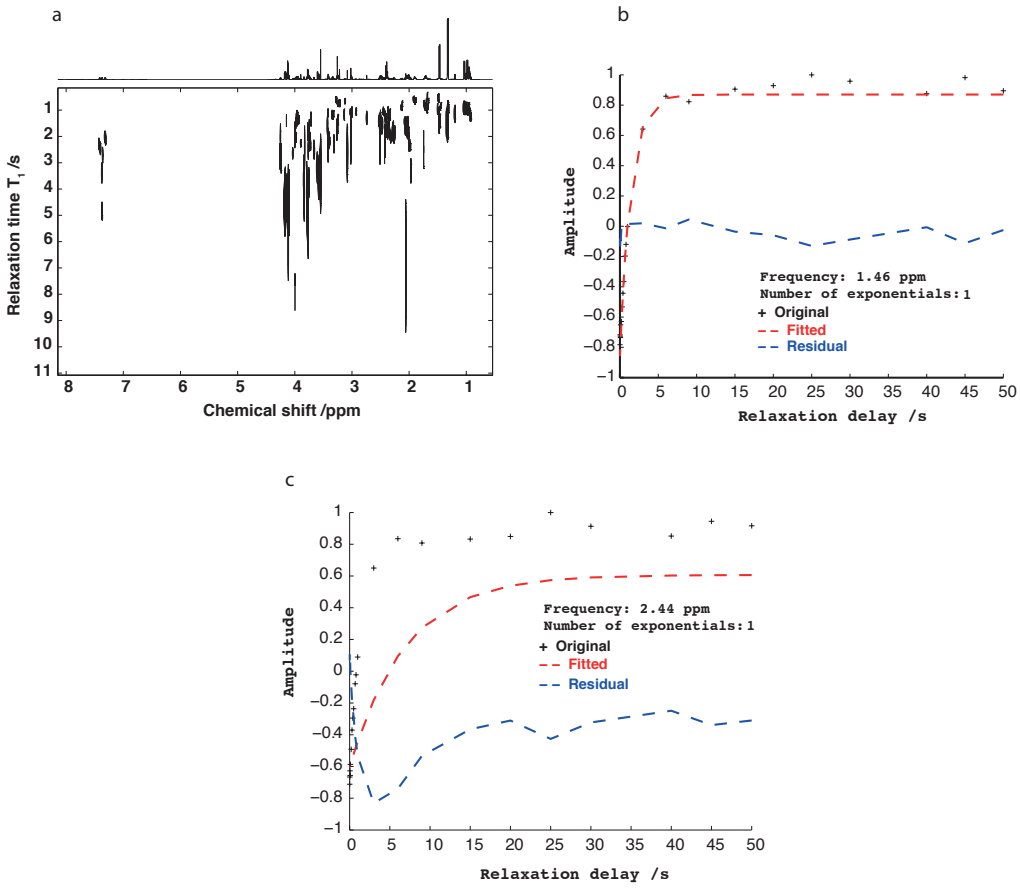


Figure 6.6. Pseudo-2D plot of the inversion-recovery data with the complex mixture in D_2O (a), with the fit and the residual plot from one peak from the alanine peak at 1.46 ppm (b) and from a glutamine peak at 2.44 ppm (c).

For this fit, one interesting note is that the first point of the experimental data has an amplitude of -0.7, while the highest amplitude is 0.9. The absolute values of maximum and

minimum amplitudes are different, while the equation 6.4 forces the minimum and the maximum amplitudes to have the same absolute values in the fit. This lack of flexibility in the equation used by the algorithm may have caused this problem with some peaks. Within Topspin, the equation used to fit the data in this T_1 module allows the possibility to have different absolute values for minimum and maximum values.

These results shows that optimization in the algorithm in the T_1 Toolbox is still needed in order to properly fit the experimental data. The first step will be to allow the absolute value of minimum and maximum amplitudes to be different. Other improvements may have to be included in order to increase the accuracy of the fit, which may decrease substantially the width of the peaks in the T_1 dimension. During the same time, the IR pulse sequence has been modified to allow the measurement of T_1 in highly protonated water.

2. IR pulse sequence with efficient solvent suppression

In order to obtain efficient water suppression, the IR pulse sequences have several requirements:

- The water residual peak should have the narrowest linewidth to distinguish as much peaks as possible close to the water peak. In the same spirit, in the case of excitation sculpting, the zone with altered intensities should also be as narrow as possible.

- Because of the variable delay between the 180° and the 90° pulse, the water suppression block should also give a water suppression that is robust against these delays. This gives an advantage to water suppression blocks placed after the mixing, like excitation sculpting.

- Since the T_1 estimation is based on the analysis of peak intensities, a flat baseline is necessary. This means that the water residual peak must have the same phase as the other peaks or impact only a small portion of the spectrum.

While several water suppression modules have already been tested with the IR pulse sequence,[298] we chose another water suppression module developed recently, the "perfect echo" WATERGATE,[77] along with presaturation. While the excitation sculpting allows a very good suppression of the water peak, the "perfect echo" was chosen to avoid the peak distortion from anti-phase build-up during the echoes. The pulse sequence was a modification of the Bruker "t1ir" pulse sequence, with the addition of the presaturation and "perfect echo" WATERGATE. The phase cycling of the 90° and 180° pulses were the same as "t1ir", while only the first step of the "perfect echo" WATERGATE phase cycling was kept.

The pulse sequence is shown in figure 6.7.a, and has been tested with one of the synthetic serum samples used in part 2 and 3. Twenty delays were used between the 180° and the 90° pulses in seconds: 0.01, 0.02, 0.05, 0.07, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 1, 3, 6, 9, 15, 20, 25, 30 and 40. The inter-scan delay was 35 s, to allow a complete relaxation of the protons between scans. The last 3 seconds were also used to presaturate the water peak. The spectra are shown in figure 6.7.b for a variable delay of 40 s and in figure 6.7.c for a variable delay of 10 ms. The selective pulses were rectangular pulses of 2 ms for these spectra.

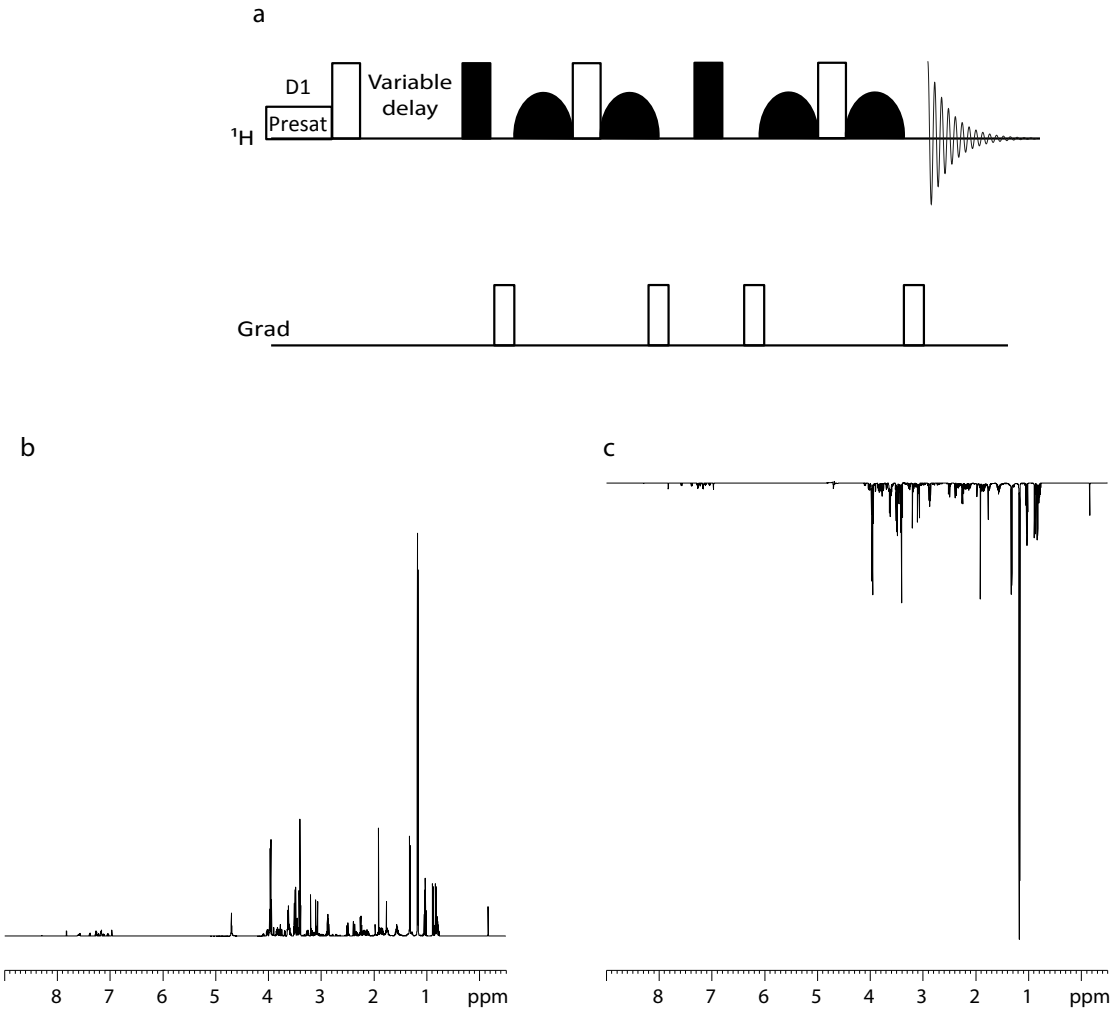


Figure 6.7. Pulse sequence for IR with presaturation and "perfect echo" WATERGATE (a) and the 1D spectrum with a complex mixture with a variable delay of 40 s (b) or 10 ms (c).

The behavior of the peaks was the expected ones, with a change of signs between short and long delays between the 180° and the 90° pulses. The spectrum with a low delay is presented with peaks with negative intensities to respect the results from Equation (6.1), where the intensity is indeed negative for short delays compared to T_1. All the peaks have a pure absorption lineshape, with no contribution from anti-phase peaks, as expected from the "perfect-echo" WATERGATE.

One interesting result is that the water residual peak has the same phase as the other peaks. As a result, it does not distort the baseline. However, the closest peak to the water residual peak, from carnitine at 4.5 ppm, is not visible on the two spectra. This is caused by the rather low delay of the selective pulse, which made the excitation profile large enough to suppress the intensity from this peak. Using longer delays for the selective pulses allows seeing the carnitine peak, but gives slightly lower water suppression.

Therefore, this pulse sequence can be used for T_1 evaluation. The data were exported to the T_1 Toolbox and subjected to the same routine described in section 1.1. The resulting plot is shown in figure 6.8, along with the residual plot from the lactate peak at 1.33 ppm and one of the ethanol peaks at 1.19 ppm. In this pseuso-2D plot, we see similar attributes to the T_1 analysis in section 1.4. There is a large range of T_1, from hundreds of ms to several seconds, the uncertainty on the T_1 value, shown by the width of the peaks in the T_1 dimension, becomes greater as the T_1 value itself is bigger. Similar to section 1.4, peaks with low T_1 tends to have fitted exponential reproducing experimental data very closely, while peaks with higher T_1 have a high residual plot.

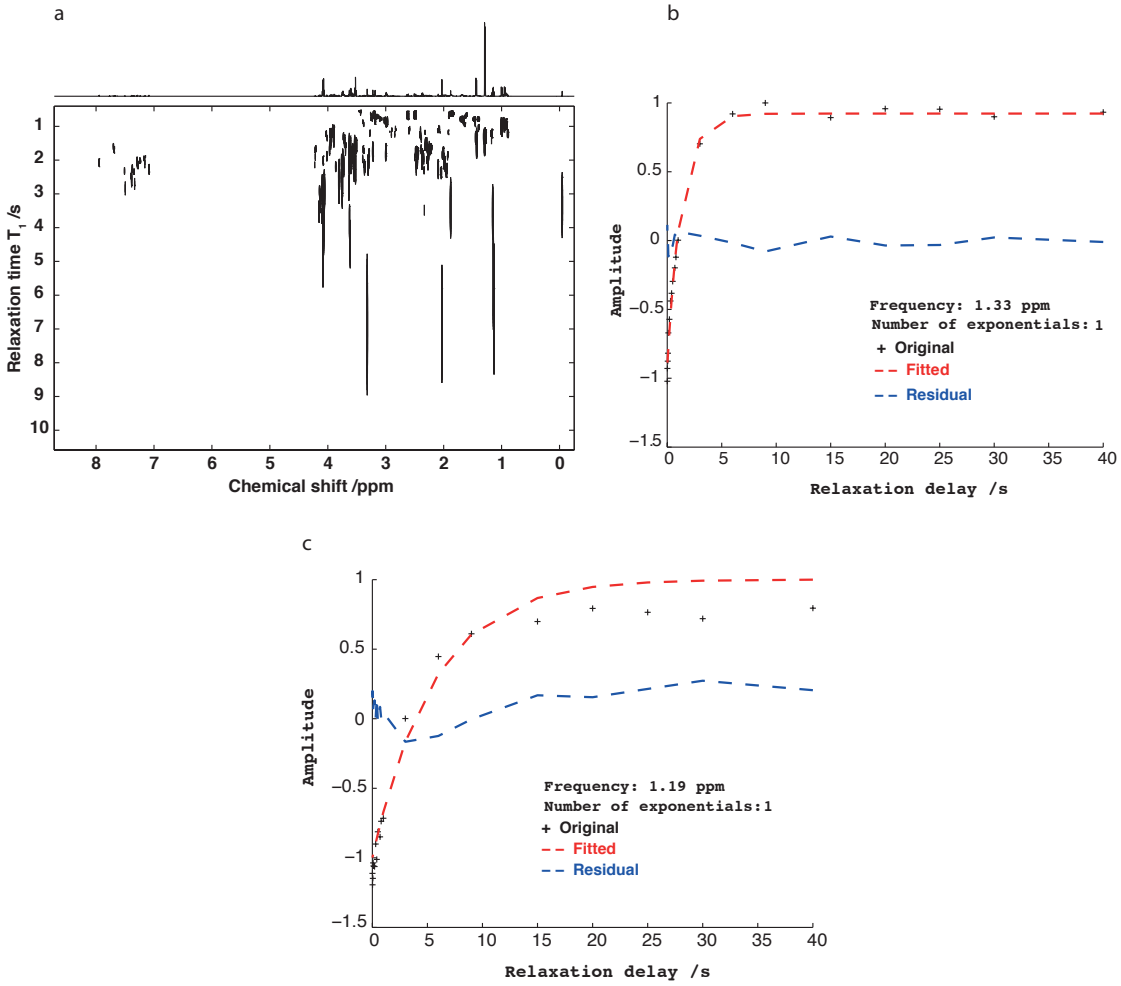


Figure 6.8. Pseudo-2D plot of the inversion-recovery data with the complex mixture in H_2O/D_2O (90/10) (a), with the fit and the residual plot (a) from one peak from the lactate peak at 1.33 ppm (b) and from the ethanol peak at 1.19 ppm (c).

Surprisingly, among the peaks with the largest T_1 values, there are the CH_3 peaks from methanol (T_1 = 7 s), the CH_3 from ethanol (T_1 = 6 s) and the CH_3 from TMS (T_1 = 3 s). These high T_1 values for CH_3 groups, which are often seen as the quickest to relax, shows that there is indeed a need to evaluate the T_1 values, since these values are not always predictable. However, the pre-mentioned problem in the algorithm is still necessary in this case, since the T_1 measurement of these protons suffers from high uncertainty due to the difficulty of the algorithm to properly fit the experimental data from these protons, as seen for one of the ethanol peaks at 1.19 ppm in figure 6.8.c.

From this analysis, we can conclude that with the T_1 Toolbox and the IR pulse sequence with "perfect echo" WATERGATE, there is a possibility to automate the T_1 analysis of small molecules in highly protonated water. But improvements in the algorithm are still necessary for a better accuracy and repeatability, along with additional improvements in the pulse sequence. For example, replacing the hard 180° pulse by an adiabatic one, in order to reduce the sensitivity to pulse imperfection. Now we can verify if other modifications to the DOSY Toolbox can also allow the estimation of T_2 values.

3. The T_2 toolbox

As far as we know, the estimation of T_2 values is even less frequent than T_1 values for small molecules. Some reasons that may explain the lack of investigations on T_2 values are the similarities between T_1 and T_2 values for small molecules, the possibility to obtain a partial estimation with peak lineshapes or that the pulse sequence to measure T_2 values, the Carr-Purcell-Meiboom-Gill (CPMG), needs to use hundreds or thousands of 180° pulses to measure high T_2 values. Compared to the IR pulse sequence, where only a delay is introduced between two peaks, the CPMG is therefore more sensitive to pulse imperfections. It should also be noted that T_2 values serve mainly to determine the acquisition times, in both dimensions, which are mainly chosen for resolution issues.

Nevertheless, T_2 estimations could also be used to set acquisition times, in order to optimize the signal-to-noise ratio (SNR) per unit of time, since this value is maximum when the acquisition time is equal to roughly 1.3 T_2*.[286] At first, the new equations will be expressed for T_2 calculations. Then this new Toolbox, which will be called the T_2 Toolbox, will be tested with the CPMG pulse sequence and finally with the PROJECT pulse sequence.

3.1. Equations for T_2 calculations

Similarly to T_1 equations, to calculate T_2 values by the DOSY Toolbox, the parameter conversion needs to be changed from gradient intensity to the variable spin-echo delay, the exponential function, its derivative and the conversion from exponential decay to the parameter T_2. The variable spin-echo delay is extracted from a variable counter delay, which is multiplied by the duration of a spin-echo to obtain the variable spin-echo delay.

For the relaxation delays, the exponential function, derived from Bloch equations,[299] is:

$$M(t, T_2) = M_0 e^{-\frac{t}{T_2}} \qquad (6.7)$$

where M_0 is the equilibrium magnetization. The derivative of the targeted parameter is:

$$\frac{dM}{dT_2} = 2M_0 \frac{t}{T_2^2} e^{-\frac{t}{T_2}} \qquad (6.8)$$

Finally, the parameter T_2 can be extracted with the equation:

$$T_2 = -\frac{t}{\ln\left(\frac{1-\frac{M(t,T_2)}{M_0}}{2}\right)} \qquad (6.9)$$

Changes for the GUI can also be added to adapt the presentation to the parameter used, in this case T_2 estimations. With these modifications, we can evaluate the T_2 Toolbox for the estimation of T_2 values.

3.2. Use of CPMG with water suppression for T_2 estimations

Like the IR pulse sequence, the CPMG pulse sequence was already tested with water suppression modules.[299] We tested some of them, but the most simple to implement (composite pulses, presaturation alone) gave spectra with the presence of faraway water. The other variants had a different phase cycle for the 180° pulse, which transformed the pulse sequence into a Carr-Purcell pulse sequence. Since the Carr-Purcell is sensitive to the calibration of the 180° pulse, we decided to use another water suppression module. Like the IR, we used the "perfect echo" WATERGATE with presaturation, by adding these two modules to the Bruker pulse sequence "cpmg". The phase cycling of the 90° and 180° pulses were the same as "cpmg", while only the first step of the "perfect echo" WATERGATE phase cycling was kept.

A sample of valine at 13 mM in H_2O/D_2O (90/10) was used for this test. Twelve different numbers of spin echoes were used: 4, 12, 20, 40, 60, 80, 100, 140, 200, 300, 400 and 800. Each 180° in the spin echo was separated by 2 ms (so the delay τ in figure 1.4 was fixed at 1 ms). The inter-scan delay was 15 s, to allow a complete relaxation of the protons between scans. The last 3 seconds were also used to presaturate the water peak. The acquired data were processed in the T_2 Toolbox, as presented in section 1.1. The pseudo-2D plot and the residual plot of one of the peaks from the CH_3 group are shown in figure 6.9.

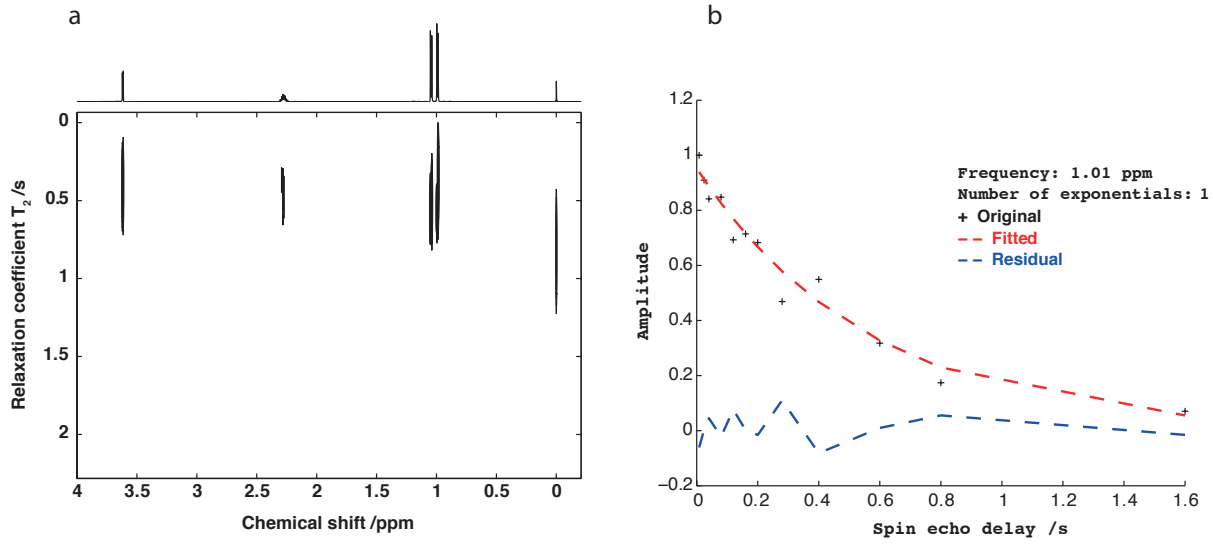


Figure 6.9. Pseudo-2D plot of the CPMG data with the valine sample at 13 mM in H_2O/D_2O (90/10) (a), with the fit and the residual plot (a) from one peak from the CH_3 group at 1.01 ppm (b).

From the pseudo-2D plot, we can see small values of T_2, all inferior to 1 s, even for the peak from TSP. However, we can see from the residual plot that the exponential decay does not fit completely the data, especially with the first points. This problem was seen in all variants of the CPMG, so the spin echo may have caused the problem. Another potential problem is the rather short values for spin echoes, with the maximum value at 1.6 s. This value is close to the maximum

recommended for delay τ chosen. But if we raise further the delay, the lineshapes will start to be distorted by anti-phase peaks. For these reasons, we chose to evaluate another pulse sequence for T_2 estimations.

3.3. Evaluation of PROJECT for T_2 estimation

As stated in section 1.3 of part 1, the PROJECT pulse sequence has already been used very recently for T_2 estimations.[83] Since the PROJECT also uses the "perfect echo" to avoid the formation of anti-phase peaks, we chose to use this pulse sequence for the estimation of T_2 in complex mixtures. That module was placed in lieu of the CPMG module, from the pulse sequence used in the previous section, with the phase cycle described in the first publication of the PROJECT pulse sequence.[82]

This part of the study is still in progress, but partial results can still be shown. The most convincing results so far were obtained with the study of the complex mixture already used in section 2. Twelve different numbers of spin echoes were used: 2, 6, 10, 20, 30, 40, 70, 100, 150, 200, 400 and 800. Two successive 180° pulses in the spin echo were separated by 2 ms (so the delay τ in figure 1.4 was fixed at 1 ms), with a 90° pulse between two spin echoes half the time. The inter-scan delay was 15 s, to allow a complete relaxation of the protons between scans. The last 3 seconds were also used to presaturate the water peak. The acquired data were processed in the T_2 Toolbox, as presented in section 1.1. The pulse sequence is shown in figure 6.10.a. The spectra are shown in figure 6.10.b for a spin echo delay of 8 ms and in figure 6.10.c for a spin echo delay of 3.2 ms.

In both spectra, the residual water peak is negligible, and the baseline is flat near the water peak. However, the spectra could not be properly phased and some peak distortions can be seen. The peak distortion remained constant for all the number of spin echoes tested. We can also see that at 3.2 s, peaks are still visible, which means that the magnetization was not completely relaxed after 3.2 s.

The acquired data were processed in the T_2 Toolbox, as presented in section 1.1. The pseudo-2D plot and the residual plot of one of the peaks from the lactate peak at 1.33 ppm are shown in figure 6.11.

In the 2D plot, we can see a large range of T_2, from hundreds of ms to a few seconds. However, the uncertainty in the T_2 value is globally greater than for T_1 values. This may be caused by the peak distortions or by the fact that magnetization have not completely relaxed. One last possibility is the fact that the exponential decay does not completely fit the experimental data, especially the shortest values of spin echo delay, as with CPMG.

The next step in that study is the optimization of the PROJECT pulse sequence with the "perfect echo" WATERGATE, in order to obtain pure absorption peaks in the whole spectral width. Another step will be to increase the delay τ to increase the maximum value of the spin echo delay without the risk of probe heating.

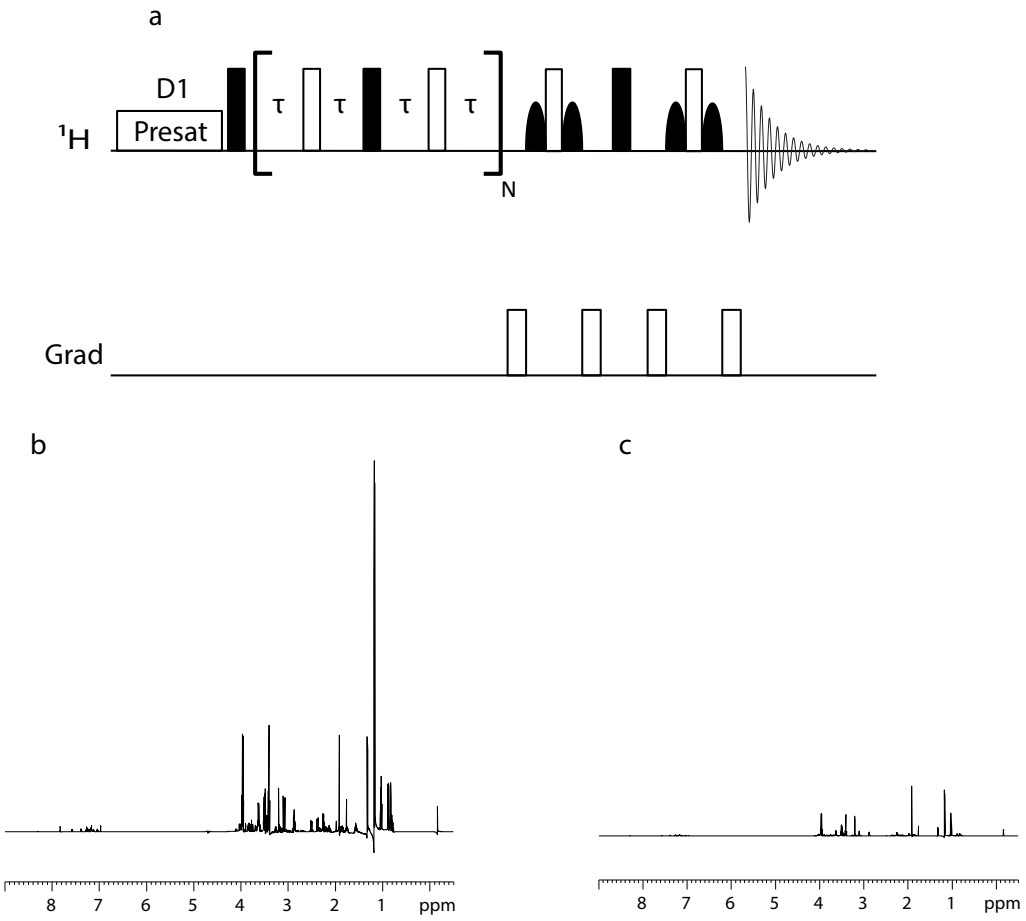


Figure 6.10. Pulse sequence for PROJECT with presaturation and "perfect echo" WATERGATE (a) and the 1D spectrum with a complex mixture with a spin echo delay of 8 ms (b) or 3.2 s (c). The delay τ was fixed at 1 ms.

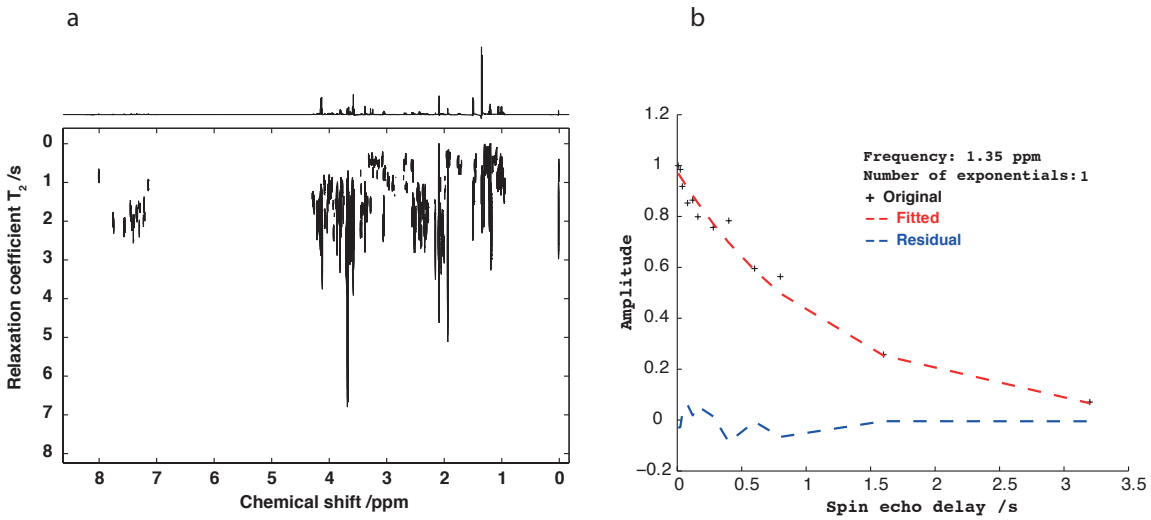


Figure 6.11. Pseudo-2D plot of the CPMG data with the complex mixture in H_2O/D_2O (90/10) (a), with the fit and the residual plot (a) from one peak from the lactate peak at 1.35 ppm (b).

In conclusion, we started to develop a convenient tool to estimate T_1 and T_2 in a pseudo-2D plot, which should make the study of complex mixtures easier. But an optimization of the algorithm is still necessary to have a reliable measurement of T_1 and T_2 within a large dynamic range. We also modified the usual pulse sequences for T_1 and T_2 calculation in order to properly estimate these values in highly protonated water.

While the IR pulse sequence seems reliable with our latest optimization, the PROJECT pulse sequence still needs optimization in order to use it for metabolomics samples. Experiments in real biological samples should also be tested to verify the robustness of the pulse sequence.

Conclusion and perspectives

During this study, several steps have been done to show the usefulness of fast 2D NMR for metabolomics. More notably, we have shown with synthetic mixtures that statistical analysis of 2D pulse sequences can be more truthful to the statistical analysis of concentrations than the statistical analysis of 1D ^{1}H spectra, thanks to the reduced overlap between peaks. Even more importantly, we have shown that fast 2D NMR variants of COSY and HSQC pulse sequences could give similar information by statistical analysis compared to their conventional counterparts, in a much shorter time. In particular, the UF COSY, at the time of writing, can be used to identify potential biomarkers with a concentration down to a few hundreds of µM in the sample with only 15 min of experimental time. For the NUS COSY and HSQC, the results of statistical analysis were very similar to those obtained by their conventional counterparts. Fast 2D NMR approaches are therefore fully relevant for high-throughput metabolomics studies. This outcome resulted in three other studies.

The first one was the development of an alternative to UF COSY in order to avoid diagonal peaks while keeping a similar sensitivity for correlation peaks. The UF variant of double-quantum spectroscopy was chosen and it effectively suppressed the diagonal peaks, reducing the risk of overlap. We also developed another variant of UF DQS in order to obtain a COSY-like spectrum, bringing the possibility to use COSY libraries to identify metabolites. In that study, we showed its usefulness to isolate correlation peaks close to the diagonal peaks. This opens the possibility to use this pulse sequence instead of UF COSY for complex mixture analysis where a lot of correlation peaks are close to the diagonal, like sugars.[242] In addition, the interleaving scheme developed by the group of Lucio Frydman at the Weizmann Institute[215] and recently optimized by the EBSI group, in the CEISAM laboratory from the University of Nantes,[265] synergizes well with UF DQS. However, some peaks seem to be affected by J-modulation and/or strong coupling. A study of these effects is currently done by Boris Gouilleux and Laetitia Rouger, PhD students in the EBSI group in Nantes.

After that, the effect of time-equivalent NUS was tested for the case of complex mixtures. At first, sensitivity-enhanced NUS was tested, by using the time gained from the unrecorded t_1 increments to increase the number of scans. But no significant improvement of the sensitivity was found with this strategy. The second test was resolution-enhanced NUS, where the maximum evolution time in the indirect was increased instead of the number of scans. This strategy gave an important gain in resolution for the HSQC spectrum (up to 32 times less width at half-height) without significant losses in sensitivity. For TOCSY, the increase in resolution allowed to see multiplet patterns in the indirect dimension, increasing the chances of identification. But sensitivity losses were seen in the resolution-enhanced spectra because of the multiplet splitting.

A last point to optimize is the study of T_1 and T_2 in complex mixtures. While a modified toolbox was used in order to produce a pseudo 2D plot of T_1 or T_2 for each peak, the algorithm still needs to be optimized. Variants of the pulse sequences for T_1 and T_2 pulse sequences have also been developed for a good and robust suppression of the water peak. The water suppression was indeed robust but tests are still needed for the accuracy of the T_1 and T_2 estimations.

While these new tools were efficient with synthetic samples, it remains necessary to test them with biological samples to show their usefulness for metabolomics. A first evaluation of UF DQS has started in a study by HR-MAS by Marion André, in order to use UF 2D NMR with magic angle spinning. The UF DQS pulse sequence was preferred because the sample studied possesses several sugars, which means . The quantification by UF DQS in complex mixtures is also evaluated by Laetitia Rouger.

For NUS, the most recent algorithms for the processing of NUS spectra should be tested, like the Low Rank algorithm,[292] to verify if the fidelity of NUS spectra to the US spectra is greater than with CS. The analysis of noise should also be deepened, in order to see if we can substantially increase sensitivity by post-processing.[300]

Finally, other approaches could be evaluated for metabolomics. One interesting approach in that respect is ASAP HSQC.[176] This approach could be evaluated on its own, for the reduction of experimental time or increase in resolution. An even more interesting potential is its combination with NUS for the acquisition of HSQC spectra with very high resolution in a minimal experimental time, or the even greater reduction of the experimental time.

Despite all these progresses for the reduction of the experimental time in 2D NMR, the use of NMR for analysis of complex mixtures remains limited by sensitivity, especially compared to mass spectrometry. This limitation has been known since the beginning of NMR and has been continually reduced by raising the magnetic field and improvement in pulse sequences. The most recent improvement in sensitivity comes from hyperpolarization [301], which can considerably raise the magnetization of a sample while keeping the same concentration. For liquid NMR, Dynamic Nuclear Polarization (DNP)[302] has shown great promises, but single scan approaches are necessary, because of the transient nature of hyperpolarized polarization.

Recently, it has been shown that DNP was compatible by UF 2D NMR, with natural products[223] or complex mixtures[303] which could resolve the sensitivity issue of UF 2D NMR. These first studies have shown that several metabolites can be hyperpolarized simultaneously, with a good repeatability and important enhancement in sensitivity. However, the transient nature of hyperpolarization does not allow the use of hybrid UF approaches and restrict its use to single-scan UF 2D NMR, at least for now.

In conclusion, this PhD thesis has opened interesting perspectives, both in terms of methodological developments and applications. Some of them are currently investigated, and we hope that the metabolomics community will be interested in evaluating these new methods in studies on real biological samples.

Tables and figures

Tables

Figures

List of publications and communications

Publications from the thesis

Le Guennec, A., Dumez, J.-N., Giraudeau, P., Caldarelli, S. ' Resolution-enhanced 2D NMR of complex mixtures by non-uniform sampling ', *Magn. Reson. Chem.*, 2015, in press

Le Guennec, A., Giraudeau, P., Caldarelli, S., and Dumez, J.-N., 'Ultrafast double-quantum NMR spectroscopy', *Chem. Commun.*, 2015, 51, 354-357

Le Guennec, A., Giraudeau, P., and Caldarelli, S., 'Evaluation of Fast 2D NMR for Metabolomics', *Anal. Chem.*, 2014, 86, 5946–5954

Other publications

Tea, I., **Le Guennec, A.**, Frasquet-Darrieux, M., Julien, M., Romek, K., Antheaume, I., Hankard, R. and Robins, R.J., 'Simultaneous determination of natural-abundance d15N values and quantities of individual amino acids in proteins from milk of lactating women and from infant hair using gas chromatography/isotope ratio mass spectrometry', *Rapid Commun. Mass Spectrom.*, **2013**, 27, 1345–1353

Le Guennec, A., Tea, I., Antheaume, I., Martineau, E., Charrier, B., Pathan, M., Akoka, S., and Giraudeau, P., 'Fast Determination of Absolute Metabolite Concentrations by Spatially Encoded 2D NMR: Application to Breast Cancer Cell Extracts', *Anal. Chem.*, **2012**, 84, 10831-10837

Oral communications from the thesis (Presenting authors are in bold)

Le Guennec, A., Rouger, L., Gouilleux, B., Kuprov, I., Levitt, M.H., Akoka, S., Caldarelli, S., Giraudeau, P., **Dumez, J.-N.**, Ultrafast double-quantum NMR spectroscopy, 55th ENC, Pacific Grove, USA, 19-24 April 2015

Le Guennec, A., Caldarelli, S., Giraudeau, P., 'Évaluation de la RMN 2D rapide pour la métabolomique à haut débit', GERM (Groupe d'Études de Résonance Magnétique), 1-6 June 2014, Biarritz, France

Le Guennec, A., Caldarelli, S., Giraudeau, P., 'Évaluation de techniques RMN 2D rapides pour la métabolomique', Réunion RMN Grand Bassin Parisien, 17th April 2014, Orsay, France

Posters from the thesis

Le Guennec, A., Caldarelli, S., Giraudeau, P., 'Evaluation of fast 2D NMR techniques for high-throughput metabolomics', AISBM 2014, 23-24 October 2014, Gif-sur-Yvette, France

Le Guennec, A., Caldarelli, S., Giraudeau, P., 'Evaluation of fast 2D NMR techniques for high-throughput metabolomics', EUROMAR 2014, 29 Juin - 3 July 2014, Zurich, Swaziland

Le Guennec, A., Caldarelli, S., Giraudeau, P., 'Évaluation de la RMN 2D rapide pour la métabolomique à haut débit', GERM (Groupe d'Études de Résonance Magnétique), 1-6 June 2014, Biarritz, France

Bibliography

[1] J. C. Lindon, J. K. Nicholson, E. Holmes, J. R. Everett, *Concept. Magnetic Res.* **2000**, Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids, 12, 289-320.

[2] J. K. Nicholson, J. Connelly, J. C. Lindon, E. Holmes, others, *Nat. Rev. Drug Discov.* **2002**, Metabonomics: a platform for studying drug toxicity and gene function, 1, 153-162.

[3] O. Beckonert, H. C. Keun, T. M. D. Ebbels, J. Bundy, E. Holmes, J. C. Lindon, J. K. Nicholson, *Nat. Protocols* **2007**, Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts, 2, 2692-2703.

[4] J. Trygg, E. Holmes, T. Lundstedt, *J. Proteome Res.* **2006**, Chemometrics in Metabonomics, 6, 469-479.

[5] K. Bingol, R. Bruschweiler, *Anal. Chem.* **2014**, Multidimensional approaches to NMR-based metabolomics, 86, 47-57.

[6] J. Nielsen, S. Oliver, *Trends Biotechnol.* **2005**, The next wave in metabolome analysis, 23, 544-546.

[7] F. Wei, K. Furihata, M. Koda, F. Hu, R. Kato, T. Miyakawa, M. Tanokura, *J. Agric. Food. Chem.* **2012**, C-13 NMR-Based Metabolomics for the Classification of Green Coffee Beans According to Variety and Origin, 60, 10118-10125.

[8] C. S. Clendinen, B. Lee-McMullen, C. M. Williams, G. S. Stupp, K. Vandenborne, D. A. Hahn, G. A. Walter, A. S. Edison, *Anal. Chem.* **2014**, 13C NMR Metabolomics: Applications at Natural Abundance.

[9] A. M. Weljie, J. Newton, P. Mercier, E. Carlson, C. M. Slupsky, *Anal. Chem.* **2006**, Targeted Profiling : Quantitative Analysis of 1H NMR Metabolomics Data, 78, 4430-4442.

[10] J. Jeener, Basko Polje, Yugoslavia, **1971**.

[11] W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **1976**, Two-dimensional spectroscopy. Application to nuclear magnetic resonance, 64, 2229-2246.

[12] M. R. Viant, *Biochem. Biophys. Res. Commun.* **2003**, Improved methods for the acquisition and interpretation of NMR metabolomic data, 310, 943-948.

[13] C. Ludwig, M. R. Viant, *Phytochem. Anal* **2010**, Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox, 21, 22-32.

[14] M. J. Thrippleton, R. A. E. Edden, J. Keeler, *J. Magn. Reson.* **2005**, Suppression of strong coupling artefacts in J-spectra, 174, 97-109.

[15] H. M. Parsons, C. Ludwig, M. R. Viant, *Magn. Reson. Chem.* **2009**, Line-shape analysis of J-resolved NMR spectra: application to metabolomics and quantification of intensity errors from signal processing and high signal congestion, 47, S86-S95.

[16] L. Braunschweiler, R. Ernst, *J. Magn. Reson.* **1983**, Coherence transfer by isotropic mixing: application to proton correlation spectroscopy, 53, 521-528.

[17] F. C. Schroeder, D. M. Gibson, A. C. L. Churchill, P. Sojikul, E. J. Wursthorn, S. B. Krasnoff, J. Clardy, *Angew. Chem. Int. Ed.* **2007**, Differential Analysis of 2D NMR Spectra: New Natural Products from a Pilot-Scale Fungal Extract Library, 46, 901-904.

[18] Q. N. Van, H. J. Issaq, Q. Jiang, Q. Li, G. M. Muschik, T. J. Waybright, H. Lou, M. Dean, J. Uitto, T. D. Veenstra, *J. Proteome Res.* **2008**, Comparison of 1D and 2D NMR Spectroscopy for Metabolic Profiling, 7, 630-639.

[19] M. Koda, K. Furihata, F. Wei, T. Miyakawa, M. Tanokura, *J. Agric. Food. Chem.* **2012**, NMR-Based Metabolic Profiling of Rice Wines by F2-Selective Total Correlation Spectra, 60, 4818-4825.

[20] S. Massou, C. Nicolas, F. Letisse, J.-C. Portais, *Phytochemistry* **2007**, NMR-based fluxomics: Quantitative 2D NMR methods for isotopomers analysis, 68, 2330-2340.

[21] S. Massou, C. Nicolas, F. Letisse, J.-C. Portais, *Metab. Eng.* **2007**, Application of 2D-TOCSY NMR to the measurement of specific 13C-enrichments in complex mixtures of 13C-labeled metabolites, 9, 252-257.

[22] I. A. Lewis, R. H. Karsten, M. E. Norton, M. Tonelli, W. M. Westler, J. L. Markley, *Anal. Chem.* **2010**, NMR Method for Measuring Carbon-13 Isotopic Enrichment of Metabolites in Complex Solutions, 82, 4558-4563.

[23] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, 69, 185-189.

[24] Y. Date, T. Iikura, A. Yamazawa, S. Moriya, J. Kikuchi, *J. Proteome Res.* **2012**, Metabolic Sequences of Anaerobic Fermentation on Glucose-Based Feeding Substrates Based on Correlation Analyses of Microbial and Metabolite Profiling, 11, 5602-5610.

[25] S. Mavel, L. Nadal-Desbarats, H. Blasco, F. Bonnet-Brilhault, C. Barthélémy, F. Montigny, P. Sarda, F. Laumonnier, C. R. Andres, P. Emond, *Talanta* **2013**, 1H-13C NMR-based urine metabolic profiling in autism spectrum disorders, 114, 95-102.

[26] I. A. Lewis, S. C. Schommer, B. Hodis, K. A. Robb, M. Tonelli, W. M. Westler, M. R. Sussman, J. L. Markley, *Anal. Chem.* **2007**, Method for Determining Molar Concentrations of Metabolites in Complex Solutions from Two-Dimensional 1H-13C NMR Spectra, 79, 9385-9390.

[27] J. Yuk, J. R. McKelvie, M. J. Simpson, M. Spraul, A. J. Simpson, *Environ. Chem.* **2010**, Comparison of 1-D and 2-D NMR techniques for screening earthworm responses to sub-lethal endosulfan exposure, 7, 524-536.

[28] J. Yuk, M. J. Simpson, A. J. Simpson, *Environ. Chem.* **2011**, 1-D and 2-D NMR metabolomics of earthworm responses to sub-lethal trifluralin and endosulfan exposure, 8, 281-294.

[29] J. Yuk, M. J. Simpson, A. J. Simpson, *Environ. Pollut.* **2013**, 1-D and 2-D NMR-based metabolomics of earthworms exposed to endosulfan and endosulfan sulfate in soil, 175, 35-44.

[30] J. Kikuchi, K. Shinozaki, T. Hirayama, *Plant Cell Physiol*. **2004**, Stable Isotope Labeling of Arabidopsis thaliana for an NMR-Based Metabolomics Approach, 45, 1099-1104.

[31] Y. Date, Y. Nakanishi, S. Fukuda, T. Kato, S. Tsuneda, H. Ohno, J. Kikuchi, *J. Biosci. Bioeng*. **2010**, New monitoring approach for metabolic dynamics in microbial ecosystems using stable-isotope-labeling technologies, 110, 87-93.

[32] S. G. Hyberts, H. Arthanari, S. A. Robson, G. Wagner, *J. Magn. Reson*. **2014**, Perspectives in magnetic resonance: NMR in the post-FFT era, 241, 60-73.

[33] R. K. Rai, N. Sinha, *Anal. Chem*. **2012**, Fast and Accurate Quantitative Metabolic Profiling of Body Fluids by Nonlinear Sampling of 1H-13C Two-Dimensional Nuclear Magnetic Resonance Spectroscopy, 84, 10005-10011.

[34] S. G. Hyberts, G. J. Heffron, N. G. Tarragona, K. Solanky, K. A. Edmonds, H. Luithardt, J. Fejzo, M. Chorev, H. Aktas, K. Colson, K. H. Falchuk, J. A. Halperin, G. Wagner, *J. Am. Chem. Soc*. **2007**, Ultrahigh-Resolution 1H-13C HSQC Spectra of Metabolite Mixtures Using Nonlinear Sampling and Forward Maximum Entropy Reconstruction, 129, 5108-5116.

[35] L. Frydman, T. Scherf, A. Lupulescu, *Proc. Natl. Acad. Sci*. **2002**, The acquisition of multidimensional NMR spectra within a single scan, 99, 15858-15862.

[36] A. Tal, L. Frydman, *Prog. Nucl. Magn. Reson. Spectrosc*. **2010**, Single-scan multidimensional magnetic resonance, 57, 241-292.

[37] M. Pathan, B. Charrier, I. Tea, S. Akoka, P. Giraudeau, *Magn. Reson. Chem*. **2013**, New practical tools for the implementation and use of ultrafast 2D NMR experiments, 51, 168-175

[38] A. Le Guennec, I. Tea, I. Antheaume, E. Martineau, B. Charrier, M. Pathan, S. Akoka, P. Giraudeau, *Anal. Chem*. **2012**, Fast Determination of Absolute Metabolite Concentrations by Spatially Encoded 2D NMR: Application to Breast Cancer Cell Extracts, 84, 10831-10837

[39] T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry, Volume 27*, Second Edition ed., Elsevier Science, **2008**.

[40] J. Keeler, *Understanding NMR Spectroscopy, Second Edition*, Wiley, **2010**.

[41] T. I. H. G. M. Consortium, *Nature* **2001**, A physical map of the human genome, 409, 934-941.

[42] R. Goodacre, S. Vaidyanathan, W. B. Dunn, G. G. Harrigan, D. B. Kell, *Trends Biotechnol*. **2004**, Metabolomics by numbers: acquiring and understanding global metabolite data, 22, 245-252.

[43] H. F. Willard, M. Angrist, G. S. Ginsburg, *Philos. T. Roy. Soc. B* **2005**, Genomic medicine: genetic variation and its impact on the future of health care, 360, 1543-1550.

[44] W. C. S. Cho, *Genomics, Proteomics & Bioinformatics* **2007**, Proteomics Technologies and Challenges, 5, 77-85.

[45] J. C. Lindon, J. K. Nicholson, *TrAC, Trends Anal. Chem*. **2008**, Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery, 27, 194-204.

[46] N. Psychogios, D. D. Hau, J. Peng, A. C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z.

Hollander, T. L. Pedersen, S. R. Smith, F. Bamforth, R. Greiner, B. McManus, J. W. Newman, T. Goodfriend, D. S. Wishart, *PLoS ONE* **2011**, The Human Serum Metabolome, 6, e16957.

[47] I. F. Duarte, A. M. Gil, *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, Metabolic signatures of cancer unveiled by NMR spectroscopy of human biofluids, 62, 51-74.

[48] R. G. Ratcliffe, Y. Shachar-Hill, *Biological Reviews* **2005**, Revealing metabolic phenotypes in plants: inputs from NMR analysis, 80, 27-43.

[49] D. Rolin, C. Deborde, M. Maucourt, C. Cabasson, F. Fauvelle, D. Jacob, C. Canlet, A. Moing, in *Metabolomics Coming of Age with Its Technological Diversity, Vol. 67* (Ed.: D. Rolin), **2013**, pp. 1-66.

[50] N. L. Kuehnbaum, P. Britz-McKibbin, *Chem. Rev.* **2013**, New Advances in Separation Science for Metabolomics: Resolving Chemical Diversity in a Post-Genomic Era, 113, 2437-2468.

[51] B. Zhang, R. Powers, *Future Medicinal Chemistry* **2012**, Analysis of bacterial biofilms using NMR-based metabolomics, 4, 1273-1306.

[52] P. Giraudeau, I. Tea, G. S. Remaud, S. Akoka, *J. Pharm. Biomed. Anal.* **2013**, Reference and normalization methods: Essential tools for the intercomparison of NMR spectra, 96, 3-16.

[53] G. Barding, R. Salditos, C. Larive, *Anal. Bioanal. Chem* **2012**, Quantitative NMR for bioanalysis and metabolomics, 404, 1165-1179.

[54] A. C. Dona, B. Jiménez, H. Schäfer, E. Humpfer, M. Spraul, M. R. Lewis, J. T. M. Pearce, E. Holmes, J. C. Lindon, J. K. Nicholson, *Anal. Chem.* **2014**, Precision High-Throughput Proton NMR Spectroscopy of Human Urine, Serum, and Plasma for Large-Scale Metabolic Phenotyping, 86, 9887-9894.

[55] P. Bernini, I. Bertini, C. Luchinat, P. Nincheri, S. Staderini, P. Turano, *J. Biomol. NMR* **2011**, Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks, 49, 231-243.

[56] J. S. McKenzie, J. A. Donarski, J. C. Wilson, A. J. Charlton, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, Analysis of complex mixtures using high-resolution nuclear magnetic resonance spectroscopy and chemometrics, 59, 336-359.

[57] K. Kjeldahl, R. Bro, *J. Chemom.* **2010**, Some common misunderstandings in chemometrics, 24, 558-564.

[58] H. Mo, D. Raftery, *J. Biomol. NMR* **2008**, Improved residual water suppression: WET180, 41, 105-111.

[59] P. Giraudeau, V. Silvestre, S. Akoka, *Metabolomics, in press* **2015**, Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review.

[60] V. Krishnan, N. Murali, *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, Radiation damping in modern NMR experiments: Progress and challenges, 68, 41-57.

[61] Z. Serber, C. Richter, D. Moskau, J.-M. Böhlen, T. Gerfin, D. Marek, M. Häberli, L. Baselgia, F. Laukien, A. S. Stern, J. C. Hoch, V. Dötsch, *J. Am. Chem. Soc.* **2000**, New Carbon-Detected Protein NMR Experiments Using CryoProbes, 122, 3554-3555.

[62] G. Zheng, W. S. Price, *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, Solvent signal suppression in NMR, 56, 267-288.

[63] R. T. McKay, *Ann. R. NMR S.* **2009**, Recent advances in solvent suppression for solution NMR: a practical reference, 66, 33-76.

[64] D. I. Hoult, *J. Magn. Reson.* **1976**, Solvent peak saturation with single phase and quadrature fourier transformation, 21, 337-347.

[65] P. S. Wu, G. Otting, *J. Biomol. NMR* **2005**, SWET for secure water suppression on probes with high quality factor, 32, 243-250.

[66] H. Mo, D. Raftery, *J. Magn. Reson.* **2008**, Pre-SAT180, a simple and effective method for residual water suppression, 190, 1-6.

[67] J. K. Nicholson, P. J. D. Foxall, M. Spraul, R. D. Farrant, J. C. Lindon, *Anal. Chem.* **1995**, 750 MHz 1H and 1H-13C NMR Spectroscopy of Human Blood Plasma, 67, 793-811.

[68] B. C. M. Potts, A. J. Deese, G. J. Stevens, M. D. Reily, D. G. Robertson, J. Theiss, *J. Pharm. Biomed. Anal.* **2001**, NMR of biofluids and pattern recognition: assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse, 26, 463-476.

[69] R. T. Mckay, *Concept. Magn. Reson. A* **2011**, How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution, 38, 197-220.

[70] N. Aranibar, K.-H. Ott, V. Roongta, L. Mueller, *Anal. Biochem.* **2006**, Metabolomic analysis using optimized NMR and statistical methods, 355, 62-70.

[71] R. J. Ogg, R. B. Kingsley, J. S. Taylor, *J. Magn. Reson., Ser B* **1994**, WET, a T1- and B1-Insensitive Water-Suppression Method for in Vivo Localized 1H NMR Spectroscopy, 104, 1-10.

[72] S. H. Smallcombe, S. L. Patt, P. A. Keifer, *J. Magn. Reson., Ser A* **1995**, WET solvent suppression and its applications to LC NMR and high-resolution NMR spectroscopy, 117, 295-303.

[73] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **1992**, Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, 2, 661-665.

[74] M. Liu, X.-a. Mao, C. Ye, H. Huang, J. K. Nicholson, J. C. Lindon, *J. Magn. Reson.* **1998**, Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy, 132, 125-129.

[75] T.-L. Hwang, A. Shaka, *J. Magn. Reson., Ser A* **1995**, Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients, 112, 275-279.

[76] A. J. Simpson, S. A. Brown, *J. Magn. Reson.* **2005**, Purge NMR: effective and easy solvent suppression, 175, 340-346.

[77] R. W. Adams, C. M. Holroyd, J. A. Aguilar, M. Nilsson, G. A. Morris, *Chem. Commun.* **2013**, "Perfecting" WATERGATE: clean proton NMR spectra from aqueous solution, 49, 358-360.

[78] K. Takegoshi, K. Ogura, K. Hikichi, *J. Magn. Reson.* **1989**, A perfect spin echo in a weakly homonuclear J-coupled two spin-12 system, 84, 611-615.

[79] M. Malet-Martino, U. Holzgrabe, *J. Pharm. Biomed. Anal.* **2011**, NMR techniques in biomedical and pharmaceutical analysis, 55, 1-15.

[80] R. L. Vold, J. S. Waugh, M. P. Klein, D. E. Phelps, *J. Chem. Phys.* **1968**, Measurement of Spin Relaxation in Complex Systems, 48, 3831-3832.

[81] S. Meiboom, D. Gill, *Rev. Sci. Instrum.* **1958**, Modified Spin-Echo Method for Measuring Nuclear Relaxation Times, 29, 688-691.

[82] J. A. Aguilar, M. Nilsson, G. Bodenhausen, G. A. Morris, *Chem. Commun.* **2011**, Spin echo NMR spectra without J modulation, 48, 811-813

[83] L. Castanar, P. Nolis, A. Virgili, T. Parella, *J. Magn. Reson.* **2014**, Measurement of T1/T2 relaxation times in overlapped regions from homodecoupled 1H singlet signals, 244, 30-35.

[84] L. M. Smith, A. D. Maher, E. J. Want, P. Elliott, J. Stamler, G. E. Hawkes, E. Holmes, J. C. Lindon, J. K. Nicholson, *Anal. Chem.* **2009**, Large-Scale Human Metabolic Phenotyping and Molecular Epidemiological Studies via 1H NMR Spectroscopy of Urine: Investigation of Borate Preservation, 81, 4847-4856.

[85] P. B. Toft, H. Leth, H. C. Lou, O. Pryds, O. Henriksen, *J. Magn. Reson. Imaging* **1994**, Metabolite concentrations in the developing brain estimated with proton MR spectroscopy, 4, 680.

[86] J. P. Munasinghe, L. D. Colebrook, J. J. Attard, T. A. Carpenter, L. D. Hall, *Magn. Reson. Chem.* **1998**, Estimation of the spin–lattice relaxation time constants of low molecular weight solutes in dilute complex aqueous solutions: application to urinary metabolites, 36, 123.

[87] D. S. Wishart, *TrAC, Trends Anal. Chem.* **2008**, Quantitative metabolomics using NMR, 27, 228-237.

[88] P. Mercier, M. J. Lewis, D. Chang, D. Baker, D. S. Wishart, *J. Biomol. NMR* **2011**, Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra, 49, 307-323.

[89] G. A. Nagana Gowda, Y. N. Gowda, D. Raftery, *Anal. Chem.* **2015**, Expanding the Limits of Human Blood Metabolite Quantitation Using NMR Spectroscopy, 87, 706-715.

[90] C. M. Slupsky, K. N. Rankin, J. Wagner, H. Fu, D. Chang, A. M. Weljie, E. J. Saude, B. Lix, D. J. Adamko, S. Shah, R. Greiner, B. D. Sykes, T. J. Marrie, *Anal. Chem.* **2007**, Investigations of the Effects of Gender, Diurnal Variation, and Age in Human Urinary Metabolomic Profiles, 79, 6995-7004.

[91] F. Wei, K. Furihata, F. Hu, T. Miyakawa, M. Tanokura, *Magn. Reson. Chem.* **2010**, Complex mixture analysis of organic compounds in green coffee bean extract by two-dimensional NMR spectroscopy, 48, 857-865.

[92] V. Ramaswamy, J. W. Hooker, R. S. Withers, R. E. Nast, W. W. Brey, A. S. Edison, *J. Magn. Reson.* **2013**, Development of a 13C-optimized 1.5-mm high temperature superconducting NMR probe, 235, 58-65.

[93] M. G. Boersma, I. Solyanikova, W. J. H. Van Berkel, J. Vervoort, L. Golovleva, I. M. C. M. Rietjens, *J. Ind. Microbiol. Biotechnol.* **2001**, 19F NMR metabolomics for the elucidation of microbial degradation pathways of fluorophenols, 26, 22-34.

[94] M. Satake, B. Dmochowska, Y. Nishikawa, J. Madaj, J. Xue, Z. Guo, D. V. Reddy, P. L. Rinaldi, V. M. Monnier, *Invest. Ophth. Vis. Sci.* **2003**, Vitamin C Metabolomic Mapping in the Lens with 6-Deoxy-6-fluoro-ascorbic Acid and High-Resolution 19F-NMR Spectroscopy, 44, 2047-2058.

[95] M. A. DeSilva, N. Shanaiah, G. A. Nagana Gowda, K. Rosa-Pérez, B. A. Hanson, D. Raftery, *Magn. Reson. Chem.* **2009**, Application of 31P NMR spectroscopy and chemical derivatization for metabolite profiling of lipophilic compounds in human serum, 47, S74-S80.

[96] D. J. States, R. A. Haberkorn, D. J. Ruben, *J. Magn. Reson.* **1982**, A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants, 48, 286-292.

[97] D. Marion, K. Wüthrich, *Biochem. Biophys. Res. Commun.* **1983**, Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins, 113, 967-974.

[98] A. L. Davis, J. Keeler, E. D. Laue, D. Moskau, *J. Magn. Reson.* **1992**, Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients, 98, 207-216.

[99] W. P. Aue, *J. Chem. Phys.* **1976**, Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy, 64, 4226-4227.

[100] G. Bodenhausen, R. Freeman, D. L. Turner, *J. Chem. Phys.* **1976**, Two-dimensional J spectroscopy: Proton-coupled carbon-13 NMR, 65, 839-840.

[101] U. Piantini, O. Sorensen, R. R. Ernst, *J. Am. Chem. Soc.* **1982**, Multiple quantum filters for elucidating NMR coupling networks, 104, 6800-6801.

[102] C. J. Turner, *J. Magn. Reson.* **1992**, Phase cycling in double-quantum-filtered COSY, 96, 551-562.

[103] M. Rance, *J. Magn. Reson.* **1987**, Improved techniques for homonuclear rotating-frame and isotropic mixing experiments, 74, 557-564.

[104] M. J. Thrippleton, J. Keeler, *Angew. Chem.* **2003**, Elimination of Zero-Quantum Interference in Two-Dimensional NMR Spectra, 115, 4068-4071.

[105] A. Wokaun, R. R. Ernst, *Chem. Phys. Lett.* **1977**, Selective detection of multiple quantum transitions in NMR by two-dimensional spectroscopy, 52, 407-412.

[106] M. G. N. Reddy, S. Caldarelli, *Anal. Chem.* **2010**, Demixing of Severely Overlapping NMR Spectra through Multiple-Quantum NMR, 82, 3266-3269.

[107] C. Dalvit, J.-M. Bohlen, in *Annual Reports on NMR Spectroscopy, Vol. Volume 37* (Ed.: G. A. Webb), Academic Press, **1999**, pp. 203-271.

[108] A. Bax, R. Freeman, T. A. Frenkiel, *J. Am. Chem. Soc.* **1981**, An NMR technique for tracing out the carbon skeleton of an organic molecule, 103, 2102-2104.

[109] V. Wray, *Prog. Nucl. Magn. Reson. Spectrosc.* **1979**, Carbon-carbon coupling constants: A compilation of data and a practical guide, 13, 177-256.

[110] D. L. Turner, *J. Magn. Reson.* **1982**, Carbon-13 autocorrelation NMR using double-quantum coherence, 49, 175-178.

[111] A. Bax, T. H. Mareci, *J. Magn. Reson.* **1983**, Practical aspects of carbon-13 double quantum NMR, 53, 360-363.

[112] A. G. Palmer III, J. Cavanagh, P. E. Wright, M. Rance, *J. Magn. Reson.* **1991**, Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy, 93, 151-170.

[113] R. K. Rai, P. Tripathi, N. Sinha, *Anal. Chem.* **2009**, Quantification of Metabolites from Two-Dimensional Nuclear Magnetic Resonance Spectroscopy: Application to Human Urine Samples, 81, 10232-10238.

[114] P. Giraudeau, *Magn. Reson. Chem.* **2014**, Quantitative 2D liquid-state NMR, 52, 272.

[115] K. Hu, J. J. Ellinger, R. A. Chylla, J. L. Markley, *Anal. Chem.* **2011**, Measurement of Absolute Concentrations of Individual Compounds in Metabolite Mixtures by Gradient-Selective Time-Zero 1H-13C HSQC with Two Concentration References and Fast Maximum Likelihood Reconstruction Analysis, 83, 9352-9360.

[116] H. Koskela, O. Heikkilä, I. Kilpeläinen, S. Heikkinen, *J. Magn. Reson.* **2010**, Quantitative two-dimensional HSQC experiment for high magnetic field NMR spectrometers, 202, 24-33.

[117] S. Heikkinen, M. M. Toikka, P. T. Karhunen, I. A. Kilpeläinen, *J. Am. Chem. Soc.* **2003**, Quantitative 2D HSQC (Q-HSQC) via Suppression of J-Dependence of Polarization Transfer in NMR Spectroscopy: Application to Wood Lignin, 125, 4362-4367.

[118] H. Koskela, I. Kilpeläinen, S. Heikkinen, *J. Magn. Reson.* **2005**, Some aspects of quantitative 2D NMR, 174, 237-244.

[119] D. J. Peterson, N. M. Loening, *Magn. Reson. Chem.* **2007**, QQ-HSQC: a quick, quantitative heteronuclear correlation experiment for NMR spectroscopy, 45, 937-941.

[120] K. Hu, W. M. Westler, J. L. Markley, *J. Am. Chem. Soc.* **2011**, Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero 1H-13C HSQC (HSQC0), 133, 1662-1665.

[121] K. Hu, T. P. Wyche, T. S. Bugni, J. L. Markley, *J. Nat. Prod.* **2011**, Selective Quantification by 2D HSQC0 Spectroscopy of Thiocoraline in an Extract from a Sponge-Derived Verrucosispora sp, 74, 2295-2298.

[122] K. Cheng, H. Sorek, H. Zimmermann, D. E. Wemmer, M. Pauly, *Anal. Chem.* **2013**, Solution-State 2D NMR Spectroscopy of Plant Cell Walls Enabled by a Dimethylsulfoxide-d6/1-Ethyl-3-methylimidazolium Acetate Solvent, 85, 3213-3221.

[123] F. Hu, K. Furihata, Y. Kato, M. Tanokura, *J. Agric. Food. Chem.* **2007**, Nondestructive Quantification of Organic Compounds in Whole Milk without Pretreatment by Two-Dimensional NMR Spectroscopy, 55, 4307-4311.

[124] W. Gronwald, M. S. Klein, H. Kaspar, S. R. Fagerer, N. Nuürnberger, K. Dettmer, T. Bertsch, P. J. Oefner, *Anal. Chem.* **2008**, Urinary Metabolite Quantification Employing 2D NMR Spectroscopy, 80, 9288-9297.

[125] E. Martineau, K. El Khantache, M. Pupier, P. Sepulcri, S. Akoka, P. Giraudeau, *J. Pharm. Biomed. Anal.* **2015**, Non-linear effects in quantitative 2D NMR of polysaccharides: Pitfalls and how to avoid them, 108, 78-85.

[126] G. A. N. Gowda, F. Tayyari, T. Ye, Y. Suryani, S. Wei, N. Shanaiah, D. Raftery, *Anal. Chem.* **2010**, Quantitative Analysis of Blood Plasma Metabolites Using Isotope Enhanced NMR Methods, 82, 8983-8990.

[127] E. Martineau, P. Giraudeau, I. Tea, S. Akoka, *J. Pharm. Biomed. Anal.* **2011**, Fast and precise quantitative analysis of metabolic mixtures by 2D 1H INADEQUATE NMR, 54, 252-257.

[128] E. Martineau, I. Tea, S. Akoka, P. Giraudeau, *NMR Biomed.* **2012**, Absolute quantification of metabolites in breast cancer cell extracts by quantitative 2D 1H INADEQUATE NMR, 25, 985-992.

[129] C. Pungaliya, J. Srinivasan, B. W. Fox, R. U. Malik, A. H. Ludewig, P. W. Sternberg, F. C. Schroeder, *Proc. Natl. Acad. Sci.* **2009**, A shortcut to identifying small molecule signals that regulate behavior and development in Caenorhabditis elegans, 106, 7708-7713.

[130] S. T. Deyrup, L. E. Eckman, P. H. McCarthy, S. R. Smedley, J. Meinwald, F. C. Schroeder, *Proc. Natl. Acad. Sci.* **2011**, 2D NMR-spectroscopic screening reveals polyketides in ladybugs, 108, 9753-9758.

[131] R. R. Forseth, E. M. Fox, D. Chung, B. J. Howlett, N. P. Keller, F. C. Schroeder, *J. Am. Chem. Soc.* **2011**, Identification of Cryptic Products of the Gliotoxin Gene Cluster Using NMR-Based Comparative Metabolomics and a Model for Gliotoxin Biosynthesis, 133, 9678-9681.

[132] B. Bommarius, A. Anyanful, Y. Izrayelit, S. Bhatt, E. Cartwright, W. Wang, A. I. Swimm, G. M. Benian, F. C. Schroeder, D. Kalman, *PLoS ONE* **2013**, A Family of Indoles Regulate Virulence and Shiga Toxin Production in Pathogenic E. coli, 8, e54456.

[133] Y. Izrayelit, S. L. Robinette, N. Bose, S. H. von Reuss, F. C. Schroeder, *ACS Chem. Biol.* **2013**, 2D NMR-Based Metabolomics Uncovers Interactions between Conserved Biochemical Pathways in the Model Organism Caenorhabditis elegans, 8, 314-319.

[134] P. Mahanti, N. Bose, A. Bethke, J. C. Judkins, J. Wollam, K. J. Dumas, A. M. Zimmerman, S. L. Campbell, P. J. Hu, A. Antebi, F. C. Schroeder, *Cell Metab.* **2014**, Comparative Metabolomics Reveals Endogenous Ligands of DAF-12, a Nuclear Hormone Receptor, Regulating C.elegans Development and Lifespan, 19, 73-83.

[135] A. E. Taggi, J. Meinwald, F. C. Schroeder, *J. Am. Chem. Soc.* **2004**, A new approach to natural products discovery exemplified by the identification of sulfated nucleosides in spider venom, 126, 10364-10369.

[136] Y. H. Choi, H. K. Kim, H. J. M. Linthorst, J. G. Hollander, A. W. M. Lefeber, C. Erkelens, J.-M. Nuzillard, R. Verpoorte, *J. Nat. Prod.* **2006**, NMR Metabolomics to Revisit the Tobacco Mosaic Virus Infection in Nicotiana tabacum Leaves, 69, 742-748.

[137] A. Khatib, E. G. Wilson, H. K. Kim, A. W. M. Lefeber, C. Erkelens, Y. H. Choi, R. Verpoorte, *Anal. Chim. Acta* **2006**, Application of two-dimensional J-resolved nuclear magnetic resonance spectroscopy to differentiation of beer, 559, 264-270.

[138] H. T. Widarto, E. Meijden, A. W. M. Lefeber, C. Erkelens, H. K. Kim, Y. H. Choi, R. Verpoorte, *J. Chem. Ecol.* **2006**, Metabolomic Differentiation of Brassica rapa Following Herbivory by Different Insect Instars using Two-Dimensional Nuclear Magnetic Resonance Spectroscopy, 32, 2417-2428.

[139] J. M. Fonville, A. D. Maher, M. Coen, E. Holmes, J. C. Lindon, J. K. Nicholson, *Anal. Chem.* **2010**, Evaluation of Full-Resolution J-Resolved 1H NMR Projections of Biofluids for Metabonomics Information Retrieval and Biomarker Identification, 82, 1811-1821.

[140] M. Hedenström, S. Wiklund-Lindström, T. Öman, F. Lu, L. Gerber, P. Schatz, B. Sundberg, J. Ralph, *Mol. Plant* **2009**, Identification of Lignin and Polysaccharide Modifications in Populus Wood by Chemometric Analysis of 2D NMR Spectra from Dissolved Cell Walls, 2, 933-942.

[141] W. Y. Kang, S. H. Kim, Y. K. Chae, *FEMS Yeast Res.* **2012**, Stress adaptation of Saccharomyces cerevisiae as monitored via metabolites using two-dimensional NMR spectroscopy, 12, 608-616.

[142] M. Liebeke, J. G. Bundy, *Biochem. Biophys. Res. Commun.* **2012**, Biochemical diversity of betaines in earthworms, 430, 1306-1311.

[143] J. Yuk, M. J. Simpson, A. J. Simpson, *Ecotoxicology* **2012**, Coelomic fluid: a complimentary biological medium to assess sub-lethal endosulfan exposure using 1H NMR-based earthworm metabolomics, 21, 1301-1313.

[144] Y. Date, K. Sakata, J. Kikuchi, *Polym. J.* **2012**, Chemical profiling of complex biochemical mixtures from various seaweeds, 44, 888-894.

[145] T. Watanabe, A. Shino, K. Akashi, J. Kikuchi, *Plant Biotechnol.* **2012**, Spectroscopic investigation of tissue-specific biomass profiling for Jatropha curcas L, 29, 163-170.

[146] M. Yamamura, S. Noda, T. Hattori, A. Shino, J. Kikuchi, K. Takabe, S. Tagane, M. Gau, N. Uwatoko, M. Mii, *Plant Biotechnol.* **2013**, Characterization of lignocellulose of Erianthus arundinaceus in relation to enzymatic saccharification efficiency, 30, 25-35.

[147] Y. Date, Y. Nakanishi, S. Fukuda, Y. Nuijima, T. Kato, M. Umehara, H. Ohno, J. Kikuchi, *Food Chem.* **2014**, In vitro evaluation method for screening of candidate prebiotic foods, 152, 251-260.

[148] Y. Sekiyama, E. Chikayama, J. Kikuchi, *Anal. Chem.* **2010**, Profiling Polar and Semipolar Plant Metabolites throughout Extraction Processes Using a Combined Solution-State and High-Resolution Magic Angle Spinning NMR Approach, 82, 1643-1652.

[149] C. Tian, E. Chikayama, Y. Tsuboi, T. Kuromori, K. Shinozaki, J. Kikuchi, T. Hirayama, *J. Biol. Chem.* **2007**, Top-down Phenomics of Arabidopsis thaliana Metabolic Profiling by One- and Two-Dimensional Nuclear Magnetic Resonance Spectroscopy and Transcriptome Analysis of Albino Mutants, 282, 18532-18541.

[150] E. Chikayama, M. Suto, T. Nishihara, K. Shinozaki, T. Hirayama, J. Kikuchi, *PLoS ONE* **2008**, Systematic NMR Analysis of Stable Isotope Labeled Metabolite Mixtures in Plant and Animal Systems: Coarse Grained Views of Metabolic Pathways, 3, e3805.

[151] Y. J. An, W. J. Xu, X. Jin, H. Wen, H. Kim, J. Lee, S. Park, *ACS Chem. Biol.* **2012**, Metabotyping of the C. elegans sir-2.1 Mutant Using in Vivo Labeling and 13C-Heteronuclear Multidimensional NMR Metabolomics, 7, 2012-2018.

[152] Y. Sekiyama, E. Chikayama, J. Kikuchi, *Anal. Chem.* **2011**, Evaluation of a Semipolar Solvent System as a Step toward Heteronuclear Multidimensional NMR-Based Metabolomics for 13C-Labeled Bacteria, Plants, and Animals, 83, 719-726.

[153] S. Fukuda, H. Toh, K. Hase, K. Oshima, Y. Nakanishi, K. Yoshimura, T. Tobe, J. M. Clarke, D. L. Topping, T. Suzuki, *Nature* **2011**, Bifidobacteria can protect from enteropathogenic infection through production of acetate, 469, 543-547.

[154] Y. Nakanishi, S. Fukuda, E. Chikayama, Y. Kimura, H. Ohno, J. Kikuchi, *J. Proteome Res.* **2011**, Dynamic Omics Approach Identifies Nutrition-Mediated Microbial Interactions, 10, 824-836.

[155] T. Ye, S. Zhang, H. Mo, F. Tayyari, G. A. N. Gowda, D. Raftery, *Anal. Chem.* **2010**, 13C-Formylation for Improved Nuclear Magnetic Resonance Profiling of Amino Metabolites in Biofluids, 82, 2303-2309.

[156] Y. Xi, J. S. de Ropp, M. R. Viant, D. L. Woodruff, P. Yu, *Anal. Chim. Acta* **2008**, Improved identification of metabolites in complex mixtures using HSQC NMR spectroscopy, 614, 127-133.

[157] M. Hedenström, S. Wiklund, B. Sundberg, U. Edlund, *Chemometrics and Intelligent Laboratory Systems* **2008**, Visualization and interpretation of OPLS models based on 2D NMR data, 92, 110-117.

[158] J. S. McKenzie, A. J. Charlton, J. A. Donarski, A. D. MacNicoll, J. C. Wilson, *Metabolomics* **2010**, Peak fitting in 2D 1H-13C HSQC NMR spectra for metabolomic studies, 6, 574-582.

[159] Chenomx, 2010, Identifying Compounds by 2D Total Correlation Spectroscopy, accessed April 15 2015, <http://www.chenomx.com/news/img/page23/Chenomx.an010.pdf>

[160] M. Koda, K. Furihata, F. Wei, T. Miyakawa, M. Tanokura, *Magn. Reson. Chem.* **2011**, F2-selective two-dimensional NMR spectroscopy for the analysis of minor components in foods, 49, 710-716.

[161] S. L. Robinette, R. Ajredini, H. Rasheed, A. Zeinomar, F. C. Schroeder, A. T. Dossey, A. S. Edison, *Anal. Chem.* **2011**, Hierarchical Alignment and Full Resolution Pattern Recognition of 2D NMR Spectra: Application to Nematode Chemical Ecology, 83, 1649-1657.

[162] G. N. M. Reddy, S. Caldarelli, *Chem. Commun.* **2011**, Maximum-quantum (MaxQ) NMR for the speciation of mixtures of phenolic molecules, 47, 4297-4299

[163] G. N. M. Reddy, S. Caldarelli, *Analyst* **2012**, Identification and quantification of EPA 16 priority polycyclic aromatic hydrocarbon pollutants by Maximum-Quantum NMR, 137, 741-746.

[164] P. Giraudeau, N. Guignard, E. Hillion, E. Baguet, S. Akoka, *J. Pharm. Biomed. Anal.* **2007**, Optimization of homonuclear 2D NMR for fast quantitative analysis: Application to tropine-nortropine mixtures, 43, 1243-1248.

[165] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds, 127, 8014-8015.

[166] P. Schanda, *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, Fast-pulsing longitudinal relaxation optimized techniques: Enriching the toolbox of fast biomolecular NMR spectroscopy, 55, 238-265.

[167] P. Schanda, E. Kupce, B. Brutscher, *J. Biomol. NMR* **2005**, SOFAST-HMQC Experiments for Recording Two-dimensional Heteronuclear Correlation Spectra of Proteins within a Few Seconds, 33, 199-211.

[168] E. Lescop, T. Kern, B. Brutscher, *J. Magn. Reson.* **2010**, Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: Example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments, 203, 190-198.

[169] R. Ernst, W. Anderson, *Rev. Sci. Instrum.* **1966**, Application of Fourier transform spectroscopy to magnetic resonance, 37, 93-102.

[170] A. Ross, M. Salzmann, H. Senn, *J. Biomol. NMR* **1997**, Fast-HMQC using Ernst angle pulses: An efficient tool for screening of ligand binding to target proteins, 10, 389-396.

[171] P. Schanda, H. Van Melckebeke, B. Brutscher, *J. Am. Chem. Soc.* **2006**, Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes, 128, 9042-9043.

[172] E. Lescop, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2007**, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment, 187, 163-169

[173] F.-X. Theillet, A. Binolfi, S. Liokatis, S. Verzini, P. Selenko, *J. Biomol. NMR* **2011**, Paramagnetic relaxation enhancement to improve sensitivity of fast NMR methods: application to intrinsically disordered proteins, 51, 487-495.

[174] E. Kupce, R. Freeman, *Magn. Reson. Chem.* **2007**, Fast multidimensional NMR by polarization sharing, 45, 2-4.

[175] J. Furrer, *Chem. Commun.* **2010**, A robust, sensitive, and versatile HMBC experiment for rapid structure elucidation by NMR: IMPACT-HMBC, 46, 3396-3398.

[176] D. Schulze-Sünninghausen, J. Becker, B. Luy, *J. Am. Chem. Soc.* **2014**, Rapid Heteronuclear Single Quantum Correlation NMR Spectra at Natural Abundance, 136, 1242-1245.

[177] Lokesh, N. Suryaprakash, *Chem. Commun.* **2014**, Sensitivity enhancement in slice-selective NMR experiments through polarization sharing, 50, 8550-8553.

[178] B. Vitorge, G. Bodenhausen, P. Pelupessy, *J. Magn. Reson.* **2010**, Speeding up nuclear magnetic resonance spectroscopy by the use of SMAll Recovery Times - SMART NMR, 207, 149-152.

[179] D. Jeannerat, *J. Magn. Reson.* **2007**, Computer optimized spectral aliasing in the indirect dimension of 1H-13C heteronuclear 2D NMR experiments. A new algorithm and examples of applications to small molecules, 186, 112-122.

[180] B. Vitorge, Applications du repliement spectral en RMN à deux dimensions **2009**, Type thesis, University of Geneva (Geneva).

[181] E. Martineau, S. Akoka, R. Boisseau, B. Delanoue, P. Giraudeau, *Anal. Chem.* **2013**, Fast Quantitative 1H-13C Two-Dimensional NMR with Very High Precision, 85, 4777-4783.

[182] B. Vitorge, S. Bieri, M. Humam, P. Christen, K. Hostettmann, O. Munoz, S. Loss, D. Jeannerat, *Chem. Commun.* **2009**, High-precision heteronuclear 2D NMR experiments using 10-ppm spectral window to resolve carbon overlap, 950-952

[183] M. Foroozandeh, D. Jeannerat, *ChemPhysChem* **2010**, Deciphered Chemical Shifts in Aliased Spectra Recorded with two Slightly Different Narrow Windows or Differential Chemical Shift Evolution, 11, 2503-2505

[184] H. Barkhuijsen, R. de Beer, W. M. M. J. Bovée, D. van Ormondt, *J. Magn. Reson.* **1985**, Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure, 61, 465-481.

[185] R. C. Breton, W. F. Reynolds, *Nat. Prod. Rep.* **2013**, Using NMR to identify and characterize natural products, 30, 501-524.

[186] A. S. Stern, K.-B. Li, J. C. Hoch, *J. Am. Chem. Soc.* **2002**, Modern Spectrum Analysis in Multidimensional NMR Spectroscopy : Comparison of Linear-Prediction Extrapolation and Maximum-Entropy Reconstruction, 124, 1982-1993

[187] W. F. Reynolds, M. Yu, R. G. Enriquez, I. Leon, *Magn. Reson. Chem.* **1997**, Investigation of the Advantages and Limitations of Forward Linear Prediction for Processing 2D Data Sets, 35, 505-519.

[188] R. Brüschweiler, F. Zhang, *J. Chem. Phys.* **2004**, Covariance nuclear magnetic resonance spectroscopy, 120, 5253-5260.

[189] N. Trbovic, S. Smirnov, F. Zhang, R. Brüschweiler, *J. Magn. Reson.* **2004**, Covariance NMR spectroscopy by singular value decomposition, 171, 277-283.

[190] F. Zhang, R. Brüschweiler, *J. Am. Chem. Soc.* **2004**, Indirect Covariance NMR Spectroscopy, 126, 13180-13181.

[191] Y. Chen, F. Zhang, W. Bermel, R. Brüschweiler, *J. Am. Chem. Soc.* **2006**, Enhanced Covariance Spectroscopy from Minimal Datasets, 128, 15564-15565.

[192] O. Lafon, B. Hu, J.-P. Amoureux, P. Lesot, *Chemistry – A European Journal* **2011**, Fast and High-Resolution Stereochemical Analysis by Nonuniform Sampling and Covariance Processing of Anisotropic Natural Abundance 2D 2H NMR Datasets, 17, 6716-6724.

[193] S. Sibisi, *Nature* **1983**, Two-dimensional reconstructions from one-dimensional data by maximum entropy, 301, 134-136.

[194] J. C. Hoch, *J. Magn. Reson.* **1985**, Maximum entropy signal processing of two-dimensional NMR data, 64, 436-440.

[195] P. J. Hore, *J. Magn. Reson.* **1985**, NMR data processing using the maximum entropy method, 62, 561-567.

[196] E. D. Laue, J. Skilling, J. Staunton, *J. Magn. Reson.* **1985**, Maximum entropy reconstruction of spectra containing antiphase peaks, 63, 418-424.

[197] E. D. Laue, M. R. Mayger, J. Skilling, J. Staunton, *J. Magn. Reson.* **1986**, Reconstruction of phase-sensitive two-dimensional NMR spectra by maximum entropy, 68, 14-29.

[198] J. C. J. Barna, E. D. Laue, *J. Magn. Reson.* **1987**, Conventional and exponential sampling for 2D NMR experiments with application to a 2D NMR spectrum of a protein, 75, 384-389.

[199] J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling, S. J. P. Worrall, *J. Magn. Reson.* **1987**, Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments, 73, 69-77.

[200] J. C. Hoch, M. W. Maciejewski, M. Mobli, A. D. Schuyler, A. S. Stern, *Acc. Chem. Res.* **2014**, Nonuniform Sampling and Maximum Entropy Reconstruction in Multidimensional NMR, 47, 708-717.

[201] S. Hyberts, D. Frueh, H. Arthanari, G. Wagner, *J. Biomol. NMR* **2009**, FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation, 45, 283-294.

[202] V. Y. Orekhov, V. A. Jaravine, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, Analysis of non-uniformly sampled spectra with multi-dimensional decomposition, 59, 271-292.

[203] D. J. Holland, M. J. Bostock, L. F. Gladden, D. Nietlispach, *Angew. Chem.* **2011**, Fast multidimensional NMR spectroscopy using compressed sensing, 123, 6678-6681.

[204] K. Kazimierczuk, V. Y. Orekhov, *Angew. Chem. Int. Ed.* **2011**, Accelerated NMR spectroscopy by using compressed sensing, 50, 5556-5559.

[205] K. Kazimierczuk, A. Zawadzka, W. Kozminski, I. Zhukov, *J. Biomol. NMR* **2006**, Random sampling of evolution time space and Fourier transform processing, 36, 157-168.

[206] S. G. Hyberts, K. Takeuchi, G. Wagner, *J. Am. Chem. Soc.* **2010**, Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data, 132, 2145-2147.

[207] K. Nagayama, P. Bachmann, K. Wuthrich, R. R. Ernst, *J. Magn. Reson.* **1978**, The use of cross-sections and of projections in two-dimensional NMR spectroscopy, 31, 133-148.

[208] E. Kupce, R. Freeman, *J. Am. Chem. Soc.* **2004**, Projection-Reconstruction Technique for Speeding up Multidimensional NMR Spectroscopy, 126, 6429-6440.

[209] S. Hiller, F. Fiorito, K. Wüthrich, G. Wider, *Proc. Natl. Acad. Sci. USA* **2005**, Automated projection spectroscopy (APSY), 102, 10876-10881.

[210] E. Kupce, R. Freeman, *J. Magn. Reson.* **2003**, Two-dimensional Hadamard spectroscopy, 162, 300-310

[211] E. Kupce, T. Nishida, R. Freeman, *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, Hadamard NMR spectroscopy, 42, 95-122.

[212] E. Kupce, R. Freeman, *J. Magn. Reson.* **2003**, Frequency-domain Hadamard spectroscopy, 162, 158-165.

[213] P. Mansfield, *Magn. Reson. Med.* **1984**, Spatial mapping of the chemical shift in NMR, 1, 370-386

[214] L. Frydman, J. Peng, *Chem. Phys. Lett.* **1994**, Non-Cartesian sampling schemes and the acquisition of 2D NMR correlation spectra from single-scan experiments, 222, 371-377.

[215] L. Frydman, A. Lupulescu, T. Scherf, *J. Am. Chem. Soc.* **2003**, Principles and Features of Single-Scan Two-Dimensional NMR Spectroscopy, 125, 9204-9217

[216] P. Pelupessy, *J. Am. Chem. Soc.* **2003**, Adiabatic Single Scan Two-Dimensional NMR Spectrocopy, 125, 12345-12350.

[217] Y. Shrot, B. Shapira, L. Frydman, *J. Magn. Reson.* **2004**, Ultrafast 2D NMR spectroscopy using a continuous spatial encoding of the spin interactions, 171, 163-170.

[218] A. Tal, B. Shapira, L. Frydman, *J. Magn. Reson.* **2005**, A continuous phase-modulated approach to spatial encoding in ultrafast 2D NMR spectroscopy, 176, 107-114.

[219] Y. Shrot, A. Tal, L. Frydman, *Magn. Reson. Chem.* **2009**, New developments in the spatial encoding of spin interactions for single-scan 2D NMR, 47, 415-422.

[220] P. Giraudeau, S. Akoka, *J. Magn. Reson.* **2008**, Resolution and sensitivity aspects of ultrafast J-resolved 2D NMR spectra, 190, 339-345.

[221] S. Matsui, K. Sekihara, H. Kohno, *J. Am. Chem. Soc.* **1985**, High-speed spatially resolved high-resolution NMR spectroscopy, 107, 2817-2818.

[222] M. Pathan, S. Akoka, P. Giraudeau, *J. Magn. Reson.* **2012**, Ultrafast hetero-nuclear 2D J-resolved spectroscopy, 214, 335-339.

[223] P. Giraudeau, Y. Shrot, L. Frydman, *J. Am. Chem. Soc.* **2009**, Multiple Ultrafast, Broadband 2D NMR Spectra of Hyperpolarized Natural Products, 131, 13902-13903.

[224] P. Giraudeau, P. Lemeunier, M. Coutand, J.-M. Doux, A. Gilbert, G. S. Remaud, S. Akoka, *J. Spectrosc. Dyn* **2011**, Ultrafast 2D NMR applied to the kinetic study of D-glucose mutarotation in aqueous solution, 1, 1-7.

[225] A. Herrera, E. Fernandez-Valle, E. M. Gutierrez, R. Martinez-Alvarez, D. Molero, Z. D. Pardo, E. Saez, *Org. Lett.* **2009**, 2D Ultrafast HMBC: A Valuable Tool for Monitoring Organic Reactions, 12, 144-147

[226] Z. D. Pardo, G. L. Olsen, M. E. Fernandez-Valle, L. Frydman, R. Martinez-Alvarez, A. Herrera, *J. Am. Chem. Soc.* **2012**, Monitoring Mechanistic Details in the Synthesis of Pyrimidines via Real-Time, Ultrafast Multidimensional NMR Spectroscopy, 134, 2706-2715.

[227] L. H. K. Queiroz, P. Giraudeau, F. A. B. dos Santos, K. T. de Oliveira, A. G. Ferreira, *Magn. Reson. Chem.* **2012**, Real-time mechanistic monitoring of an acetal hydrolysis using ultrafast 2D NMR, 50, 496-501

[228] L. H. K. Q. Junior, D. P. K. Queiroz, L. Dhooghe, A. G. Ferreira, P. Giraudeau, *Analyst* **2012**, Real-time separation of natural products by ultrafast 2D NMR coupled to on-line HPLC, 137, 2357-2361.

[229] P. Giraudeau, S. Akoka, *J. Magn. Reson.* **2008**, Sources of sensitivity losses in ultrafast 2D NMR, 192, 151-158.

[230] P. Giraudeau, S. Akoka, *J. Magn. Reson.* **2008**, Sensitivity losses and line shape modifications due to molecular diffusion in continuous encoding ultrafast 2D NMR experiments, 195, 9-16

[231] P. Giraudeau, L. Frydman, *Annu. Rev. Anal. Chem.* **2014**, Ultrafast 2D NMR: An Emerging Tool in Analytical Spectroscopy, 7, 129-161.

[232] P. Giraudeau, S. Akoka, *Magn. Reson. Chem.* **2011**, Sensitivity and lineshape improvement in ultrafast 2D NMR by optimized apodization in the spatially encoded dimension, 49, 307-313.

[233] P. Pelupessy, L. Duma, G. Bodenhausen, *J. Magn. Reson.* **2008**, Improving resolution in single-scan 2D spectroscopy, 194, 169-174.

[234] Y. Shrot, L. Frydman, *J. Chem. Phys.* **2009**, Spatial/spectral encoding of the spin interactions in ultrafast multidimensional NMR, 131, 224516-224516-224511.

[235] P. Schanda, B. Brutscher, *J. Magn. Reson.* **2006**, Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR, 178, 334-339.

[236] M. Gal, P. Schanda, B. Brutscher, L. Frydman, *J. Am. Chem. Soc.* **2007**, UltraSOFAST HMQC NMR and the Repetitive Acquisition of 2D Protein Spectra at Hz Rates, 129, 1372-1377

[237] M. Gal, T. Kern, P. Schanda, L. Frydman, B. Brutscher, *J. Biomol. NMR* **2009**, An improved ultrafast 2D NMR experiment: Towards atom-resolved real-time studies of protein kinetics at multi-Hz rates, 43, 1-10

[238] S. Akoka, P. Giraudeau, *Magn. Reson. Chem.* **2015**, Fast hybrid multi-dimensional NMR methods based on ultrafast 2D NMR, in press.

[239] C. Ludwig, D. G. Ward, A. Martin, M. R. Viant, T. Ismail, P. J. Johnson, M. J. O. Wakelam, U. L. Günther, *Magn. Reson. Chem.* **2009**, Fast targeted multidimensional NMR metabolomics of colorectal cancer, 47, S68-S73.

[240] C. Pontoizeau, T. Herrmann, P. Toulhoat, B. Elena-Herrmann, L. Emsley, *Magn. Reson. Chem.* **2010**, Targeted projection NMR spectroscopy for unambiguous metabolic profiling of complex mixtures, 48, 727-733.

[241] M. Pathan, S. Akoka, I. Tea, B. Charrier, P. Giraudeau, *Analyst* **2011**, "Multi-scan single shot" quantitative 2D NMR: a valuable alternative to fast conventional quantitative 2D NMR, 136, 3157-3163.

[242] T. Jézéquel, C. Deborde, M. Maucourt, V. Zhendre, A. Moing, P. Giraudeau, *Metabolomics* **2015**, Absolute quantification of metabolites in tomato fruit extracts by fast 2D NMR, 1-12.

[243] P. Giraudeau, S. Massou, Y. Robin, E. Cahoreau, J.-C. Portais, S. Akoka, *Anal. Chem.* **2011**, Ultrafast Quantitative 2D NMR: An Efficient Tool for the Measurement of Specific Isotopic Enrichments in Complex Biological Mixtures, 83, 3112-3119.

[244] P. Giraudeau, E. Cahoreau, S. Massou, M. Pathan, J.-C. Portais, S. Akoka, *ChemPhysChem* **2012**, UFJCOSY: A Fast 3D NMR Method for Measuring Isotopic Enrichments in Complex Samples, 13, 3098-3101.

[245] R. Boisseau, B. Æ. Charrier, S. Massou, J.-C. Portais, S. Akoka, P. Giraudeau, *Anal. Chem.* **2013**, Fast Spatially Encoded 3D NMR Strategies for 13C-Based Metabolic Flux Analysis, 85, 9751-9757.

[246] F. Zhang, A. T. Dossey, C. Zachariah, A. S. Edison, R. Brüschweiler, *Anal. Chem.* **2007**, Strategy for Automated Analysis of Dynamic Metabolic Mixtures by NMR. Application to an Insect Venom, 79, 7748-7752.

[247] F. Zhang, L. Bruschweiler-Li, S. L. Robinette, R. Brüschweiler, *Anal. Chem.* **2008**, Self-Consistent Metabolic Mixture Analysis by Heteronuclear NMR. Application to a Human Cancer Cell Line, 80, 7549-7553.

[248] F. Kaplan, D. V. Badri, C. Zachariah, R. Ajredini, F. J. Sandoval, S. Roje, L. H. Levine, F. Zhang, S. L. Robinette, H. T. Alborn, W. Zhao, M. Stadler, R. Nimalendran, A. Dossey, R. Brüschweiler, J. M. Vivanco, A. Edison, *J. Chem. Ecol.* **2009**, Bacterial Attraction and Quorum Sensing Inhibition in Caenorhabditis elegans Exudates, 35, 878-892.

[249] F. Zhang, L. Bruschweiler-Li, R. Brüschweiler, *J. Am. Chem. Soc.* **2010**, Simultaneous de Novo Identification of Molecules in Chemical Mixtures by Doubly Indirect Covariance NMR Spectroscopy, 132, 16922-16927.

[250] K. Bingol, R. Brüschweiler, *Anal. Chem.* **2011**, Deconvolution of Chemical Mixtures with High Complexity by NMR Consensus Trace Clustering, 83, 7412-7417.

[251] K. Bingol, F. Zhang, L. Bruschweiler-Li, R. Brüschweiler, *J. Am. Chem. Soc.* **2012**, Carbon Backbone Topology of the Metabolome of a Cell, 134, 9006-9011.

[252] F. Zhang, L. Bruschweiler-Li, R. Bruschweiler, *J. Magn. Reson.* **2012**, High-resolution homonuclear 2D NMR of carbon-13 enriched metabolites and their mixtures, 225, 10-13.

[253] K. Bingol, F. Zhang, L. Bruschweiler-Li, R. Brüschweiler, *Anal. Chem.* **2013**, Quantitative Analysis of Metabolic Mixtures by Two-Dimensional 13C Constant-Time TOCSY NMR Spectroscopy, 85, 6414-6420.

[254] I. Bertini, S. Cacciatore, B. V. Jensen, J. V. Schou, J. S. Johansen, M. Kruhoffer, C. Luchinat, D. L. Nielsen, P. Turano, *Cancer Res.* **2012**, Metabolomic NMR Fingerprinting to Identify and Predict Survival of Patients with Metastatic Colorectal Cancer, 72, 356-364.

[255] G. Musumarra, D. F Condorelli, C. G Fortuna, *Comb. Chem. High T. Scr.* **2011**, OPLS-DA as a suitable method for selecting a set of gene transcripts discriminating RAS-and PTPN11-mutated cells in acute lymphoblastic leukaemia, 14, 36-46.

[256] K. Khan, M. Elia, *Clin. Nutr.* **1991**, Factors affecting the stability of L-glutamine in solution, 10, 186-192.

[257] Sigma-Aldrich, 2015, Glutamine in Cell Culture, accessed 26 February 2015, <http://www.sigmaaldrich.com/life-science/cell-culture/learning-center/media-expert/glutamine.html>

[258] M. A. McCoy, L. Mueller, *J. Am. Chem. Soc.* **1992**, Selective shaped pulse decoupling in NMR: homonuclear [carbon-13]carbonyl decoupling, 114, 2108-2112.

[259] V. Sklenar, *J. Magn. Reson., Ser A* **1995**, Suppression of radiation damping in multidimensional NMR experiments using magnetic field gradients, 114, 132-135.

[260] S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* **1993**, The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements, 115, 12593-12594.

[261] Y. K. Chae, S. H. Kim, J. E. Ellinger, J. L. Markley, *Bull. Korean Chem. Soc.* **2013**, Dosage Effects of Salt and pH Stresses on Saccharomyces cerevisiae as Monitored via Metabolites by Using Two Dimensional NMR Spectroscopy, 34, 3602-3608.

[262] A. Motta, D. Paris, D. Melck, *Anal. Chem.* **2010**, Monitoring real-time metabolism of living cells by fast two-dimensional NMR spectroscopy, 82, 2405-2411.

[263] M. Shukla, K. Dorai, *J. Magn. Reson.* **2011**, Resolving overlaps in diffusion encoded spectra using band-selective pulses in a 3D BEST-DOSY experiment, 213, 69-75.

[264] Y. Shrot, L. Frydman, *J. Chem. Phys.* **2008**, The effects of molecular diffusion in ultrafast two-dimensional nuclear magnetic resonance, 128, 164513-164513-164515.

[265] L. Rouger, B. Charrier, M. Pathan, S. Akoka, P. Giraudeau, *J. Magn. Reson.* **2014**, Processing strategies to obtain clean interleaved ultrafast 2D NMR spectra, 238, 87-93.

[266] P. Giraudeau, S. Akoka, *J. Magn. Reson.* **2007**, A new detection scheme for ultrafast 2D J-resolved spectroscopy, 186, 352-357.

[267] A. Le Guennec, P. Giraudeau, S. Caldarelli, *Anal. Chem.* **2014**, Evaluation of Fast 2D NMR for Metabolomics, 86, 5946-5954.

[268] M. Pathan, Methodological Developments in Ultrafast 2D NMR for Metabolic Studies, **2013**, Type thesis, University of Nantes (Nantes).

[269] T. H. Mareci, R. Freeman, *J. Magn. Reson.* **1983**, Mapping proton-proton coupling via double-quantum coherence, 51, 531-535.

[270] C. D. Ridge, L. Borvayeh, J. D. Walls, *J. Chem. Phys.* **2013**, Spatially encoded multiple-quantum excitation, 138, 204202.

[271] Y. Lin, Z. Zhang, S. Cai, Z. Chen, *J. Am. Chem. Soc.* **2011**, High-Resolution 2D J-Resolved Spectroscopy in Inhomogeneous Fields with Two Scans, 133, 7632-7635.

[272] Z. Zhang, H. Chen, C. Wu, R. Wu, S. Cai, Z. Chen, *J. Magn. Reson.* **2013**, Spatially encoded ultrafast high-resolution 2D homonuclear correlation spectroscopy in inhomogeneous fields, 227, 39-45.

[273] G. Wagner, E. R. P. Zuiderweg, *Biochem. Biophys. Res. Commun.* **1983**, Two-dimensional double quantum 1H NMR spectroscopy of proteins, 113, 854-860.

[274] A. Banerjee, N. Chandrakumar, *J. Chem. Phys.* **2014**, Communication: Ultrafast homonuclear correlation spectroscopy with diagonal suppressiona), 140, 231103.

[275] S. B. Reeder, E. Atalar, B. D. Bolster, E. R. McVeigh, **1997**, Quantification and reduction of ghosting artifacts in interleaved echo-planar imaging, 38, 439.

[276] M. Mishkovsky, M. Gal, L. Frydman, *J. Biomol. NMR* **2007**, Spatially encoded strategies in the execution of biomolecular-oriented 3D NMR experiments, 39.

[277] S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue, J. Staunton, *Nature* **1984**, Maximum entropy signal processing in practical NMR spectroscopy, 311, 446-447.

[278] D. L. Donoho, I. M. Johnstone, A. S. Stern, J. C. Hoch, *Proc. Natl. Acad. Sci. USA* **1990**, Does the maximum entropy method improve sensitivity?, 87, 5066-5068.

[279] J. A. Jones, P. J. Hore, *J. Magn. Reson.* **1991**, The maximum entropy method. Appearance and reality, 92, 363-376.

[280] M. R. Palmer, B. R. Wenrich, P. Stahlfeld, D. Rovnyak, *J. Biomol. NMR* **2014**, Performance tuning non-uniform sampling for sensitivity enhancement of signal-limited biological NMR, 1-12.

[281] C. Suiter, S. Paramasivam, G. Hou, S. Sun, D. Rice, J. Hoch, D. Rovnyak, T. Polenova, *J. Biomol. NMR* **2014**, Sensitivity gains, linearity, and spectral reproducibility in nonuniformly sampled multidimensional MAS NMR spectra of high dynamic range, 59, 57-73.

[282] S. Paramasivam, C. L. Suiter, G. Hou, S. Sun, M. Palmer, J. C. Hoch, D. Rovnyak, T. Polenova, *J. Phys. Chem. B* **2012**, Enhanced Sensitivity by Nonuniform Sampling Enables Multidimensional MAS NMR Spectroscopy of Protein Assemblies, 116, 7416-7427.

[283] S. G. Hyberts, S. A. Robson, G. Wagner, *J. Biomol. NMR* **2013**, Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra, 55, 167-178.

[284] R. Linser, B. Bardiaux, L. B. Andreas, S. G. Hyberts, V. K. Morris, G. Pintacuda, M. Sunde, A. H. Kwan, G. Wagner, *J. Am. Chem. Soc.* **2014**, Solid-State NMR Structure Determination from Diagonal-Compensated, Sparsely Nonuniform-Sampled 4D Proton-Proton Restraints, 136, 11002-11010.

[285] S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, G. Wagner, *J. Biomol. NMR* **2012**, Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling, 52, 315-327.

[286] D. Rovnyak, M. Sarcone, Z. Jiang, *Magn. Reson. Chem.* **2011**, Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling, 49, 483-491.

[287] C. J. Turner, P. J. Connolly, A. S. Stern, *J. Magn. Reson.* **1999**, Artifacts in Sensitivity-Enhanced HSQC, 137, 281-284.

[288] V. Dotsch, R. E. Oswald, G. Wagner, *J. Magn. Reson., Ser B* **1995**, Water Suppression by Coherence Selection with Absorptive Lineshape without Loss in Sensitivity, 108, 285-288.

[289] P. C. Aoto, R. B. Fenwick, G. J. A. Kroon, P. E. Wright, *J. Magn. Reson.* **2014**, Accurate scoring of non-uniform sampling schemes for quantitative NMR, 246, 31-35.

[290] S. G. Hyberts, H. Arthanari, G. Wagner, in *Novel Sampling Approaches in Higher Dimensional NMR, Vol. 316* (Eds.: M. Billeter, V. Orekhov), Springer Berlin Heidelberg, **2012**, pp. 125-148.

[291] L. Paudel, R. W. Adams, P. Király, J. A. Aguilar, M. Foroozandeh, M. J. Cliff, M. Nilsson, P. Sándor, J. P. Waltho, G. A. Morris, **2013**, Simultaneously Enhancing Spectral Resolution and Sensitivity in Heteronuclear Correlation NMR Spectroscopy, 52, 11619.

[292] X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, V. Orekhov, *Angew. Chem. Int. Ed.* **2015**, Accelerated NMR Spectroscopy with Low-Rank Reconstruction, 54, 852-854.

[293] Y. Nishiyama, M. H. Frey, S. Mukasa, H. Utsumi, *J. Magn. Reson.* **2010**, 13 C solid-state NMR chromatography by magic angle spinning 1 H T 1 relaxation ordered spectroscopy, 202, 135-139.

[294] U. Holzgrabe, I. Wawer, B. Diehl, *NMR spectroscopy in pharmaceutical analysis, Part II, chapter 6*, Elsevier, **2011**.

[295] M. Nilsson, *J. Magn. Reson.* **2009**, The DOSY Toolbox: A new tool for processing PFG NMR diffusion data, 200, 296-302.

[296] E. O. Stejskal, J. E. Tanner, *J. Chem. Phys.* **1965**, Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient, 42, 288-292.

[297] M. Urbanczyk, D. Bernin, W. Kozminski, K. Kazimierczuk, *Anal. Chem.* **2013**, Iterative Thresholding Algorithm for Multiexponential Decay Applied to PGSE NMR Data, 85, 1828-1833.

[298] M. M. Hoffmann, H. S. Sobstyl, S. J. Seedhouse, *Magn. Reson. Chem.* **2008**, T1 relaxation measurement with solvent suppression, 46, 660-666.

[299] M. M. Hoffmann, H. S. Sobstyl, V. A. Badali, *Magn. Reson. Chem.* **2009**, T2 relaxation measurement with solvent suppression and implications to solvent suppression in general, 47, 593-600.

[300] B. Jiang, F. Luo, Y. Ding, P. Sun, X. Zhang, L. Jiang, C. Li, X.-a. Mao, D. Yang, C. Tang, M. Liu, *Anal. Chem.* **2013**, NASR: An Effective Approach for Simultaneous Noise and Artifact Suppression in NMR Spectroscopy, 85, 2523-2528.

[301] N. Chandrakumar, *J. Indian I. Sci.* **2012**, Dynamic nuclear polarization in NMR, 90, 133-143.

[302] J. H. Ardenkjaer-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. H. Lerche, R. Servin, M. Thaning, K. Golman, *Proc. Natl. Acad. Sci.* **2003**, Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR, 100, 10158-10163.

[303] M. Yon, J. Lalande-Martin, T. Harris, I. Tea, P. Giraudeau, L. Frydman, *ScienceJet* **2015**, 13C NMR detection of metabolic mixtures enhanced by dynamic nuclear polarization, 4.

Appendices

Summary of materials and methods used during this thesis

These appendices will assemble all the parameters for NMR experiments, statistical analysis and the sample preparations in order to make the reproduction of experiments easier for readers. The order in the annex reflects the order in the manuscript.

Part 2. Testing conventional 1D and 2D NMR experiments for metabolomics

Preparation of synthetic samples for statistical analysis

In order to create the 40 samples from the information in table I, the following protocol was applied:

- 1 L of a solution of H_2O/D_2O (90/10) was prepared, along with NaH_2PO_4 and Na_2HPO_4, in order to obtain a phosphate buffer of 100 mM with a pH of 7. Sodium azide (around 4 mM) was also added to avoid bacterial proliferation and 0.5 mM of TrimetylSilylPropionate (TSP) for internal calibration of the chemical shift.

- From this solution, 30 stock solutions were prepared, each containing one metabolite at a concentration 30 times higher than the mean concentration given in table I.

- The 40 synthetic serum samples were obtained by dilutions from the 30 stock samples. Each individual concentration for a sample was obtained by using an homemade program aimed at multiplying the mean concentration by a random number, with all the random numbers for one metabolite centered around 1 and respecting the standard deviation for this metabolite.

- The NMR samples were obtained by putting 600 μL of a synthetic serum into a 5 mm NMR tube from Wilmad Lab Glass. The 40 samples were stored in a refrigerator at 5°C until further use.

Parameters for 1D ^{1}H NMR acquisition and processing

All the experiments described for 1D NMR and 2D conventional NMR experiments in this part have been acquired on a Bruker Avance I 600 MHz equipped with a 5 mm TXI cryoprobe triple resonance (^{1}H/^{13}C/^{15}N). For this study, the Bruker pulse sequence "noesypr1d" was used, since it is one of the most commonly used pulse sequence in metabolomics, as said in section 1.1 of part 1.[67-69] The mixing time was 1 ms, contrary to most cases where the mixing time is usually 100 to 150 ms. This value was chosen from a review of that pulse sequence, where short values were recommended.[69] Each spectrum was acquired with 4 dummy scans, 8 scans, 32768 points acquired during 2.2 s and an inter-scan delay of 3 s for relaxation and presaturation of the water peak. The spectral width was 12 ppm. The experimental time was 1 min and 3 s for each spectrum.

Before Fourier transformation (FT), the FID was weighted by an exponential function with a decay rate of 0.3 Hz, but no zero-filling was applied. After FT and phase correction, the baseline

was automatically corrected by a polynomial function of order 2 and the chemical shifts were calibrated to the TSP.

1D NMR bucketing

Areas of the 1D NMR spectra were extracted by manual bucketing into the integration tool of Topspin. Thanks to the DQF-COSY and the HSQC spectra, the attribution of each multiplet was known, and this knowledge was used during bucketing. Each time that a region of the spectrum contained an isolated multiplet or an isolated part of a multiplet, a bucket was centered on that region. With this mindset, 82 buckets were defined. Otherwise, a bucket was centered in a region with overlapping peaks, while still trying to reduce as much as possible the number of overlapping peaks in that bucket. The raw areas were normalized relatively to the total area, and then exported to SIMCA, where the analysis was similar to the one with the concentrations. The score plot and the loading plot for the OPLS-DA are presented in figure 2.3.

For this statistical analysis, 1 of the 40 samples was a clear outlier. After analysis of the data and the spectra, it turned out that this sample had a leucine concentration twice higher than intended. It has been probably an error during sample preparation, and the most likely reason is that during the dilution of the stock solutions, the sample has been spiked twice with the leucine solution. As a result, that sample was not included in any statistical analysis and only 39 of the 40 samples were used.

Parameters for 2D ^{1}H J-RES NMR acquisition and processing

The DSE-J-RES pulse sequence was obtained by modifying the Bruker pulse sequence with the addition of a second 180° hard pulse during the mixing time and a change in delays to agree with the pulse sequence shown in figure 1.8. All the DSE J-RES spectra were acquired with 17480 points in the direct dimension during 1.37 s and 32 points in the indirect dimension, with a spectral width of 11 ppm in the direct dimension and 50 Hz in the indirect dimension. 16 dummy scans were used for magnetization equilibrium and 8 scans for cycle phasing. The delay between scans was 3 s, which was also used for presaturation of the water peak. The shaped pulses were 90° SElective Decoupling Using Crafted Excitation (SEDUCE) pulses,[258] which lasted 41 ms in order to target the water peak. The experimental time for each spectrum was 21 min and 30 s.

After acquisition, the FID were weighted with a $\pi/2$ shifted sine-bell in each dimension, then the number of t_1 increments was doubled by linear prediction to reduce the width of the peaks in the indirect dimension, making the bucketing process easier. The FID was then zero-filled to 32768 points in the direct dimension and 128 points in the indirect dimension. After FT, tilting and symmetrization were applied to suppress the majority of the strong coupling artifacts. The chemical shifts were calibrated to the TSP and a baseline correction of order 1 was applied.

2D NMR bucketing and statistical analysis

For statistical analysis, the peak volumes were extracted by manual bucketing in the integration tool of Topspin. Then the data were normalized relatively to the total volume and exported to SIMCA. Data were scaled to the UV and analyzed with OPLS-DA. This strategy was used for all 2D spectra, both fast and conventional.

The manual bucketing was done with a similar mindset to the manual bucketing of the 1D spectrum. When a correlation peak was completely resolved, the entirety of the peak was taken as one bucket. A rectangular zone containing all the peak volume above the noise level was used to delimit the bucket. In case of overlap, if it was partial, the focus was to delimit a zone of the spectrum containing the volume from only one correlation peak. That fraction needed to include at least the center of the peak. The overlapped fraction constituted another bucket. If it was impossible to separate even a fraction of a correlation peak, the entirety of the overlapping signal was considered as only one bucket.

Parameters for 2D ^{1}H DQF-COSY NMR acquisition and processing

The Bruker sequence used for this study was “cosydfphpr”, with the phase cycling for the presaturation pulse taken from the pulse sequence “cosydfgpph19”, for reasons explained in the next section. 4096 points were acquired during 0.31 s, with a spectral width of 11 ppm for both dimensions. Sixteen dummy scans were executed for magnetization equilibrium, then 256 t_1 increments were acquired with and 8 scans per increment. The delay between pulses was 3 s, which was also used for presaturation of the water peak. The experimental time for each spectrum was 1 h 55 min and 19 s.

After acquisition, the FID was weighted with a sine-bell function in each dimension, then the number of t_1 increments was doubled by linear prediction. The FID was then zero-filled to 8192 points in the direct dimension and 1024 points in the indirect dimension. After FT, the chemical shifts were calibrated to the TSP and an automatic baseline correction of order 2 was applied in both dimensions. Phase correction was applied, but the spectrum was presented in magnitude mode in order to avoid the cancellation of positive and negative peaks in the total volume for the statistical analysis.

Parameters for 2D ^{1}H HSQC NMR acquisition and processing

Similar to the DQF-COSY, one of the most basic variants of HSQC has been chosen for this study, “hsqcgpph”, with the addition of presaturation during the inter-scan delay. During acquisition, 4096 points were recorded in the direct dimension during 0.31 s, with a spectral width of 11 ppm. 16 dummy scans were executed to obtain a steady state magnetization, then 8 scans were acquired for 128 t_1 increments. The spectral width for the ^{13}C dimension was 165 ppm. The inter-scan delay was 3 s, which was also used for presaturation. The delays during the INEPT transfer were 1.72 ms, which corresponds to an average $^{1}J_{C\text{-}H}$ of 145 Hz. $J_{C\text{-}H}$ decoupling during acquisition was done with the Globally optimized Alternating phase Rectangular Pulse (GARP) supercycle. The total experimental time was 57 min and 37 s.

After acquisition the FID were weighted with a π/2 shifted squared sine-bell in each dimension, then the number of t_1 increments was doubled by linear prediction. Then the FID was zero-filled to 1024 points in the indirect dimension. After FT, the chemical shifts were calibrated to the TSP, a baseline correction of order 2 was applied and the resulting spectrum was phase corrected.

Part 3. Evaluation of fast 2D experiments for metabolomics

Parameters for 2D NMR NUS DQF-COSY acquisition and processing

All the experiments described for 2D NUS NMR experiments in this part have been acquired on a Bruker Avance I 600 MHz equipped with a 5 mm TCI cryoprobe triple resonance ($^{1}H/^{13}C/^{15}N$). For the NUS DQF-COSY, the spectra were acquired in a similar manner to the conventional DQF-COSY in part 2. In short, the pulse sequence was the "cosydfphpr" with the phase cycling of "cosydfgpph19" for the presaturation pulse. 256 t_1 increments were acquired with 8 scans and an inter-scan delay of 3 s. With NUS, only 30% of t_1 increments were acquired, reducing the experimental time to 34 min and 36 s. The exponential sampling schedule was automatically generated by Topspin at the beginning of the acquisition, with the parameter "T1" set to 1 s.

The processing of NUS DQF-COSY was also similar to conventional DQF-COSY, with the inclusion of Compressed Sensing (CS) to reconstruct the missing points in the indirect FID. After CS, the direct and the indirect FID were weighted with a sine-bell function. 256 t_1 increments were added by linear prediction, to reduce the volume of peaks for an easier bucketing, similarly to the conventional counterpart. Finally the FID was zero-filled to 8192 points in the direct dimension and 1024 in the indirect dimension before FT. The chemical shifts were referenced to the TSP and a baseline correction with a polynomial function of order 2 was applied.

Parameters for 2D NMR NUS HSQC acquisition and processing

As with NUS DQF-COSY, the parameters of NUS HSQC were kept similar to those of conventional HSQC. Specifically, the pulse sequence "hsqcgpph" was chosen, with the addition of presaturation during the inter-scan delay. 128 t_1 increments were recorded with 8 scans and an inter-scan delay of 3 s. With NUS, only half of the t_1 increments were acquired with an exponential sampling schedule generated by Topspin at the beginning of the acquisition. The experimental time was 29 min and 15 s.

The processing of NUS HSQC spectra was also kept similar to conventional HSQC. At first, CS was used to reconstruct the missing points in the indirect FID. After CS, the direct and indirect FIDs were weighted with a π/2 shifted squared sine-bell function. The number of t_1 increments was doubled by linear prediction, to reduce the width of peaks for a better bucketing, similarly to the conventional counterpart. Then the indirect FID was zero-filled to 1024 points before FT. The chemical shifts were referenced to the TSP and a baseline correction with a polynomial function of order 2 was applied.

Parameters for 2D NMR UF COSY

The UF 2D experiments were acquired with a Bruker Avance I 500 MHz equipped with a 5 mm $^{1}H/^{13}C$ dual cryoprobe.

The M3S UF COSY was developed during my Master internship for quantification in complex mixtures[38], so the parameters were similar to that publication. For the acquisition, the pulse sequence was depicted in figure 1.21. The spatial encoding was performed by a pair of 180° smooth chirp pulses, each one lasting 15 ms with a sweep range of 60 kHz, with the simultaneous application of a 5.33 G/cm bipolar gradient pair (8.3% of the maximum gradient strength available). During the acquisition, 128 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 512 μs and had an amplitude of 52 G/cm (80% of the maximum gradient strength available). 1024 points were acquired for each gradient pair. The number of scans was 176 and the inter-scan delay 5 s, for an experimental time of 14 min and 57 s. This experimental time was optimized to obtain at least one correlation peak for all the metabolites in the mixture.

The processing was done using a home-written routine from the EBSI group,[37] used within the commercial software. Briefly, Inverse Fourier Tranformation (IFT) of the ultrafast dimension was performed, then an optimized Gaussian apodization was applied to eliminate non-idealities at the border of the detection coil and the noise from outside the coil.[232] A FT was applied afterwards to return to the ultrafast dimension. The conventional dimension was weighted with a $\pi/4$ shifted sine-bell function, and then zero-filled to 1024 before FT in this dimension. After FT, the chemical shifts were referenced to the two diagonal peaks from lactate (1.32 and 4.10 ppm) and a baseline correction of order 2 was applied.

Parameters for 2D NMR UF J-RES

The UF J-RES pulse sequence has a different way to acquire the data than most UF pulse sequences.[266] Compared to the original way to acquire the UF J-RES spectrum,[218] this allows the chemical shifts and the J-couplings to be already separated during the acquisition. Tilting the resulting spectrum is thus not needed. Compared to UF COSY, UF J-RES has several advantages. Since the spectral width in the conventional dimension is small, long gradient durations can be applied during the acquisition, to obtain a high spectral width in the ultrafast dimension. Since the gradient duration is high, the amplitude of the gradients during acquisition can be lowered, which is beneficial for sensitivity because the receiver bandwidth can be reduced, lowering the noise level in the process. The reduced amplitude during acquisition can also be used to improve the resolution in the indirect dimension by increasing the duration of spatial encoding. In order to limit the diffusion, a multi-echo excitation scheme can be used.[220, 230]

For these reasons, the UF J-RES was tested for metabolomics. The pulse sequence used has been shown in figure 1.19. Spatial encoding was performed with two pairs of smooth chirp pulses, each one lasting 15 ms, with a bandwidth of 60 kHz, with the simultaneous application of two bipolar gradients with an amplitude of 5.33 G/cm (8.2% of the maximum available strength). The acquisition was done with 128 gradients to refocus the dephasing caused by the spatial encoding.

Each gradient lasted 6.6 ms with an amplitude of 5.2 G/cm (8% of the maximum available strength). 4096 points were acquired for each gradient. Two successive acquisition gradients were separated with a hard 180° pulse. Within each scan, successive 180° pulses were phase-cycled (y; y; -y; -y) to avoid spurious simulated echoes. The number of scans was 96 and the inter-scan delay 5 s, for an experimental time of 9 min and 36 s.

The processing was done with the same home-written routine as UF COSY. The main difference was the apodization of the conventional dimension with a $\pi/4$ shifted squared sine-bell function. After FT, the chemical shifts were referenced to the two peaks from lactate and a baseline correction of order 2 was applied.

Parameters for the evaluation of T_1 with the inversion-recovery with selective pulses

This study was conducted on a Avance 600 MHz Bruker spectrometer with a 5 mm TXI cryoprobe and the sample was serine in water. For the IR with hard pulses, 12288 points were recorded during 1 s of acquisition, with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 30 or 40 s, for reasons explained in the main text. 12 different delays between the 90° and the 180° pulse were acquired, in seconds: 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 4, 8, 15, 30 and 40. Later, the delays 45 seconds and 50 seconds were added, which will be also explained later. The IR with selective pulses was recorded with the same parameters, with the use of a 13 ms Reburp 90° pulse and of a 14 ms 180° Eburp pulse. The results of the T_1 estimation for the two proton groups visible on the spectrum with different parameters are shown in figure 3.5. The T_1 for each peak was estimated with the T_1/T_2 Topspin module.

Parameters for the evaluation of BEST for the reduction of T_1 of small molecules in water

This study was also conducted on a Avance 600 MHz Bruker spectrometer with a 5 mm TXI cryoprobe and the sample was phenylalanine in water. Four experiments were conducted on this sample. The first was the IR with hard pulses, where 12288 points were recorded during 1 s of acquisition, with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 30 s and the 14 delays between the 90° and the 180° pulse used in the previous chapter were used during acquisition.

The three other experiments were IR with soft pulses, each selecting only one of the three groups of protons in phenylalanine, and each having the same parameters as the IR with hard pulses, except for the selective pulses, which were a Reburp pulse for the 180° pulse and a Gaussian pulse for the 90° pulse. The first experiment used a 180° pulse of 13 ms, a 90° pulse of 9 ms and was centered at 7.28 ppm for the aromatic peaks. The second experiment used a 180° pulse of 23 ms, a 90° pulse of 9 ms and was centered at 3.89 ppm for the peaks from the -CH- group. The third experiment used a 180° pulse of 23 ms, a 90° pulse of 5 ms and was centered at 3.11 ppm for the peaks from the $-CH_2-$ group. For each experiment, the T_1 was also estimated from the T_1/T_2 Topspin toolbox.

Part 4. Ultrafast double-quantum spectroscopy for diagonal-free 2D spectra

All the NMR experiments in part 4 were acquired with a Bruker Avance I 600 MHz equipped with a 5 mm TCI cryoprobe triple resonance ($^{1}H/^{13}C/^{15}N$).

Calibration of UF experiments for a 600 MHz

The calibration was made by recording a UF COSY spectrum of sucrose in water. Spatial encoding was obtained using the scheme described in Figure 1.23. Two chirp pulses of 15 ms with a sweep range of 10 kHz were used, together with two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). During the acquisition, 64 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 718 μs and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available). These parameters were calculated from the website developed by the EBSI group.[37] 1048 points were acquired for each gradient pair. Only one scan was acquired, but presaturation of the water peak during 2 s was added to obtain a clean spectrum.

The processing was done in a similar manner to the UF spectra in part C. Briefly; a home-written routine from the EBSI group was used, developed by Benoît Charrier,[37] where apodization of the ultrafast dimension was done with an optimized Gaussian function. For the apodization of the conventional dimension, a sine-bell function was used. After FT, the chemical shifts were referenced to those described in the HMDB database.

Acquisition and processing parameters for conventional and UF DQS experiments on a sucrose sample

The conventional DQS spectrum was acquired with the pulse sequence described in figure 4.2.a. The last pulse before acquisition was a 120° pulse instead of a 90° pulse to maximize the intensity of the correlation peaks, as described elsewhere.[107] The delay τ in the spin echo was fixed to a value of 25 ms, for an average $^{3}J_{H\text{-}H}$ of 10 Hz. Coherence selection was done by gradients of 1 ms, to select the echo pathway for the magnetization that was in the DQ coherence before the last pulse and became SQ coherence after the last pulse. The acquisition lasted 0.34 s with 4096 points acquired. 512 t_1 increments were recorded with two scans per increment and an inter-scan delay of 2.5 s for relaxation and presaturation of the water peak. The experimental time was 50 min and 41 s.

Before FT, the FID was weighted with a sine-bell function in both dimensions, and then the FID was zero-filled to 8192 points in the direct dimension. After FT, the chemical shifts were referenced to those of the HMDB in the direct dimension and to the sum of chemical shifts in the indirect dimension.

For the UF DQS, the pulse sequence is described in figure 4.2.c. Spatial encoding was done similarly to the UF COSY, with two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The delay τ in the spin echo was also fixed to a value of 25 ms, for an average $^3J_{H-H}$ of 10 Hz. The final hard pulse before acquisition was a 120° pulse, to maximize the intensity of the correlation peaks, and was flanked by the same gradients as conventional DQS for coherence selection. During the acquisition, 64 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 718 μs and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available), similarly to the UF COSY. Only one scan was acquired, but the water peak was presaturated during 2 s.

The processing was done using the same home-written routine as for UF COSY, with the same parameters. After FT, the chemical shifts were referenced to those in the conventional DQS spectrum.

Acquisition and processing parameters for conventional and UF DQSSY experiments on a sucrose sample

For the conventional DQSSY pulse sequence, in figure 4.4.a, the same parameters as for the conventional DQS were chosen. Briefly, 512 t_1 increments were acquired with 2 scans per increment and the inter-scan delay was 2.5 s. The delay τ was also fixed at 25 ms, and the acquisition lasted 0.34 s with 4096 points recorded during acquisition.

The processing was also similar to the conventional DQS spectrum. The FID was weighted with a sine-bell function in each dimension, and then zero-filled to 8192 points in the direct dimension. After FT, the spectrum was referenced to the peaks in the COSY spectrum.

For the UF DQSSY pulse sequence, in figure 4.4.c, the parameters are also similar to the UF DQS pulse sequence. Spatial encoding was done with two series of two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two pairs of two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The first two chirp pulses were placed before the last hard pulse, of 120°, and the last two after the last hard pulse. The delay τ in the spin echo was also fixed to a value of 25 ms, for an average $^3J_{H-H}$ of 10 Hz. During the acquisition, 64 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 718 μs and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available), similarly to the UF COSY. Only one scan was acquired, but the water peak was presaturated during 2 s.

The processing was also similar to the UF DQS spectrum, with the same home-written routine, with the same parameters as UF COSY. After FT, the chemical shifts were referenced to those in the conventional DQSSY spectrum.

Acquisition and processing parameters for the 2D NMR conventional and UF COSY on a complex mixture

For the UF COSY acquisition, the parameters were similar to the UF COSY of sucrose, in section 1.1. Spatial encoding was obtained using two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). During the acquisition, 32 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 1.436 ms and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available). These parameters were calculated from the website developed by the EBSI laboratory[37], using the same calculations introduced in section 1.1. 1048 points were acquired for each gradient pair. Interleaving was used to double the spectral width in both dimension, which resulted in the acquisition of 4 datasets, with 1 scan recorded by data set. The inter-scan delay was 5 s and 4 dummy scans were added to avoid as much as possible the apparition of ghost peaks. The total experimental time was 41 s.

The processing of the UF COSY spectra was also very similar to the processing of the UF spectrum of sucrose. A home-written routine was used, with an additional step to combine the four datasets into one FID. The apodization of the ultrafast dimension was done with an optimized Gaussian function. For the apodization of the conventional dimension, a sine-bell function was used. After FT, the chemical shifts were referenced to the two diagonal peaks of lactate.

Acquisition and processing parameters for the 2D NMR conventional and UF DQS on a complex mixture

The conventional DQS spectrum was acquired with the pulse sequence described in figure 4.2.a. The last pulse before acquisition was a 120° pulse to maximize the intensity of the correlation peaks, as described elsewhere.[107] The delay τ in the spin echo was fixed to a value of 25 ms, for an average $^3J_{H\text{-}H}$ of 10 Hz. Coherence selection was done by gradients of 1 ms, to select the echo pathway for the magnetization that was in the DQ coherence before the last pulse and became SQ coherence after the last pulse. The acquisition lasted 0.34 s with 4096 points acquired. 1024 t_1 increments were acquired with four scans per increment and an inter-scan delay of 2.5 s for relaxation and presaturation of the water peak. The experimental time was 3 h 23 min and 4 s.

Before FT, the FID was weighted with a sine-bell function in both dimensions, and then the FID was zero-filled to 8192 points in the direct dimension. After FT, the chemical shifts were referenced to the two peaks from lactate in the direct dimension and the sum of the chemical shifts of lactate in the indirect dimension.

For the UF DQS, the parameters for acquisition were similar to the parameters in the UF DQS spectrum of figure 4.2.b. Spatial encoding was done with two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The delay τ in the spin echo was also fixed to a value of 25 ms, for an average $^3J_{H\text{-}H}$ of 10 Hz. The final hard pulse before acquisition was a 120° pulse, to maximize the intensity of the correlation peaks, and was flanked by the same

gradients as conventional DQS for coherence selection. During the acquisition, 32 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 1.436 ms and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available), similarly to the UF COSY. Interleaving was used to double the spectral width in both dimensions, which resulted in the acquisition of 4 datasets, with 1 scan recorded per dataset. The inter-scan delay was 5 s and 4 dummy scans were added to avoid as much as possible the apparition of ghost peaks. The acquisition time was 41 s.

The processing was done using the same home-written routine as UF COSY, with the same parameters and the step to combine the 4 datasets into one FID. After FT, the chemical shifts were referenced to those in the conventional DQS spectrum.

Acquisition and processing parameters for the 2D NMR conventional and UF DQSSY on a complex mixture

For the conventional DQSSY pulse sequence, the same parameters as for the conventional DQS pulse sequence were chosen. 1024 t_1 increments were acquired with 4 scans per increment and the inter-scan delay was 2.5 s. The delay τ was also fixed at 25 ms, and the acquisition lasted 0.34 s with 4096 points recorded during acquisition. The acquisition time was 3 h 31 min and 48 s.

The processing was also similar to the conventional DQS spectrum. The FID was weighted with a sine-bell function in each dimension, and then zero-filled to 8192 points in the direct dimension. After FT, the spectrum was referenced to the peaks in the COSY spectrum.

For the UF DQSSY pulse sequence, the parameters are also similar to the UF DQS pulse sequence. Spatial encoding was done with two series of two chirp pulses of 15 ms and a sweep range of 10 kHz at the same time as two series of two gradients of opposite signs with an amplitude of 1.33 G/cm (2.5% of the maximum gradient strength available). The delay τ in the spin echo was also fixed to a value of 25 Hz, for an average $^1J_{H\text{-}H}$ of 10 Hz. During the acquisition, 32 gradients pairs were applied to refocus the dephasing introduced by the spatial encoding. Each acquisition gradient pair lasted 1.436 ms and had an amplitude of 42.4 G/cm (80% of the maximum gradient strength available). Interleaving was used to double the spectral width in both dimensions, which resulted in the acquisition of 4 datasets, with 1 scan recorded by data set. The inter-scan delay was 5 s and 4 dummy scans were added to avoid as much as possible the apparition of ghost peaks. The acquisition time was 41 s.

The processing was also similar to the UF DQS spectrum, with the same home-written routine, with the same parameters as UF COSY. After FT, the chemical shifts were referenced to those in the conventional DQSSY spectrum.

Part 5. Time-equivalent non-uniform sampling for complex-mixture analysis

All the NMR experiments in part 5 were acquired with a Bruker Avance I 600 MHz equipped with a 5 mm TCI cryoprobe triple resonance ($^{1}H/^{13}C/^{15}N$).

Analysis of sensitivity in sensitivity-enhanced NUS

This study was performed with a sample of valine at 1.26 mM in 600 μL of H_2O/D_2O (90/10), with 0.5 mM of TSP for chemical shift reference. The pulse sequence chosen was "hsqcetgpsp", an HSQC variant using 180° adiabatic pulses for the INEPT and retro-INEPT blocks in the heteronuclear channel (^{13}C in this case). The US HSQC spectrum was acquired with 256 t_1 increments, 8 scans per increment and an inter-scan delay of 2.45 s. During acquisition, 2048 points were acquired for an acquisition time of 0.155 s. The delay for the INEPT transfer was fixed to 1.72 ms for a mean value of $^{1}J_{C\text{-}H}$ of 145 Hz). Adiabatic refocusing was performed with chirp pulses of 500 μs with a bandwidth of 60 kHz and decoupling during acquisition was done with the GARP supercycle. The total experimental time was 1 h 30 min and 14 s.

Before FT, the FID was weighted with a $\pi/2$ shifted squared sine-bell function in both dimensions, then zero-filling was applied to 4096 points in the direct dimension and 512 points in the indirect dimension. After FT, the chemical shifts were referenced to the TSP and an automatic baseline correction with a polynomial function of order 2 was applied in both dimensions.

For the time-equivalent NUS spectra, the same parameters were applied, with different values of the percentage of NUS of 66%, 50%, 33%, 25%, 20% and 12.5%. The time savings given by recording less t_1 increments was used to record more scans, respectively 12, 16, 24, 32, 40 and 64. The experimental time remained at 1 h 30 min and 14 s and the same processing was applied.

For each spectrum, the Signal-to-Root-Mean-Square-Noise Ratio (SNR_{RMS}) was calculated for 2 of the 4 correlation peaks from valine, one the CH_3 peaks at 1.04 ppm in the proton dimension and the CH peak at 2.28 ppm. The S/N toolbox from Topspin was used to calculate the S/N_{RMS} in 1D slices.

Samples for the evaluation of resolution-enhanced NUS

We chose to work on model mixtures for which the intensity of the less sensitive peaks was very close to the noise level ($S/N_{RMS} < 10$). Since the sensitivity of the two pulse sequences is different, two samples were used. These two samples were dilutions of one of the samples used in part 2 and 3 for the metabolomics studies from 1D and 2D NMR spectra. For TOCSY, the sample was diluted 16 times, while for HSQC, the sample was diluted 4 times.

In order to be even closer to the metabolic profile of serum samples, glucose was added in each sample. Two stock samples of 125 mM and 250 mM of glucose were prepared with the solution of H_2O/D_2O (90/10) in phosphate buffer (100 mM) with 0.5 mM of TSP and 4 mM of NaN3 described in section 1.1 of part 2. Then 25 μL of the 125 mM glucose sample were added in

the sample for TOCSY, for a final concentration of 5 mM. Finally, 50 μL of the 250 mM glucose sample were added in the sample for HSQC, for a final concentration of 20 mM. The dilution of the other metabolites was taken into account and the final concentration for each metabolite in the two samples is shown in table III.

Parameters for acquisition and processing of NUS HSQC and TOCSY spectra for the evaluation of resolution-enhanced NUS

The important parameters for acquisition used in both SE HSQC and z-filtered (zf) TOCSY are summarized in table 1A.

Table 1A. Main experimental parameters for acquisition of HSQC and TOCSY spectra

Parameters	Sequence	D1 (s)	SW2 (ppm)	SW1 (ppm)	o1p (ppm)	DS	NS	AQ (s)	Sampling schedule
HSQC	Sensitivity enhanced with presaturation	3	11	165.6	4.702	16	8	0.155	Poisson gap
TOCSY	DIPSI with excitation sculpting and z-filter	3	11	11	4.694	16	2	0.155	Poisson gap

D1: delay for relaxation, SW: spectral width, o1p: offset of the transmitter frequency, DS: number of dummy scans, NS: number of scans per t_1 increment, AQ: acquisition time in the direct dimension

Along with these parameters, others specific to SE HSQC or zf-TOCSY have to be mentioned. For SE HSQC, the INEPT delays were fixed to 1.72 ms, corresponding to an average $^1J_{C\text{-}H}$ coupling constant of 145 Hz. Decoupling of $^1J_{C\text{-}H}$ during acquisition was done with the GARP supercycle and 2096 points were recorded during acquisition. The total experimental time was 1 h 50 min 25 s.

For zf TOCSY, a Decoupling In the Presence of Scalar Interactions 2 (DIPSI2) motif of 80 ms was used for the spin lock. A z-filter was added to suppress zero-quantum coherences, with a chirp pulse of 50 ms before the spin-lock and another chirp pulse of 30 ms after the spin-lock. Both chirp pulses had a 20 kHz bandwidth and were applied together with 2.3 G/cm gradients. Like SE HSQC, 2096 points were recorded during acquisition. The total experimental time was 58 min and 39 s.

To evaluate the impact of resolution-enhanced NUS for the two pulse sequences, different percentages of NUS were evaluated. These percentages are seen in table 2A, along with the corresponding compression rates and number of t_1 increments in the indirect dimension after the CS reconstruction for both pulse sequences. For each experiment, the sampling schedule was

constructed using a program to generate Poisson gap sampling schedules develop recently,[289] allowing us to choose a sampling schedule with little to no chances to generate sampling artifacts or unfaithfully reconstruct the peaks volumes[290]. The sampling schedule with the best score from the program was chosen for the experiment and was placed as a variable counter list in the Topspin folder.

Table 2A. Compression rates used for NUS spectra and respective % of NUS (% of points acquired on the spectrometer), TD_1 and maximum evolution time for each spectrum.

Compression rate	1	2	3	4	5	8	16	32	50	100
% NUS	100	50	33	25	20	12.5	6.25	3.125	2	1
TD_1 HSQC	256	512	776	1024	1280	2048	4096	8192	12800	25600
TD_1 TOCSY	512	1024	1552	2048	2560	4096	n/a	n/a	n/a	n/a
Maximum evolution time TOCSY (ms)	38.7	77.4	117	155	194	298	n/a	n/a	n/a	n/a
Maximum evolution time HSQC (ms)	5.12	10.2	15.5	20.5	25.6	39.4	85.4	171	256	512

For SE HSQC, after acquisition, the indirect dimension was reconstructed with the CS algorithm, except for the US spectrum. Then the FID were weighted with a $\pi/2$ shifted squared sine-bell function in both dimensions and the direct FID was zero-filled to 4096 points, while the indirect FID was zero-filled to 16384 points for all experiments, to evaluate fairly the volumes when the compression rate is changed. After FT, the chemical shifts were referenced to the peak of TSP and an automatic baseline correction with a polynomial function of order 2 was applied.

For zf TOCSY, the processing was similar to SE HSQC. After acquisition, the indirect dimension was reconstructed with the CS algorithm, except for the US spectrum. Then the FID were weighted with a $\pi/2$ shifted squared sine-bell function and zero-filled to 4096 points in both dimensions, to evaluate fairly the volumes when the compression rate is changed. After FT, the chemical shifts were referenced to the peak of TSP and an automatic baseline correction with a polynomial function of order 2 was applied.

Part 6. Estimation of T_1 and T_2 in complex mixtures

All the NMR experiments in part 6 were acquired with a Bruker Avance I 600 MHz equipped with a 5 mm TCI cryoprobe triple resonance ($^{1}H/^{13}C/^{15}N$).

First test for T_1 estimation: phenylalanine sample

In order to test the capacity to measure the T_1 from IR data, a first test was carried out with a sample with no overlapping peaks, in order to compare the estimated T_1 from the T_1 Toolbox to the estimated T_1 from the T_1/T_2 calculation module from Topspin. The same data that have been recorded for Figure 3.6 have been used for the test of the T_1 Toolbox. For this spectrum, 12288 points were recorded during 1 s of acquisition, with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 50 s, in order to let all the protons relax completely between peaks, as seen in the next figure. Twelve different delays between the 90° and the 180° pulses were used, in seconds: 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 4, 8, 15, 30, 40, 45 and 50.

T_1 estimations for a complex mixture in D_2O

For this test, the IR pulse sequence was used for a complex mixture similar to the samples used in part 2 and 3, but more concentrated (roughly 5 times more concentrated) and in D_2O instead of a buffered H_2O/D_2O (90/10) solvent. The parameters for the IR pulse sequence were the same as in section 1.3.1 of part 3. Briefly, 12288 points were recorded during 1 s of acquisition, with a spectral width of 10 ppm and 4 scans. The inter-scan delay was 50 s, in order to let all the protons relax completely between peaks, as seen in the next figure. Twelve different delays between the 90° and the 180° pulses were used, in seconds: 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 4, 8, 15, 30, 40, 45 and 50.

T_1 estimations for a complex mixture in H_2O/D_2O (90/10)

The pulse sequence is shown in figure 6.7.a, and has been tested with one of the synthetic serum samples used in part 2 and 3. Twenty delays were used between the 180° and the 90° pulses in seconds: 0.01, 0.02, 0.05, 0.07, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 1, 3, 6, 9, 15, 20, 25, 30 and 40. The inter-scan delay was 35 s, to allow a complete relaxation of the protons between scans. The last 3 seconds were also used to presaturate the water peak. The spectra are shown in figure 6.7.b for a variable delay of 40 s and in figure 6.7.c for a variable delay of 10 ms. The selective pulses were rectangular pulses of 2 ms for these spectra.

First test for T_2 estimation: valine sample

A sample of valine at 13 mM in H_2O/D_2O (90/10) was used for this test. Twelve different numbers of spin echoes were used: 4, 12, 20, 40, 60, 80, 100, 140, 200, 300, 400 and 800. Each 180° pulses in the spin echo were separated by 2 ms (so the delay τ in figure 1.4 was fixed at 1 ms).

The inter-scan delay was 15 s, to allow a complete relaxation of the protons between scans. The last 3 seconds were also used to presaturate the water peak. The acquired data were processed in the T_2 Toolbox, as presented in section 1.1.

T_2 estimations for a complex mixture in H_2O/D_2O (90/10)

This part of the study is still in progress, but partial results can still be shown. The most convincing results so far were obtained with the study of the complex mixture already used in section 2. Twelve different numbers of spin echoes were used: 2, 6, 10, 20, 30, 40, 70, 100, 150, 200, 400 and 800. Two successive 180° in the spin echo were separated by 2 ms (so the delay τ in figure 1.4 was fixed at 1 ms), with a 90° pulse between two spin echoes half the time. The inter-scan delay was 15 s, to allow a complete relaxation of the protons between scans. The last 3 seconds were also used to presaturate the water peak. The acquired data were processed in the T_2 Toolbox, as presented in section 1.1. The pulse sequence is shown in figure 6.10.a.