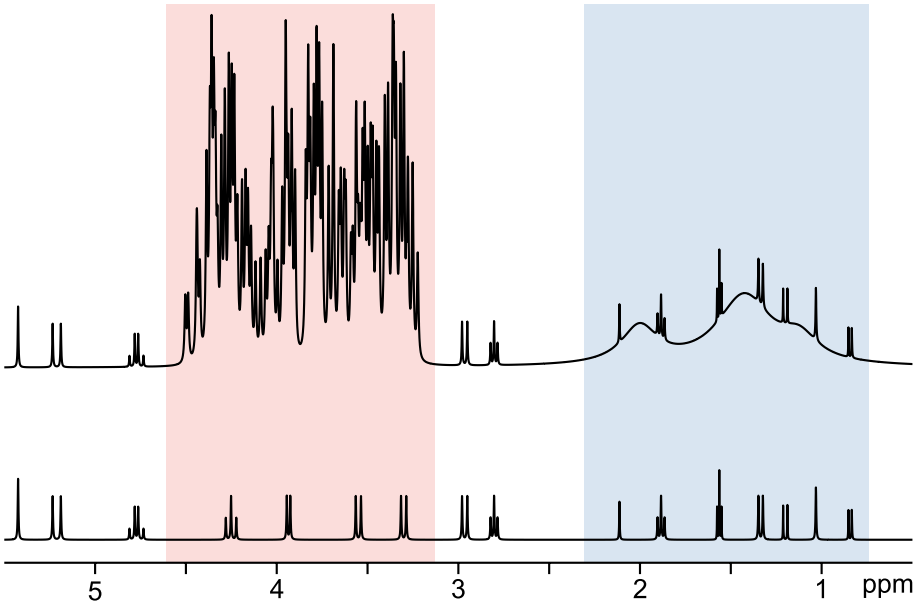
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:49
Faculty of Natural Resources and Agricultural Sciences

Reducing signal interference in complex NMR spectra

Characterization and quantification of metabolites in mixtures

Elin Alexandersson



Reducing signal interference in complex NMR spectra

Characterization and quantification of metabolites in mixtures

Elin Alexandersson
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Sciences
Uppsala

SLU
SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Uppsala 2024

Acta Universitatis Agriculturae Sueciae
2024:49

Cover: Illustration of an NMR spectrum with two types of signal interferences before (top) and after (bottom) application of the methods described in this thesis.
(Elin Alexandersson, 2024)

ISSN 1652-6880
ISBN (print version) 978-91-8046-038-5
ISBN (electronic version) 978-91-8046-039-2
https://doi.org/10.54612/a.1fkijr63lg

Swedish University of Agricultural Sciences, Department of Molecular Sciences, Uppsala, Sweden

Print: SLU Grafisk service, Uppsala 2024

Reducing signal interference in complex NMR spectra. Characterization and quantification of metabolites in mixtures

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a versatile analytical technique that can be used for identification, characterization, and quantification of various compounds. However, many sample types, including most biological samples, are mixtures of numerous different compounds with large differences in concentration and physico-chemical properties. When such samples are studied with NMR spectroscopy, they typically give rise to complicated NMR spectra that are difficult to analyse due to large dynamic range and extensive signal overlap. Methods are therefore needed that reduce signal overlap and other interferences in NMR spectra of complex mixtures so that individual compounds can be characterized and quantified. The work in this thesis revolves around two such methods.

One common form of signal interference is caused by intense, unwanted signals that overlap with, and sometimes obscure, signals of interest. Here, an NMR experiment was developed that enables selective suppression of the unwanted signals so that other signals can be studied. After evaluation and optimization of the experiment, it was used in the first complete NMR spectral assignment of the minor furanose forms of glucose.

Broad signals from lipids or macromolecules is another type of interference. Such signals are frequently encountered in NMR spectra of metabolomics samples, where they prevent accurate quantification of certain metabolites. In this thesis, an automated workflow was devised and optimized that first eliminates the interference from broad signals and then calculates absolute metabolite concentrations. The entire workflow was performed in less than one second per spectrum.

Keywords: NMR spectroscopy, signal interference, selective NMR experiments, metabolomics, baseline correction, automated quantification

Reducering av signalinterferens i komplexa NMR-spektra. Karakterisering och kvantifiering av metaboliter i blandningar

Sammanfattning

Kärnmagnetisk resonansspektroskopi, förkortat NMR-spektroskopi, är en mångfasetterad analysteknik som kan användas för identifiering, karakterisering och kvantifiering av olika substanser. Ett problem är att många provtyper, däribland de flesta biologiska prover, är blandningar av en mängd olika substanser med stora skillnader i koncentration och fysikalisk-kemiska egenskaper. När dessa prover studeras med NMR-spektroskopi ger de ofta upphov till komplicerade spektra som är svåra att analysera på grund av omfattande signalöverlapp och stora intensitetsskillnader mellan olika signaler. Det finns därför ett behov av metoder som på olika sätt minskar förekomsten av signalöverlapp och andra interferenser i NMR-spektra av komplexa blandningar så att enskilda substanser kan karakteriseras och kvantifieras. Den här avhandlingen handlar om två sådana metoder.

En vanlig form av signalinterferens orsakas av intensiva, oönskade signaler som överlappar med, och ibland helt täcker över, intressanta signaler. Därför utvecklades ett NMR-experiment som selektivt kan avlägsna oönskade signaler från NMR-spektra. Efter att experimentet utvärderats och optimerats användes det för att för första gången studera och beskriva alla NMR-signaler från de ovanliga furanosformerna av glukos.

En annan typ av interferens utgörs av breda signaler från lipider eller makromolekyler. Sådana signaler förekommer ofta i NMR-spektra som analyseras inom metabolomik och är problematiska eftersom de förhindrar absolutkvantifiering av vissa metaboliter. Den här avhandlingen beskriver ett automatiskt arbetsflöde som först tar bort interferenser från breda signaler och därefter beräknar korrekta metabolitkoncentrationer, allt på mindre än en sekund per spektrum.

Nyckelord: NMR-spektroskopi, signalinterferens, selektiva NMR-experiment, metabolomik, baslinjekorrektion, automatisk kvantifiering

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Elin Alexandersson, Corine Sandström, Lena C.E. Lundqvist and Gustav Nestor (2020). Band-selective NMR experiments for suppression of unwanted signals in complex mixtures. *RSC Advances*, 10 (54), 32511-32515.

II. Elin Alexandersson and Gustav Nestor (2022). Complete ^{1}H and ^{13}C NMR spectral assignment of D-glucofuranose. *Carbohydrate Research*, 511, 108477.

III. Elin Alexandersson, Corine Sandström, Johan Meijer, Gustav Nestor, Anders Broberg and Hanna E. Röhnisch (2024). Extended automated quantification algorithm (AQuA) for targeted ^{1}H NMR metabolomics of highly complex samples: application to plant root exudates. *Metabolomics*, 20, 11.

The contribution of Elin Alexandersson to the papers included in this thesis was as follows:

I. Project planning together with co-authors. Acquisition, analysis, and interpretation of all experimental data. Method optimization together with co-author. Wrote the majority of the manuscript and prepared all figures.

II. Project planning together with co-author. Planning, acquisition, analysis, and interpretation of all experimental data, including spin simulations. Wrote the majority of the manuscript, prepared all figures, and constructed the supplementary information. Submitting author and shared corresponding author.

III. Project planning together with co-authors. Planning, acquisition, analysis, and interpretation of all experimental data. Design and generation of simulated data together with co-author. Algorithm set-up together with co-author. Evaluation of method performance. Wrote the manuscript, prepared all figures, and constructed the supplementary information. Submitting and corresponding author.

Abbreviations

airPLS	Adaptive iteratively reweighted penalized least squares
AQuA	Automated quantification algorithm
BURP	Band-selective, uniform response, pure-phase
CPMG	Carr-Purcell-Meiboom-Gill
DOSY	Diffusion-ordered spectroscopy
DPFGSE	Double pulsed field gradient spin echo
DSS	Sodium 3-(trimethylsilyl)propane-1-sulfonate
FID	Free induction decay
GABA	γ-aminobutyric acid
HMBC	Heteronuclear multiple bond correlation
HSQC	Heteronuclear single quantum coherence
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PFG	Pulsed field gradient
RF	Radiofrequency
S/N	Signal-to-noise ratio
SMolESY	Small molecule enhancement spectroscopy

SPE	Solid phase extraction
SUN	Suppression of unwanted signals
TMS	Tetramethylsilane
TOCSY	Total correlation spectroscopy
TSP	3-(Trimethylsilyl)propanoic acid

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful, non-destructive analytical technique that can be used for structural elucidation, compound quantification, and to study dynamic processes. Therefore, NMR spectroscopy is an indispensable tool in scientific fields such as organic chemistry, analytical chemistry, structural biology, and medicine. However, the practical utility of the technique is often compromised by signal interference, or spectral overlap, that leads to complicated NMR spectra. In spectra with a high degree of signal interference, it is difficult to distinguish individual signals and thus the information presented by the signals about e.g. molecular structure or compound concentration cannot be utilized. The problem is frequently encountered when complex mixtures of small molecules are studied, and it is further exacerbated by large concentration differences between the mixture components. The NMR spectra of most biological samples consequently display extensive signal overlap.

The components of a mixture can be separated physically to reduce the sample complexity, but this is both costly and time consuming and affects the sample integrity. A more attractive alternative is to reduce the interference artificially, for instance by using certain NMR experiments or computational methods that remove unwanted signals from NMR spectra so that signals of interest can be studied. Two such approaches, one NMR experiment and one computational method, were investigated in this thesis.

The following three chapters aim to build a theoretical framework and to summarize previous research relevant to the work in this thesis. After a short description of the thesis aims, the research outcomes of papers **I-III** are presented and discussed in three successive chapters, followed by general conclusions and future perspectives.

2. Introduction to NMR spectroscopy

The theory behind NMR spectroscopy is quite complex and can only be fully described using quantum mechanics. Here, the principles will be presented in a simplified manner based on the vector model. As the work in this thesis has been performed using liquid-state ^{1}H and ^{13}C NMR spectroscopy, the following text will focus on these techniques.

2.1 Fundamental theory

As indicated by its name, NMR spectroscopy is a technique that focuses on atomic nuclei. Each type of atomic nucleus, i.e. each elemental isotope, can be characterized by a nuclear spin quantum number I that can be either equal to zero or a positive multiple of ½. Only nuclei with I greater than zero possess nuclear spin and can be observed in NMR spectroscopy, which excludes common isotopes such as ^{12}C and ^{16}O. The most studied nuclei in organic chemistry are instead ^{1}H, ^{13}C, and ^{15}N that all have $I = ½$.

Nuclei with spin (often simply called "spins") generate small magnetic fields, described by a vector called the magnetic moment (μ). Normally, the magnetic moments in a sample are randomly oriented due to thermal molecular motion. However, inside the NMR magnet the sample is exposed to a static, external magnetic field B_0, which makes it slightly more energetically favourable for the magnetic moments to be aligned in the direction of this field. The sample will thereby obtain a net magnetization, usually represented as a bulk magnetization vector, parallel to B_0. By convention, B_0 is depicted along the $+z$-axis in a right-handed Cartesian coordinate system. At equilibrium there is no net magnetization in the xy-plane.

In the presence of B_0, the magnetic moments begin to rotate about the +z-axis in a process called Larmor precession. The precession frequency, called the Larmor frequency, is proportional both to the B_0 field strength and to the magnetogyric (or gyromagnetic) ratio γ, a constant, fundamental property of each nucleus. All work in this thesis has been performed with $B_0 = 14.1$ T, which corresponds to a Larmor frequency of 600 MHz for ^{1}H and 150 MHz for ^{13}C.

2.1.1 Excitation and detection

Before an NMR signal can be recorded, the net magnetization needs to be excited, i.e. relocated from the +z-axis into the xy-plane where the detector is situated. To achieve this, a second magnetic field B_1 is applied as a radiofrequency (RF) pulse perpendicular to B_0. The magnitude of B_1 is much smaller than B_0, but because B_1 is oscillating at the Larmor frequency of the investigated nucleus, i.e. is resonant with the magnetic moments, it is able to shift the magnetization away from equilibrium. Once the magnetization moves into the xy-plane, the bulk magnetization vector as a whole starts to precess around the z-axis at the Larmor frequency, which induces a current in the detector. After the RF pulse, the magnetization gradually returns to equilibrium in a process known as relaxation. The recorded signal thus decays with time and is referred to as a time-dependent free induction decay (FID), which is the raw NMR data. Interpreting the FID is difficult and the data is therefore converted into a frequency-dependent spectrum through Fourier transformation, a mathematical process, before analysis.

2.1.2 Relaxation

The decay of the NMR signal with time is the result of two distinct relaxation processes. The longitudinal, or spin-lattice, relaxation T_1 corresponds to the restoration of magnetization along the +z-axis and occurs primarily because of interactions with local magnetic fields induced by neighbouring spins, so called dipole-dipole relaxation. As a result of the nuclear Overhauser effect (NOE), this type of relaxation can enhance the intensity of the signal arising from the neighbouring magnetic dipole.

The transverse, or spin-spin, relaxation T_2 describes the loss of net magnetization in the xy-plane that occurs due to magnetic field differences across the sample, caused either by intra- and intermolecular interactions or by inhomogeneities in the static magnetic field. For small molecules, T_1 and

T_2 are equally long whereas for large molecules T_2 is often much shorter than T_1. Because the signal line width is inversely proportional to T_2, NMR signals belonging to large molecules are broader than signals from small molecules.

2.1.3 Sensitivity

The intensity of the NMR response is influenced by several factors, including the nuclear spin quantum number, the magnetogyric ratio, and the natural abundance of the isotope in question. The magnitude of B_0 also influences the response, with a higher field strength resulting in both higher sensitivity and signal resolution. However, even with the highest field strengths available, the excess of magnetic moments aligned with B_0 is very small and NMR is therefore much less sensitive than most other spectroscopic methods.

The proton (^{1}H) is undoubtedly the most important nucleus in organic chemistry NMR studies as it is ubiquitously present in organic molecules, has a high γ, and a natural abundance of 99.99 %. In contrast, the NMR active isotope of carbon, ^{13}C, has a natural abundance of just 1.07 % and its γ is almost four times lower than that of ^{1}H. Therefore, ^{13}C NMR is far less sensitive than ^{1}H NMR. Nevertheless, ^{13}C data is required in most studies and different strategies are available to increase the sensitivity of ^{13}C NMR experiments. Molecules can be enriched in ^{13}C since a higher abundance both improves the sensitivity and allows for shorter experiment times. Alternatively, experiments can be utilized that detect ^{13}C resonances indirectly by taking advantage of the superior sensitivity of ^{1}H.

2.2 The NMR spectrum

The frequency-dependent data that is obtained when the FID has been Fourier transformed is the actual NMR spectrum (Figure 1). Usually, one isotope at a time is studied. The Larmor frequencies of individual atomic nuclei of the same isotope can differ slightly if the nuclei are situated in different local electrochemical environments, and therefore there is usually more than one signal in a spectrum even if only one compound is analysed. Nuclei surrounded by dense electron clouds are said to be shielded and resonate at a lower frequency than nuclei located next to electron withdrawing groups, because the latter nuclei are more exposed to the static magnetic field. These different nuclei are then observed as separate signals

in the NMR spectrum. Nuclei that resonate at identical frequencies cannot be distinguished from each other and are observed as one single signal. The area under each signal, the integral, is proportional to the number of nuclei giving rise to the signal and NMR spectroscopy can therefore be used for quantification purposes.

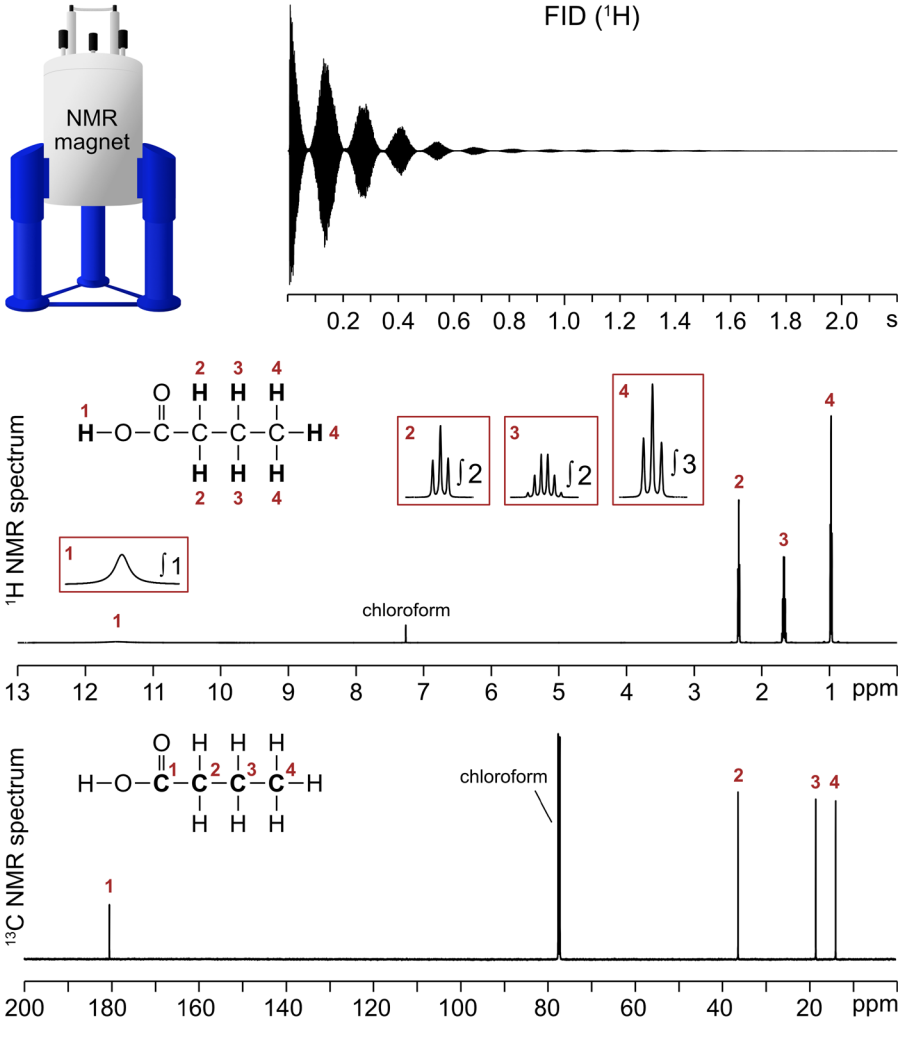


Figure 1. NMR analysis of butanoic acid in $CDCl_3$. Top: Illustration of an NMR magnet (left) and FID from the ^1H NMR analysis (right). Middle: 1D-^1H NMR spectrum. The insets display the splitting pattern and integral of each signal. Bottom: 1D-^{13}C NMR spectrum, acquired with proton decoupling and NOE enhancement.

In NMR spectroscopy, frequency differences between nuclei are of interest, rather than the actual frequencies. The signal positions, called chemical shifts (δ), are therefore reported in ppm (parts per million) as the frequency difference relative to a reference nucleus (Harris et al. 2001):

$$\delta = \frac{\nu_{signal} - \nu_{reference}}{\nu_{reference}}$$

where *v* is the frequency in hertz. Tetramethylsilane (TMS) has been defined as the primary reference compound for calculating chemical shifts, with an assigned chemical shift of 0 ppm for both ^{1}H and ^{13}C (Harris et al. 2001). In aqueous solutions, sodium 3-(trimethylsilyl)propane-1-sulfonate (DSS) or 3-(trimethylsilyl)propanoic acid (TSP) are commonly used instead due to the limited solubility of TMS (although the chemical shift of TSP is sensitive to pH (De Marco 1977)). Expressing the results in ppm enables comparison between spectrometers with different field strengths because the position of a certain signal in ppm will be the same despite the difference in actual frequency. ^{1}H NMR chemical shifts typically range between 0 and 12 ppm whereas the corresponding range for ^{13}C NMR signals is about 0-220 ppm.

NMR signals have different shapes, called splitting pattern, depending on their neighbouring spins. *J*-coupling, also called scalar or spin-spin coupling, is mediated through covalent chemical bonds and induces a splitting in the observed signal. The *J*-coupling can typically be observed over one, two, and three bonds, but more long-range couplings also exist. The size of the splitting, the coupling constant, is reported in hertz and is independent of the magnetic field strength. Coupling constants, chemical shifts, and signal integrals are all important indicators of the structure of a molecule. Both chemical shifts and coupling constants can often be read directly from an NMR spectrum with reasonable accuracy. However, to obtain the exact values it might be necessary to use computational line-shape analysis tools, especially for signals with complex splitting patterns or when signals overlap with each other.

To maximize sensitivity and take advantage of the NOE from neighbouring protons, ^{13}C NMR spectra are usually recorded with proton decoupling. As a consequence, ^{13}C signals typically appear as singlets and cannot be reliably integrated unless the spectra were recorded without NOE enhancement.

2.3 NMR experiments

During an NMR experiment, one or more RF pulses are applied to a sample to manipulate its magnetization in different ways. The number of RF pulses, as well as their respective amplitudes and durations, affect what type of information that can be obtained from the resulting NMR spectrum. Most NMR experiments also include time delays, the lengths of which further influence the outcome. Additionally, it is possible to apply magnetic field gradients during NMR experiments to create spatial differences in B_0 across the sample. A specific combination of RF pulses, time delays, and gradients is called a pulse sequence.

The simplest NMR experiment was described in section 2.1.1 and consists of a 90° pulse, i.e. an RF pulse that shifts the net magnetization from the +*z*-axis onto the *xy*-plane, followed by a delay during which the decaying signal is recorded (Figure 2). Many pulse sequences also contain 180° pulses, which invert the orientation of the magnetization (e.g. from the +*y*-axis to the -*y*-axis).

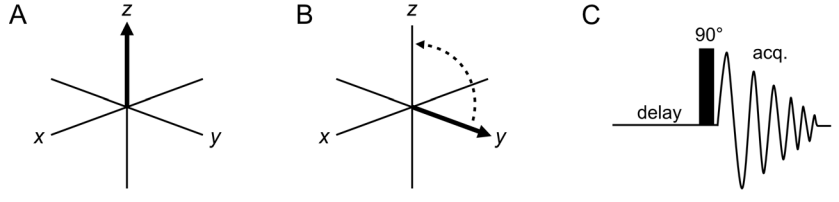


Figure 2. The basic NMR experiment. A) At equilibrium, the bulk magnetization is aligned with the +*z*-axis. B) After the RF pulse, the magnetization is located in the *xy*-plane and then gradually returns to equilibrium. C) Schematic illustration of the pulse sequence, which includes a relaxation delay, a 90° excitation pulse, and an acquisition time during which the decaying signal is recorded.

Because the NMR signal is quite weak, an NMR experiment is usually repeated several times and the FIDs summed together before Fourier transformation. This improves the signal-to-noise ratio (S/N) since the signals add up proportionally whilst the noise, because it is random, is not amplified as much. A relaxation delay in-between each repetition of the experiment allows the magnetization to return to its equilibrium position along the *z*-axis before the next round of excitation.

2.3.1 Hard and soft pulses

As described in section 2.2, individual nuclei of the same isotope have slightly different Larmor frequencies depending on their local electronic environment. It is therefore not possible to set the RF frequency to be exactly on resonance with all the spins in a sample. However, if the RF field is strong enough, a single RF pulse can still be used to excite all resonances in a sample simultaneously. Such a high-power RF pulse is called a "hard", or non-selective, pulse and these are the most commonly utilized pulses in NMR experiments.

The difference between a spin's Larmor frequency and the frequency of the RF field is called offset. When the RF field strength is much larger than the offset, the latter becomes negligible and the spin is excited just as if the pulse was on resonance. On the contrary, if the RF field strength is comparable to the size of the offset, the pulse will only excite signals over a limited frequency range. The weaker the RF field is, the narrower the excitation range becomes; it is thus possible to select what range of frequencies to excite by adjusting the RF field strength. RF pulses that act on a limited frequency span are called "soft", or selective, pulses. Soft pulses that target a range of frequencies, as opposed to a narrow span, are called band-selective.

The angle by which the magnetization vector is tipped away from the z-axis, the flip angle Θ, depends on both the power and the duration (t_p) of the RF pulse:

$$\Theta = 360 \times \gamma \times B_1 \times t_p \text{ degrees}$$

A low-power pulse thus needs to be longer than a high-power pulse to achieve the same flip angle. Consequently, soft pulses are much longer than hard pulses, typically being in the range of milliseconds and microseconds, respectively. From the equation above it can also be inferred that if the field strength is kept constant but the pulse length is increased, the flip angle will increase proportionally. Thus, a 180° pulse is twice as long as a 90° pulse with the same excitation profile.

Hard pulses are rectangular in their time-domain profile, meaning that the RF power is abruptly turned on and off when the pulse is applied. Selective pulses, on the other hand, are usually more elaborately shaped. There are numerous different types of shaped pulses with different excitation profiles.

Two examples are shown in Figure 3. The Gaussian shaped pulse (Bauer et al. 1984) has a narrow excitation profile and is therefore suitable for targeting small spectral regions. However, due to its broad tails, it will partly excite any resonances situated near the targeted region. Additionally, phase errors often appear in the resulting spectrum. There are other shaped pulses specifically designed to produce a "top-hat" excitation profile, with uniform, pure-phase excitation inside the selected region and negligible impact on the rest of the frequency range. One such example is the BURP (band-selective, uniform response, pure-phase) pulse family (Geen & Freeman 1991), illustrated in Figure 3 with the inversion variant I-BURP-2. The shapes of this kind of pulses are more intricate than the Gaussian pulse and so the pulses often need to be longer to selectively target a given bandwidth (Figure 3). Therefore, they are mostly used in band-selective applications as they can get excessively long when applied to narrow frequency ranges.

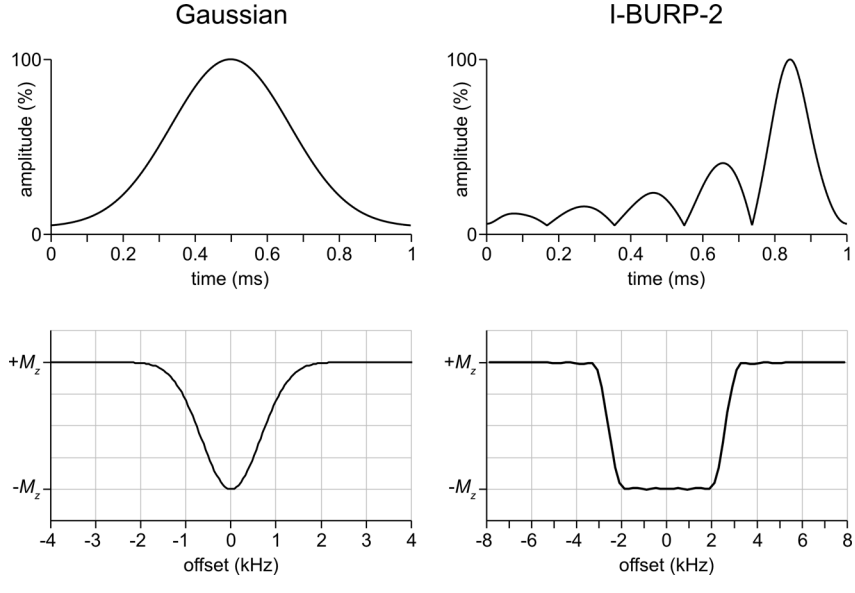


Figure 3. The pulse time domain profiles (top) and resulting magnetization inversion profiles (bottom) of a Gaussian and an I-BURP-2 selective inversion pulse, respectively. The length of both pulses is 1 ms. Note the difference in the inversion bandwidth obtained using the two pulses.

2.3.2 The spin echo

One of the most common building blocks in NMR pulse sequences is the spin echo, depicted in Figure 4. The spin echo contains a central 180° pulse surrounded by two time delays τ of equal length. Magnetization that is defocused during the first delay, due to frequency offset or B_0 inhomogeneities, is inverted by the 180° pulse and refocused by the end of the second delay.

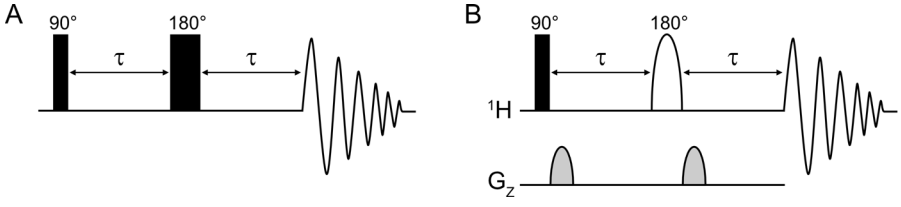


Figure 4. The spin echo pulse sequence, which can be performed without (A) or with PFGs (B) (trace labelled G_Z). The inversion pulse can be either hard (black rectangle) or soft (shaped white bar).

Spin echoes can be combined with pulsed field gradients (PFGs) for better performance and a wider field of application (Figure 4B). PFG spin echoes are commonly used e.g. for efficient selection of specific signals, in which case the 180° pulse usually is selective for the spectral region of interest, or to measure molecular diffusion. The two gradient pulses in a PFG spin echo are typically of equal amplitude and polarity, as in Figure 4B. For spins inverted by the 180° pulse, the defocusing in phase caused by the first gradient pulse is effectively reversed by the second gradient pulse and refocused by the end of the sequence. Remaining spins will experience the cumulative defocusing effect of both gradient pulses and will thus not appear in the resulting spectrum.

The PFG spin echo sequence can be further improved by the introduction of a second spin echo that uses a different gradient strength but where the sequence is otherwise the same (Figure 5). A double PFG spin echo (DPFGSE), or excitation sculpting, sequence results in spectra with significantly better phase properties than with a single echo (Hwang & Shaka 1995). In one common version of the DPFGSE sequence, the refocusing element consists of a soft and a hard 180° pulse in conjunction (Figure 5). The frequencies that are targeted by the selective pulse experience both 180° pulses, i.e. execute no net rotation, and therefore become defocused by the

gradient pulses. Remaining signals are refocused in the spin echoes and thus appear in the resulting spectrum. This approach thus enables selective suppression of unwanted signals and the pulse sequence is commonly used for solvent suppression, especially to suppress the signal from water.

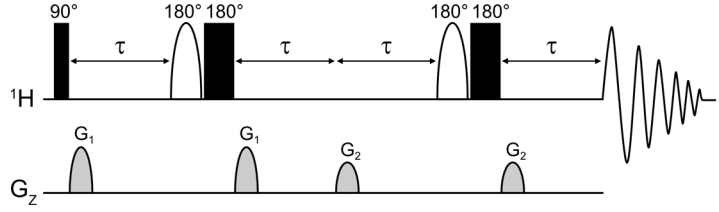


Figure 5. The DPFGSE sequence used for selective signal suppression (Hwang & Shaka 1995). Black rectangles represent hard pulses and shaped white bars represent selective pulses. PFGs are denoted G_Z.

2.3.3 Two-dimensional NMR experiments

NMR experiments can be further extended by the introduction of a second, indirect, frequency dimension. Thereby, connections between different nuclei are revealed, something that is essential in molecular structure elucidation. Many different 2D NMR experiments have been developed, making it possible to detect correlations through chemical bonds or through space, either between nuclei of the same isotope (homonuclear experiments) or between different types of nuclei (heteronuclear experiments). An additional benefit of 2D NMR spectra is that the signals are more dispersed than in the corresponding 1D spectra, thus problems with spectral overlap are less frequent.

One common 2D experiment is the total correlation spectroscopy (TOCSY) experiment (Braunschweiler & Ernst 1983), which is a homonuclear ^{1}H NMR method used to reveal connections between all protons in a spin system. TOCSY includes a mixing scheme, also called spin-lock, during which magnetization is transferred throughout a chain of *J*-coupled spins, thereby generating cross-peaks even between spins that are not directly coupled to each other.

The heteronuclear single quantum coherence (HSQC) experiment (Bodenhausen & Ruben 1980) displays one-bond connections between ^{1}H and ^{13}C (or ^{15}N) atoms. In HSQC, ^{1}H is the observed nucleus whilst the ^{13}C signals are detected indirectly. The indirect detection improves the ^{13}C

sensitivity substantially compared to when ^{13}C is detected directly, which it is in a 1D-^{13}C spectrum. Two- and three-bond ^{1}H-^{13}C connections can be detected with the heteronuclear multiple-bond correlation (HMBC) experiment (Bax & Summers 1986) which also uses indirect detection of ^{13}C.

Most 2D NMR experiments can be transformed into analogous 1D experiments that are faster to record and have higher digital resolution. The selective 1D-TOCSY experiment (Davis & Bax 1985) starts by exciting one signal of interest using a selective pulse. During the subsequent spin-lock, magnetization is transferred from the excited signal to its coupled neighbours until the magnetization has been propagated to the entire spin system. After Fourier transformation, only signals that have received magnetization from the initially excited spin will be visible in the spectrum. The length of the mixing scheme affects how far the magnetization is transferred from the excited spin. Thereby, 1D-TOCSY can be used for step-wise assignment of a spin system by recording a series of spectra with incremented mixing time.

2.4 Quantitative NMR spectroscopy

It is often stated that NMR spectroscopy is an inherently quantitative technique; the intensity of the response is directly proportional to the number of underlying nuclei. While this is in principle correct, in reality the quantitative quality is often compromised by the pulse sequences and parameters used for spectra acquisition. For a spectrum to be truly quantitative, a number of criteria need to be fulfilled. The 90° pulse has to be calibrated prior to acquisition to ensure uniform excitation throughout the spectral frequency range. During data collection, the relaxation delay must be sufficiently long for the magnetization to relax fully between each repetition of the experiment. Ideally, the relaxation delay should be at least five times longer than the T_1 of the slowest relaxing nucleus in the sample. Furthermore, the spectrum needs to be collected with a sufficient digital resolution and S/N so that signals can be reliably integrated. The latter is usually improved by increasing the number of experiment transients.

Because of its high sensitivity, 1D-^{1}H NMR is used in the vast majority of quantitative NMR studies. As mentioned in section 2.2, ^{13}C NMR spectra are usually not quantitative due to the desirable NOE enhancement achieved by proton decoupling. Quantitative ^{13}C NMR spectra can be obtained if the NOE is eliminated, for example by using inverse-gated decoupling where the

decoupling is applied only during the acquisition time. However, the low sensitivity of ^{13}C NMR makes this isotope much less suitable than ^{1}H for quantitative applications. 2D NMR experiments can also be used and a number of pulse sequences for e.g. quantitative HSQC have been developed (Heikkinen et al. 2003; Peterson & Loening 2007; Hu et al. 2011). Because peak intensities in 2D spectra are dependent not only on the number of nuclei but also on spin-specific coupling constants and T_2 relaxation times, calibration with each quantified compound as a pure reference needs to be performed for absolute quantification based on 2D spectra. Furthermore, 2D spectra take longer to acquire and analyse than the corresponding 1D spectra.

3. NMR-based metabolomics

3.1 Introduction to metabolomics

The entire set of metabolites, i.e. molecules of low molecular weight, present in a particular cell, tissue, or organism at a certain time is called metabolome (Oliver et al. 1998; Tweeddale et al. 1998). The research field aimed at studying the metabolome is consequently termed metabolomics (Raamsdonk et al. 2001) or, in certain instances, metabonomics (Nicholson et al. 1999). The number of metabolites is known to vary widely between different organisms, although no metabolome has yet been fully characterized. Plant metabolomes are especially complex with numerous different metabolites, both primary and secondary (also called specialized), and large variations in chemical structures and molar concentrations amongst the metabolites (Deborde et al. 2017).

Compared to the genome or proteome, the metabolome is more closely connected to the biochemical activity and thus the current state of a cell or an organism. Because the metabolome is dynamic, i.e. its set-up varies as a result of internal processes and external influences, metabolomics is a useful tool for as diverse applications as drug toxicity assessment (Nicholson et al. 1999), disease risk biomarker discovery (Holmes et al. 2008), and plant breeding (Razzaq et al. 2022). The samples analysed in metabolomics consist of various types of biofluids and tissues, although the most commonly studied sample types are human blood and urine.

There are two main analytical approaches in metabolomics, commonly called targeted and untargeted, respectively. Targeted metabolomics refers to the quantification of a number of preselected metabolites, whereas the aim of untargeted metabolomics is to investigate as many metabolites as possible without prior identification. The analytical techniques predominantly used in

the metabolomics field are NMR spectroscopy and mass spectrometry (MS), the latter typically preceded by liquid or gas chromatography. Most studies use only one of these techniques, although NMR and MS are often considered complimentary due to their respective strengths and weaknesses. Whilst NMR spectroscopy is non-destructive and highly reproducible, MS is more sensitive and able to detect a higher number of metabolites in a given sample. In an NMR spectrum recorded under quantitative conditions, signals belonging to different compounds are scaled according to their molar concentrations. Therefore, NMR spectroscopy is particularly suitable for quantitative metabolomics.

3.2 Data generation

There is currently no unified convention in the NMR-based metabolomics field regarding sample preparation, acquisition parameters, or spectral processing, although standardized protocols have been proposed (Beckonert et al. 2007; Kim et al. 2010; Emwas et al. 2016, 2018; Deborde et al. 2019). Common practices are described in the following sections.

3.2.1 Sample preparation

One major advantage of NMR spectroscopy as an analytical technique in metabolomics is that it can be applied after very little sample preparation. Many biofluids are prepared by the addition of a buffer solution, to control the pH of the samples, and possibly an internal standard for chemical shift referencing. When tissues are analysed with solution-state NMR spectroscopy, it is usually necessary to extract metabolites from the tissue using different solvents (Beckonert et al. 2007; Kim et al. 2010). Samples are generally kept cold (-80 °C) before analysis.

In addition to metabolites, many sample types also contain compounds such as proteins, lipids, polysaccharides, and nucleic acids. These compounds give rise to broad signals in NMR spectra and may also physically interact both with certain metabolites and with internal standard compounds. The presence of lipids and macromolecules is thus problematic for metabolite quantification. Macromolecules can be physically removed from a sample before NMR analysis using either ultrafiltration or precipitation with organic solvents (Daykin et al. 2002; Nagana Gowda & Raftery 2014). These techniques are routinely applied to blood samples,

although results are inconclusive regarding their effect on sample metabolite composition (Daykin et al. 2002; Tiziani et al. 2008; Nagana Gowda & Raftery 2014; Madrid-Gambin et al. 2023).

3.2.2 Quantification references

Calibration with a reference with known concentration is necessary for accurate absolute quantification. In NMR spectroscopy, as opposed to MS, a single reference compound can be used to quantify all other compounds in a spectrum. The reference can be an external standard compound (Burton et al. 2005; Wider & Dreier 2006) or even an electronically generated signal (Barantin et al. 1997; Akoka et al. 1999), but internal standards are the most common in metabolomics.

Over the years, a number of different substances have been used as internal standards in quantitative NMR analyses. A good reference compound should give rise to an isolated NMR signal, preferably a singlet, that can be used for quantification. Furthermore, the reference compound should not have other signals that overlap with metabolite signals, and the compound itself must not interact with any components of the sample.

The most commonly used internal standard compounds in NMR-based metabolomics are the chemical shift references DSS and TSP. The partially deuterated forms of DSS and TSP both produce a single, relatively intense, singlet at 0 ppm. However, both DSS and TSP are known to interact with certain aromatic compounds and can therefore not be used for reliable quantification in samples that contain e.g. macromolecules (Hand & Cohen 1965; Lam & Kotowycz 1977; Bell et al. 1989; Shimizu et al. 1994). Other reference compounds have been proposed, including 4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate (Nowick et al. 2003; Alum et al. 2008) and the naturally occurring metabolites formic acid (Kriat et al. 1992), fumaric acid, and maleic acid (Rundlöf et al. 2010; Nagana Gowda et al. 2021), but these are not broadly used in quantitative NMR-based metabolomics.

3.2.3 Spectra acquisition

The most commonly analysed data in NMR-based metabolomics is 1D-^{1}H NMR spectra, although other nuclei as well as 2D NMR experiments have been used in some studies. Most sample types are aqueous solutions and therefore require the use of water suppression schemes for the metabolites to

be observable in the NMR spectra. The most frequently used NMR experiment is the 1D NOE spectroscopy (NOESY) presaturation sequence since it offers good water suppression while still being relatively robust and easy to set up. Broad signals from lipids or macromolecules can be spectroscopically filtered away using the Carr-Purcell-Meiboom-Gill experiment (CPMG) (Carr & Purcell 1954; Meiboom & Gill 1958) or 1D diffusion-edited experiments (see Chapter 4), but these approaches are not necessarily quantitative and may increase the experiment time.

Metabolomics data sets often contain hundreds or even thousands of samples. To enable high sample throughput, the experiments are kept short and are extensively automated. The limit of detection and limit of quantification are often defined as 3 × S/N and 10 × S/N, respectively, hence the number of experimental transients needs to be high enough to ensure sufficient sensitivity. Normally, at least 64 transients are recorded in each experiment, although this needs to be adjusted for each sample type (Emwas et al. 2016). Currently, both the limit of detection and limit of quantification of NMR spectroscopy are in the micromolar range (Wishart et al. 2022).

For an NMR spectrum to be quantitative, the relaxation delay needs to be at least 5 × T_1 for all metabolite signals. Adhering to this criterion results in very long experiments since most metabolites have T_1 values of several seconds. The common metabolite formic acid, for example, has a ^{1}H T_1 of around 8 seconds at 600 MHz (Saude et al. 2006), meaning that the delay must be over 40 seconds for it to relax fully. To reduce the experimental time, shorter relaxation delays are usually used after which a T_1 correction factor is applied before calculating the concentrations (Saude et al. 2006; Bharti et al. 2008). Alternatively, the T_1 times can be shortened by the addition of a paramagnetic relaxation agent to the sample (Mulder et al. 2019).

3.2.4 Spectral processing

In addition to Fourier transformation, a number of processing steps need to be applied to the NMR data before it can be analysed. The processing procedure usually includes zero filling to increase the number of data points, exponential line broadening for improved S/N, phase correction, baseline correction, and chemical shift referencing. Spectral processing is most often carried out in a semi-automated manner, although the entire workflow can

be automated to save time and avoid user bias (Ravanbakhsh et al. 2015; Rout et al. 2023).

Correct phase and a flat spectral baseline are prerequisites for accurate quantification. Unfortunately, both phase correction and baseline correction can be difficult to execute properly, especially for complex spectra. Automatic phase correction is commonly used but is often followed by manual fine-tuning (Emwas et al. 2018). Similarly, the starting points for baseline correction are usually set manually by defining signal-free regions that can be used to calculate the fitted baseline. Several automatic baseline correction functions have been proposed (Golotvin & Williams 2000; Cobas et al. 2006; Zhang et al. 2010; Bao et al. 2012; Wang et al. 2013; Liu et al. 2014), most of which are based on smoothing algorithms such as the Whittaker smoother (Whittaker 1922; Eilers 2003). Automated baseline correction methods tend to perform well when applied to spectra of limited complexity but might be less accurate for complicated spectra. Since manual baseline correction is not straightforward either in the case of highly complex spectra, the best solution can be to combine automated baseline correction with subsequent manual inspection and fine-tuning, or to use a semi-automated approach (Emwas et al. 2018).

3.3 Metabolite identification

Metabolite identification can be a challenging task. For unambiguous compound identification, it is necessary to perform at least two orthogonal analyses (e.g. both ^{1}H and ^{13}C NMR) and compare the results with an authentic standard analysed the same way (Sumner et al. 2007). Rigorous standardized reporting criteria have been proposed (Sumner et al. 2007; Joesten & Kennedy 2019) but are not universally applied. Often, metabolites are assigned based on reference chemical shifts, coupling constants, and signal intensities found in various databases such as the Human Metabolome Database (Wishart et al. 2007), the Biological Magnetic Resonance Data Bank (Ulrich et al. 2008), or the commercial Chenomx library (Weljie et al. 2006). However, no database is complete and there are substantial discrepancies between different databases that may affect the assignment accuracy (Ross et al. 2023). For sample types that have been thoroughly characterized in previous studies, it might not be necessary to perform *de novo* identification on each new data set. For instance, the majority of signals

in ^{1}H NMR spectra of human blood samples have been assigned in several studies (Psychogios et al. 2011; Nagana Gowda et al. 2022).

3.4 Metabolite quantification

Metabolomics NMR spectra are typically complicated with numerous signals, extensive signal overlap, and large dynamic range. To accurately quantify metabolites based on overlapped signals it is necessary to perform peak fitting or spectral deconvolution, in which the spectrum is decomposed into the individual underlying metabolite signals, before quantification (Figure 6). Manual peak fitting (Weljie et al. 2006) is still the most common approach in metabolomics, although numerous automated spectral deconvolution methods exist, including recent deep neural network-based approaches (Li et al. 2023; Schmid et al. 2023). To obtain absolute concentrations, the splitting pattern of each metabolite signal must be known as well as the number of nuclei contributing to the signal. This information is usually collected from a database. Concentrations can thereafter be calculated based on the metabolite signal intensities relative to the intensity of the quantification reference signal.

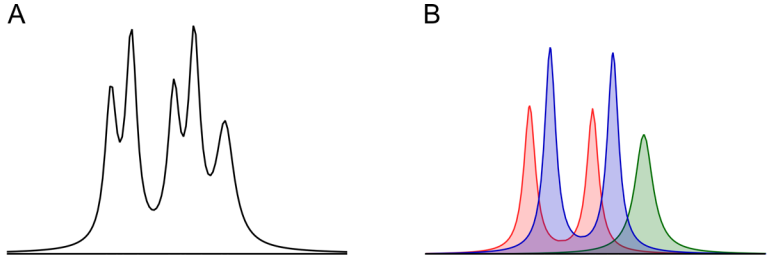


Figure 6. Illustration of spectral deconvolution. The observed spectrum (A) is the sum of three overlapping NMR signals (B) that can be resolved by spectral deconvolution.

Manual metabolite identification and quantification is time consuming and the results can be operator dependent (Tredwell et al. 2011; Canlet et al. 2023). Therefore, several automated methods have been developed that quantify metabolites based on 1D-^{1}H NMR spectra (Zheng et al. 2011; Hao et al. 2012; Ravanbakhsh et al. 2015; Tardivel et al. 2017; Röhnisch et al. 2018; Lefort et al. 2019; Häckl et al. 2021; Rout et al. 2023), typically using predefined spectral libraries. Some of these methods also perform automated

compound identification (Zheng et al. 2011; Hao et al. 2012; Ravanbakhsh et al. 2015; Tardivel et al. 2017; Lefort et al. 2019; Rout et al. 2023). Most automated methods are based on curve fitting and thus they use all experimental signals as input to estimate metabolite concentrations. As a result, the computational time increases considerably with spectral complexity as well as with the number of experimental spectra. In contrast, the automated quantification algorithm (AQuA) (Röhnisch et al. 2018, 2021) employs a unique data reduction strategy that reduces the computational time and workload substantially. AQuA calculates metabolite concentrations based on the height of one single signal apex per metabolite, while still accounting for signal interferences and variations in signal positions between different spectra. For this to be possible, AQuA collects information about signal positions, relative signal intensities, and metabolite concentrations from a spectral library and uses this as a basis to model signal interferences and subsequently compute accurate concentration estimates (Figure 7).

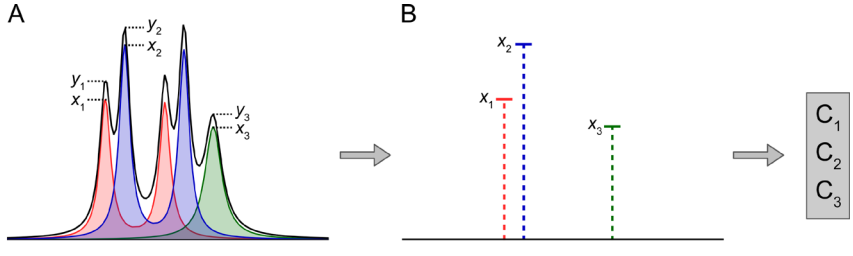


Figure 7. The principle of AQuA. A) AQuA uses a spectral library to model metabolite signals and their interferences. Thereby, the height of individual metabolite signals (x) can be calculated from the experimental signal heights (y). B) The experimental spectrum is reduced to the height x of one selected signal apex per metabolite, based on which absolute concentrations are calculated.

Most automated quantification methods, including AQuA, are intended for a specific type of sample and cannot be applied to other sample types without further method development and optimization. Furthermore, the chemical shifts and linewidths of many metabolite signals are sensitive to changes in pH, temperature, concentration, and ionic strength, which might impair automated assignments and quantification if not properly corrected for (Bhinderwala et al. 2022).

4. Reduction of signal interference in NMR spectra

As has been touched upon in the previous chapters, signal interferences, or overlap, is a common issue in NMR spectroscopy. The level of interference typically increases with the complexity of the sample, due to the large number of electrochemically different nuclei present, and becomes more problematic when the concentration differences within the sample is large, especially when low-concentration compounds are of interest. Because signal interference is most likely to occur when the degree of signal splitting is high and the possible range of chemical shifts is small, it is most frequently encountered in ^{1}H NMR spectra, particularly 1D spectra.

Spectral overlap can obstruct both identification and quantification of affected compounds. Therefore, considerable effort has been made to reduce or resolve problematic interferences. The most common strategies are summarized in this chapter.

4.1 Physical removal of interfering compounds

Perhaps the most apparent way of reducing signal interference is to remove the interfering compound or compounds from the sample altogether. Two such approaches have been mentioned already, namely ultrafiltration and protein precipitation that are both commonly employed to remove macromolecules from metabolomics samples. Other physical methodologies include liquid-liquid extraction, where compounds are separated based on their relative solubility in different liquids, and various types of chromatography where the affinity of a compound for a stationary phase compared to a mobile phase is used as the basis for separation. A more destructive approach is to chemically degrade unwanted compounds, as can

be done with carbohydrates using the oxidative agent sodium periodate (Yuan et al. 2018).

Because physical methods can be very efficient for reducing spectral overlap and improving dynamic range, they are indispensable in many studies. However, the techniques all disrupt the sample integrity, often irreversibly, and are more or less laborious to perform. Sample alteration is often undesirable, especially in quantitative studies, as it is difficult to ensure complete reproducibility and that other compounds are unaffected by the procedure. Thus, these methods should be used with caution when the aim is absolute quantification or complete sample characterization.

4.2 NMR experiments to reduce signal interferences

4.2.1 Reducing interference by enhancing resolution

A common way to avoid signal interferences in NMR spectra is to utilize certain NMR experiments that improve the spectral resolution. The most straightforward approach is to increase the spectral width so that the signals become more dispersed, e.g. by analysing ^{13}C instead of ^{1}H or by recording 2D spectra instead of 1D spectra. However, some of the information contained in a 1D-^{1}H NMR spectrum, such as signal integrals and coupling constants, is often lost with these approaches. Furthermore, the lower sensitivity of both 2D NMR spectra and heteronuclear 1D NMR spectra necessitates longer experiments than when 1D-^{1}H NMR spectra are recorded.

Removal of signal splitting also improves the spectral resolution. This is achieved using certain decoupling schemes. As already mentioned, ^{13}C NMR spectra are routinely acquired with heteronuclear proton decoupling to eliminate the effect of ^{1}H-^{13}C couplings. In ^{1}H NMR spectra, the effects of ^{1}H-^{1}H couplings can be avoided using a homonuclear broadband decoupling scheme that condenses all signals into singlets (Zangger & Sterk 1997; Foroozandeh et al. 2014). The resulting “pure shift” ^{1}H NMR spectra have very high resolution but relatively poor sensitivity.

4.2.2 Spectral editing based on relaxation and diffusion

In many cases, especially when dealing with biological samples, broad signals from macromolecules is a problematic form of interference. Certain

NMR experiments have been developed that remove these interferences based on size-dependent molecular properties.

The transverse relaxation time T_2 is usually significantly shorter in large molecules than in small molecules. If the durations of the experimental transients are sufficiently long the resulting spectrum will thus only contain signals from small molecules (Campbell et al. 1975). This type of T_2-filtering is often achieved using the CPMG sequence (Carr & Purcell 1954; Meiboom & Gill 1958), which is a 1D NMR experiment consisting of a loop of spin echoes, and can be very useful in e.g. metabolomics studies (Liu et al. 1996). However, the approach might not be suitable for absolute quantification due to frequent signal intensity modulations in CPMG spectra.

The molecular diffusion rate is also related to the molecular size, with small molecules diffusing faster than larger molecules, and this is the fundament of diffusion-ordered spectroscopy (DOSY) NMR experiments (Morris & Johnson 1992). By recording several 1D spectra with a varying gradient strength, or a varying delay to allow diffusion, a "pseudo-2D" spectrum is obtained where one axis displays the chemical shifts and the other axis separates signals based on the diffusion coefficients of their parent molecules. Diffusion can also be used as a spectral editing tool in 1D spectra, provided that there are pronounced differences in diffusion rate among the sample components. When the gradient is strong enough, or the diffusion delay long enough, the resulting spectrum will consist solely of signals from slowly diffusing species. By subtracting this diffusion-edited spectrum from a conventional 1D spectrum, a spectrum without signals from large molecules is obtained. This technique, although not innately quantitative, has been used in metabolomics to analyse protein-containing plasma samples (Liu et al. 1996; de Graaf & Behar 2003; Bliziotis et al. 2020).

4.2.3 Selective NMR experiments for improved resolution

Issues with both signal interferences and spectral dynamic range can be reduced considerably by using different kinds of selective NMR experiments, in which a spectral region containing one or several signals is selectively excited. The selective pulse can be incorporated into other pulse sequences, e.g. TOCSY, to visualize signals that are connected to the initially excited signal. Selective experiments are thus very useful for identifying and characterizing individual chemical species in highly overlapped spectra. Because the remaining spectrum is left untargeted, selective experiments can

also improve the sensitivity and visibility of minor signals substantially. The experimental time can therefore often be reduced compared to analogous non-selective experiments. It should be noted, however, that selective NMR experiments are typically not suitable for absolute quantification because the relative signal intensities are often distorted compared to non-selective spectra.

Numerous selective NMR experiments exist, including different versions of the 1D-TOCSY experiment. The classic 1D-TOCSY approach is to excite one isolated signal of interest to observe the other signals in the spin system. This strategy has for example been used for detection and relative quantification of different amino acids in honey, despite them being about 1000 times less abundant than the dominating sugars glucose and fructose (Sandusky & Raftery 2005). Even better performance of the 1D-TOCSY experiment, in terms of selectivity and signal line shapes, is obtained when a DPFGSE is used for excitation (Xu & Evans 1996).

The degree of selectivity can be adjusted depending on the research objective and the spectral properties. In cases where there are no isolated signals for a compound of interest, "ultra-selective" excitation of desired signals can be achieved using a chemical shift selective filter (Hall & Norwood 1988; Kiraly et al. 2021). The combination of chemical shift selective filters and 1D-TOCSY enabled visualization of individual low-concentration sugars in honey, despite extensive spectral overlap in the non-selective spectrum (Schievano et al. 2017). In other cases, when the overlap is less severe and several compounds are of interest, band-selective approaches can be more efficient. For instance, band-selective 2D DPFGSE-TOCSY and DPFGSE-NOESY experiments have been used to analyse minor components in mango juice, honey, and sake (Koda et al. 2011).

Band-selective DPFGSE-TOCSY for selective signal suppression

In 1999, an NMR experiment was presented that combines a DPFGSE sequence (Figure 5) with a TOCSY spin-lock to enable suppression of undesirable NMR signals while retaining other signals that reside in the same spectral region (Figure 8) (Rutherford et al. 1999). In this experiment, the DPFGSE selective 180° pulse is set to target the spectral region containing the unwanted signals, so that all signals in this region initially become defocused by the gradient pulses. Signals in remaining parts of the spectrum are refocused in the spin echo. During the subsequent TOCSY spin-lock, defocused signals that are *J*-coupled to at least one other signal not affected

by the selective pulse are reintroduced to the spectrum by magnetization transfer (Figure 8). Signals without any neighbours outside of the suppressed frequency range remain defocused and do not appear in the resulting spectrum. Obviously, all signals belonging to the same spin system as the unwanted signals need to be targeted by the selective pulse for the unwanted signals to be removed from the spectrum.

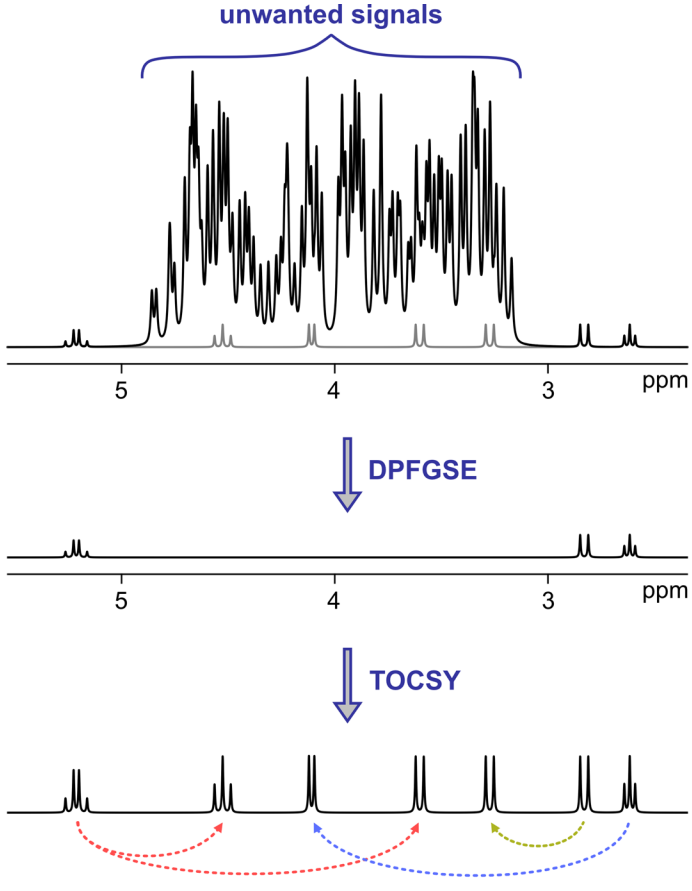


Figure 8. Illustration of the effect of the DPFGSE-TOCSY experiment on an imaginary spectrum containing signals of interest (grey) obscured by intense, unwanted signals. The entire region containing unwanted signals is initially suppressed, after which signals are reintroduced by through-bond magnetization transfer from non-suppressed spins.

This DPFGSE-TOCSY experiment was originally developed to suppress signals from a benzyl ether protecting group, which allowed observation of

previously obscured anomeric carbohydrate signals (Rutherford et al. 1999). Two-dimensional extensions of the pulse sequence were later published, including HSQC and TOCSY, for the same application (Kövér et al. 2000). Seemingly independent of these studies, analogous pulse sequences have been used to reveal signals otherwise covered by intense signal from water (Liu et al. 2001) or polyethylene glycol (Prosa et al. 2013). Apart from these examples the approach has been very sparsely used, despite its potential versatility.

4.3 Computational approaches

A third major strategy is to reduce or resolve signal interferences computationally, after the spectrum has been acquired. Compared to physical and spectroscopic methods, computational approaches are often faster and more suitable for automation, thus they enable higher sample throughput.

In targeted metabolomics, spectral deconvolution is the most common way to resolve signal interferences (introduced in section 3.4). The aim of spectral fitting, also called targeted profiling (Weljie et al. 2006), is to combine reference spectra of individual compounds to get a model spectrum highly similar to the experimental spectrum. Spectral fitting is suitable for both identification and quantification of compounds with overlapped signals and the approach is widely used in metabolomics. Unknown metabolite signals or signals from macromolecules that interfere with signals of interest can be modelled using e.g. wavelets (Hao et al. 2012) or Lorentzians (de Graaf et al. 2015) to enable absolute metabolite quantification. Alternatively, broad background signals can be regarded as a form of baseline distortion, thus they can be modelled using baseline correction functions (Zheng et al. 2011; Jacob et al. 2017).

The small molecule enhancement spectroscopy (SMolESY) method (Takis et al. 2020) uses a totally different approach to reduce signal interferences. By calculating the derivative of the imaginary part of the NMR data, broad signals are selectively suppressed whilst the line widths of other signals are significantly reduced. As a result, most interferences are eliminated in a SMolESY spectrum.

5. Thesis aims

The overarching aim of this thesis was to design, optimize, and implement methods for reduction of different types of signal interferences in NMR spectra of complex mixtures. In particular, the methods should be computationally straightforward and require minimal sample preparation and manipulation. The resulting simplification of the spectra could greatly facilitate compound identification, characterization, or quantification.

The specific aims of the thesis were as follows:

- Develop an NMR approach for suppression of unwanted signals in NMR spectra of complex mixtures (paper **I**)

- Complete ^{1}H and ^{13}C NMR spectral assignment and signal characterization of the minor furanose forms of D-glucose, with the help of NMR experiments developed in paper **I** (paper **II**)

- Develop a workflow for automated metabolite quantification in highly complex 1D-^{1}H NMR spectra from plant root exudates (paper **III**)

6. Suppression of unwanted signals

Paper **I** of this thesis introduces a band-selective DPFGSE-TOCSY NMR experiment called SUN (suppression of unwanted signals). SUN was intended as a tool for identifying and characterizing minor components of complex mixtures, specifically when signals from one or a few abundant compounds cause severe overlap in a limited spectral region. For example, biological samples often contain large amounts of sugar, whose signals can completely dominate the region 3-4 ppm in ^{1}H NMR spectra. With SUN, it is possible to selectively suppress unwanted signals and uncover other, previously obscured, signals in crowded spectral regions. The performance of SUN was demonstrated on different samples that contained large amounts of sugars such as glucose and fructose.

6.1 SUN pulse sequences

The general SUN pulse sequence, which is based on earlier DPFGSE-TOCSY sequences (Rutherford et al. 1999; Kövér et al. 2000), is shown in Figure 9. Depending on the sample type and research question at hand it might be more appropriate to employ band-selective excitation instead of suppression, i.e. to select a spectral region of interest rather than focusing on the signals to be suppressed (Figure 10). Two versions of SUN were therefore developed utilizing the two different strategies; the only difference is whether the DPFGSE contains a hard 180° pulse or not.

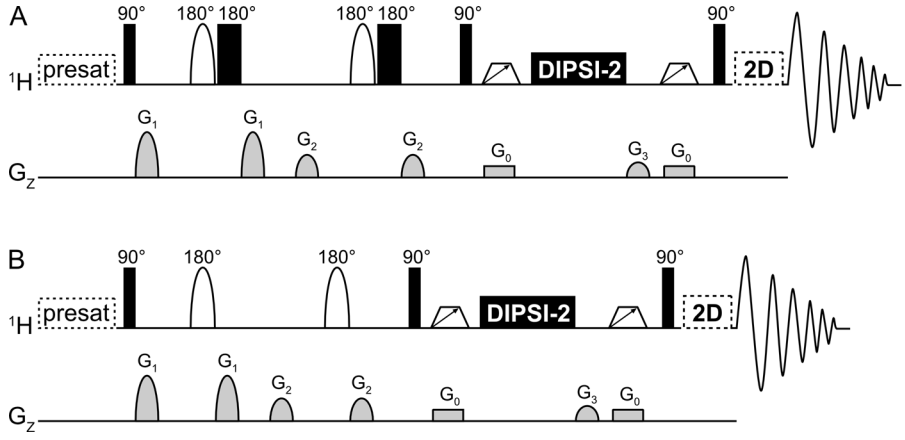


Figure 9. The SUN pulse sequence with (A) band-selective suppression and (B) band-selective excitation. Black rectangles represent hard pulses and shaped white bars represent selective pulses. By default the selective pulses are I-BURP-2 shapes (Geen & Freeman 1991), although other shapes may be used instead. PFGs are denoted G_Z. White trapezoids with arrows represent swept-frequency 180° pulses that together with the two surrounding 90° pulses and the gradient pulses G_0 and G_3 constitute the *z*-filter (Thrippleton & Keeler 2003). The inclusion of a solvent presaturation step is optional. The pulse sequences may be extended into analogous 2D experiments.

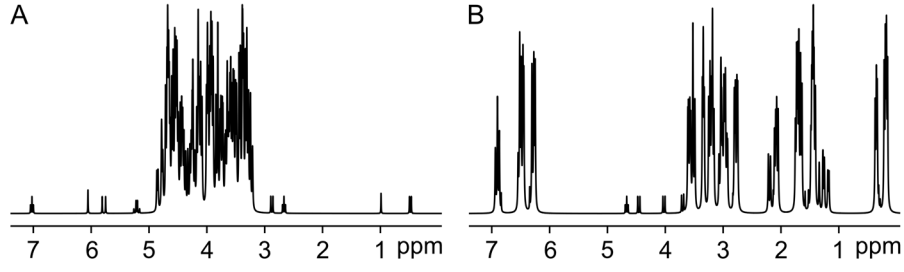


Figure 10. The distribution of unwanted signals (here, the intense signals) determines which version of the SUN experiment that is more appropriate to use in order to remove these signals. A) When the unwanted signals are located in a limited region of the spectrum (here 3-5 ppm) they can easily be removed using band-selective suppression. B) When the unwanted signals are spread out across the spectrum, it is difficult to target them all with a selective pulse. Here, it is better to use the opposite strategy and selectively excite a spectral region containing only signals of interest (e.g. 4-5 ppm).

In the SUN pulse sequences, the MLEV-17 spin-lock (Bax & Davis 1985) previously used for TOCSY transfer has been exchanged to a DIPSI-2 isotropic mixing scheme (Rucker & Shaka 1989). Before and after the TOCSY step, an efficient "*z*-filter" (Thrippleton & Keeler 2003) has been

introduced to suppress zero-quantum coherences, i.e. unwanted *z*-magnetization that would otherwise cause anti-phase distortions in the spectrum. The combination of the *z*-filter and DIPSI-2 mixing scheme led to improved line shapes in the SUN spectra.

The selective pulses are I-BURP-2 shapes (Geen & Freeman 1991) by default, for both suppression and excitation, but they can be exchanged to other shapes if desired. Furthermore, these pulses may be tuned to target two or more parts of the spectrum simultaneously if the unwanted (or wanted in the case of excitation) signals are not all localized to one spectral region. A solvent presaturation step can also be included if necessary.

In addition to the 1D experiments, several 2D and pseudo-2D extensions of SUN were developed, namely TOCSY, HSQC, HMBC, DOSY, and *J*-resolved spectroscopy. With these experiments, the interference from unwanted signals can be further reduced by dispersing the signals in two dimensions.

Finally, it should be noted that it is possible to replace the TOCSY mixing scheme with a NOESY sequence to retain signals based on their spatial proximity, rather than through-bond connectivity, to non-suppressed signals. This mechanism for magnetization transfer is less efficient than the TOCSY type, at least for small molecules, and was not explored in paper **I**. However, the approach might be useful in certain cases and it has been successfully implemented before using a different pulse sequence (Liu et al. 2001).

6.2 Application to complex mixtures

The SUN experiments were applied to a number of complex mixtures to evaluate the performance of the approach. All mixtures contained high concentrations of sugar that caused more or less severe spectral overlap and dynamic range problems in the non-selective NMR spectra. The sugar signals were reduced substantially in the SUN spectra, which increased the visibility of other signals.

6.2.1 Orange juice

Orange juice is a sample consisting mainly of water (~90 %) and the sugars sucrose, glucose, and fructose (Figure 11A). Additionally, it contains other metabolites such as amino acids and organic acids. Using either version of the 1D SUN experiment, the sugar ring proton signals could be completely

eliminated from the orange juice spectrum which revealed other, previously obscured, metabolite signals in the region 3.2-4.3 ppm (Figure 11B). No apparent difference in sugar suppression or retention of other signals was observed between the spectra obtained using band-selective suppression of the sugar region (3.0-5.5 ppm) and the spectra obtained by exciting the regions surrounding the sugar signals (here -0.1-2.7 ppm and 6.0-8.8 ppm). 2D SUN-TOCSY experiments were also conducted, with analogous sugar suppression as with the 1D versions.

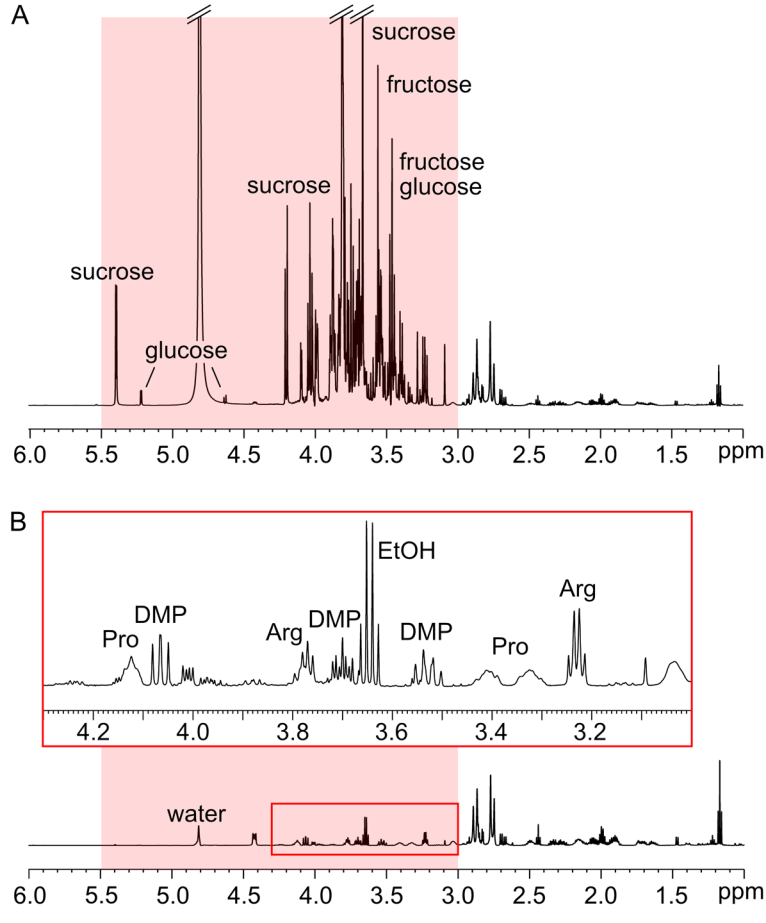


Figure 11. NMR spectra of orange juice mixed with 10 % D_2O. A) 1D-^{1}H spectrum, recorded using excitation sculpting for water suppression. B) 1D SUN spectrum with band-selective suppression of the region 3.0-5.5 ppm (highlighted in red). DMP = dimethylproline.

6.2.2 Artificial mixtures

To evaluate the performance of SUN when applied to even more challenging samples, artificial mixtures were prepared. All mixtures were dissolved in D_2O and contained 1 mM each of ascorbic acid, choline, citric acid, γ-amino-butyric acid (GABA), histidine, isoleucine, leucine, malic acid, phenyl-alanine, proline, and valine. Some of the mixtures also contained 1 mM sinigrin or 1 mM DSS-d_6. When DSS was included, its signal was used as a chemical shift reference, signal shape indicator, and to normalize spectra for intensity comparisons. Additionally, an excess of glucose was added to the mixtures, resulting in glucose concentrations of 10 mM, 100 mM, and 1000 mM, respectively. The vast span in glucose content and dynamic range among the mixtures is visualized in Figure 12A-C. For reference a mixture without glucose, but otherwise identical to the other samples, is also included (Figure 12D). In the SUN spectrum recorded with band-selective suppression of the glucose region, the glucose signals are almost entirely suppressed and the spectral dynamic range is dramatically improved compared to the non-selective spectrum (Figure 12E and F). As evidenced by the similar signal intensities in Figure 12D and F, the spectrum was largely unaffected by the SUN pulse sequence outside of the targeted region.

Three different SUN approaches (see below) were evaluated on each artificial mixture. In all experiments, the glucose signals were treated as unwanted. Different experiments were executed for each SUN approach but because the glucose suppression was highly similar in analogous 1D and 2D experiments, only 1D spectra are shown here for simplicity.

Band-selective suppression of the entire glucose region (3.1-5.5 ppm) efficiently suppressed the glucose signals but, unsurprisingly, caused the loss of certain metabolite signals that were not recovered in the TOCSY step (Figure 13). Utilizing the opposite approach, i.e. selective excitation of the regions upfield and downfield from glucose, produced very similar spectra. Interestingly, the suppression efficiency of the glucose signals was the same for all three mixtures (≥ 99 %) but because of the larger concentration ratios in the 100 and 1000 mM glucose samples, the remnants of the glucose signals are more prominent in these SUN spectra, compared to the other metabolite signals, than they are in the spectrum of the 10 mM glucose mixture (Figure 13). Accordingly, compound identification became increasingly difficult with increasing concentration differences due to more severe interference from residual glucose signals.

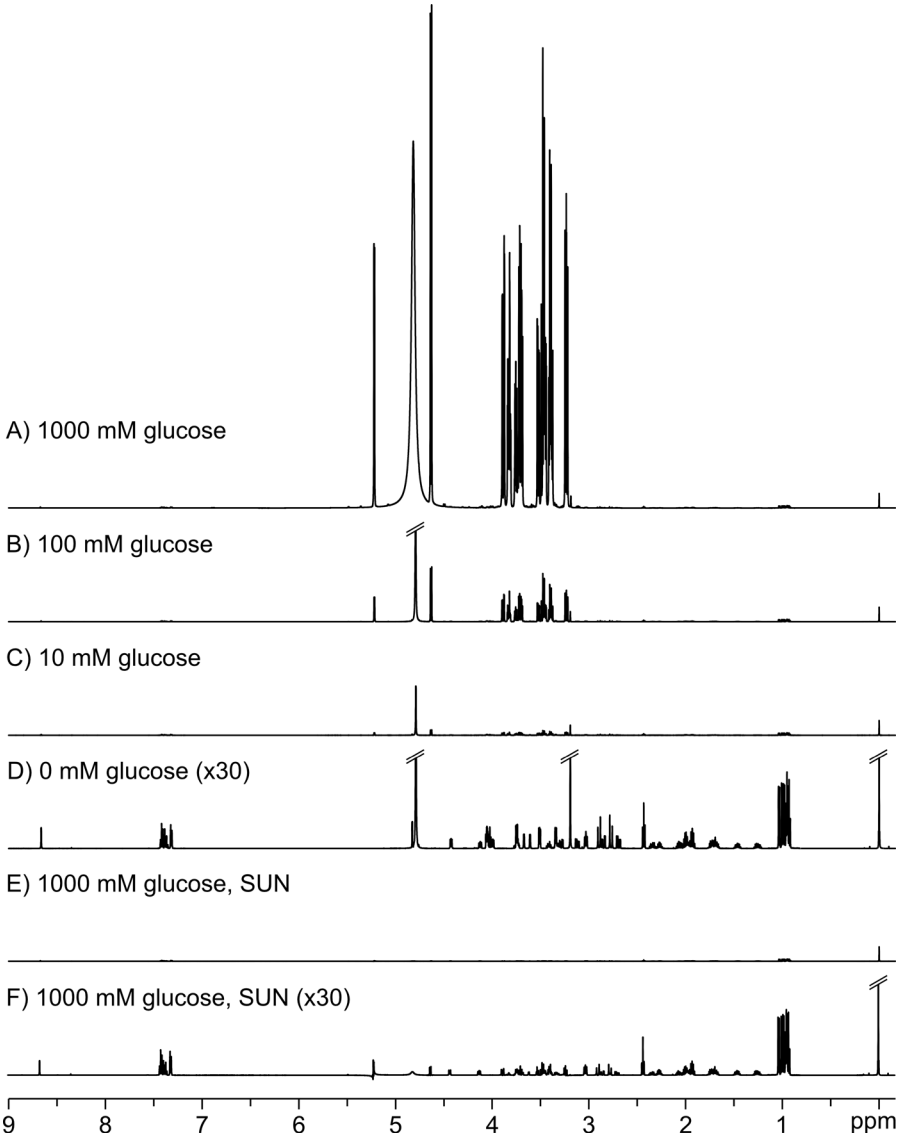


Figure 12. 1D-^{1}H NMR spectra of artificial mixtures containing glucose in different concentrations and 1 mM each of ascorbic acid, choline, citric acid, GABA, histidine, isoleucine, leucine, malic acid, phenylalanine, proline, valine, and DSS-d_6. Each sample was prepared in D_2O. Spectra A-D are non-selective whereas E and F show a SUN spectrum acquired with band-selective suppression of the spectral region 3.1-5.5 ppm. The intensities of all spectra were normalized to the DSS signal. Spectra D and F have been magnified 30 times compared to the other spectra to enhance the visibility of the low-intensity signals.

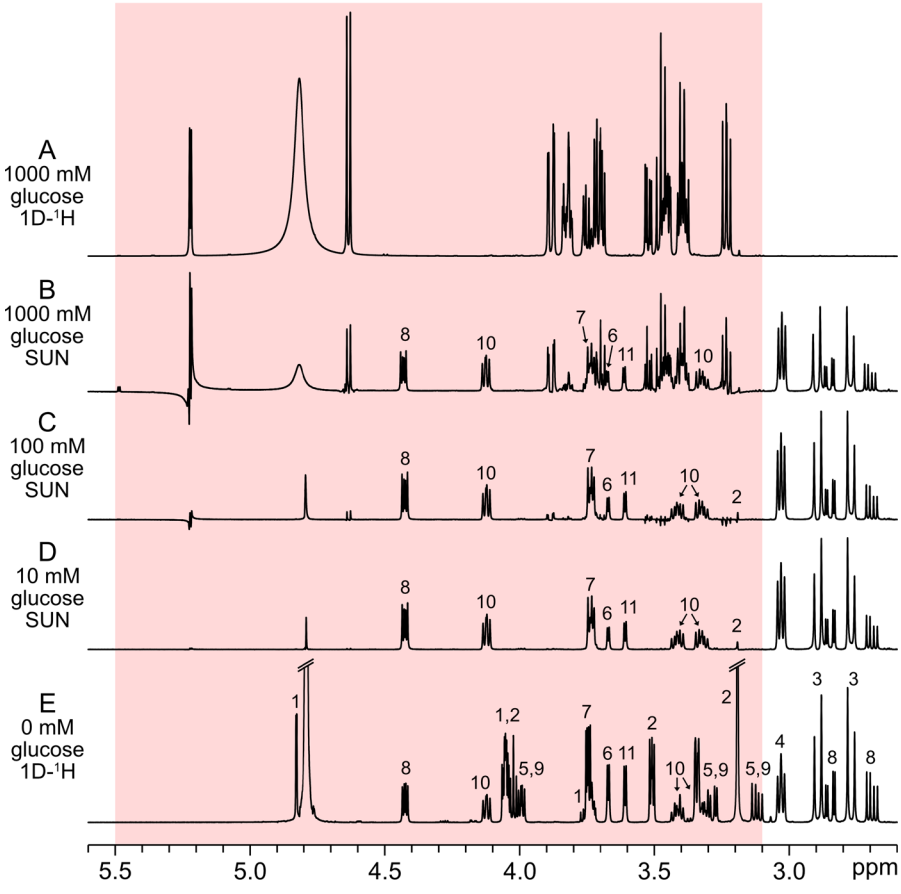


Figure 13. 1D-^{1}H NMR spectra of artificial mixtures containing glucose in different concentrations and 1 mM each of a number of other compounds (1: ascorbic acid, 2: choline, 3: citric acid, 4: GABA, 5: histidine, 6: isoleucine, 7: leucine, 8: malic acid, 9: phenylalanine, 10: proline, 11: valine, and DSS-d_6). Spectra A and E are non-selective whereas B, C, and D show SUN spectra acquired with band-selective suppression of the spectral region 3.1-5.5 ppm (highlighted in red) and a TOCSY mixing time of 50 ms. Compared to the displayed intensity of spectrum A, the vertical scale has been magnified 500 times in spectra B-D and 250 times in spectrum E for better visualization.

In another SUN experiment, the two major glucose signal regions (i.e. the regions containing the anomeric signals and the ring proton signals, respectively) were targeted separately to obtain a more specific suppression of the glucose signals (Figure 14). All non-glucose signals were then retained and could easily be identified in the spectra recorded on the 10 mM and 100 mM glucose mixtures (Figure 14C and D). Due to the very large

concentration differences in the 1000 mM glucose sample, the traces of the glucose signals in this spectrum obscured several of the other signals despite the fact that the glucose suppression was again ≥ 99 % (Figure 14B). The residual glucose is actually not the only source of interference in Figure 14B; there are also a number of other, relatively intense, signals present that neither correspond to glucopyranose (the six-membered ring forms that account for > 99 % of glucose molecules) nor to any other compound added to the mixture. These signals were later assigned to the furanose forms of glucose and were thoroughly investigated in paper **II** (discussed in Chapter 7). As can be seen in Figure 14, the glucopyranose residues in the SUN spectra displayed more pronounced phase distortions when two regions were targeted compared to when one single region was suppressed as in Figure 13.

As could be expected, the restored signals in the SUN spectra are generally less intense than the signals outside of the suppressed region. This is especially apparent when the spectra are compared with the non-selective spectrum of the mixture without glucose (Figures 12-14). Both the number of chemical bonds over which the magnetization needs to be transferred and the size of the coupling constants between a suppressed signal and its non-suppressed neighbours influence the intensity of the reintroduced signals. The TOCSY mixing time also determines what signals are reintroduced, as well as their intensity. Consequently, the SUN experiments are not directly quantitative.

6.3 Optimizing SUN

The SUN experiments were relatively straightforward to set up. Ideally, it should be sufficient to specify the spectral region or regions that contain unwanted signals (or wanted signals in the case of the band-selective excitation version) before executing an experiment. In most cases, however, the experimental parameters needed to be optimized to a greater or lesser extent before satisfactory suppression of the unwanted signals was achieved. In general, it became more crucial to optimize the settings the more intense the unwanted signals were compared to the signals of interest.

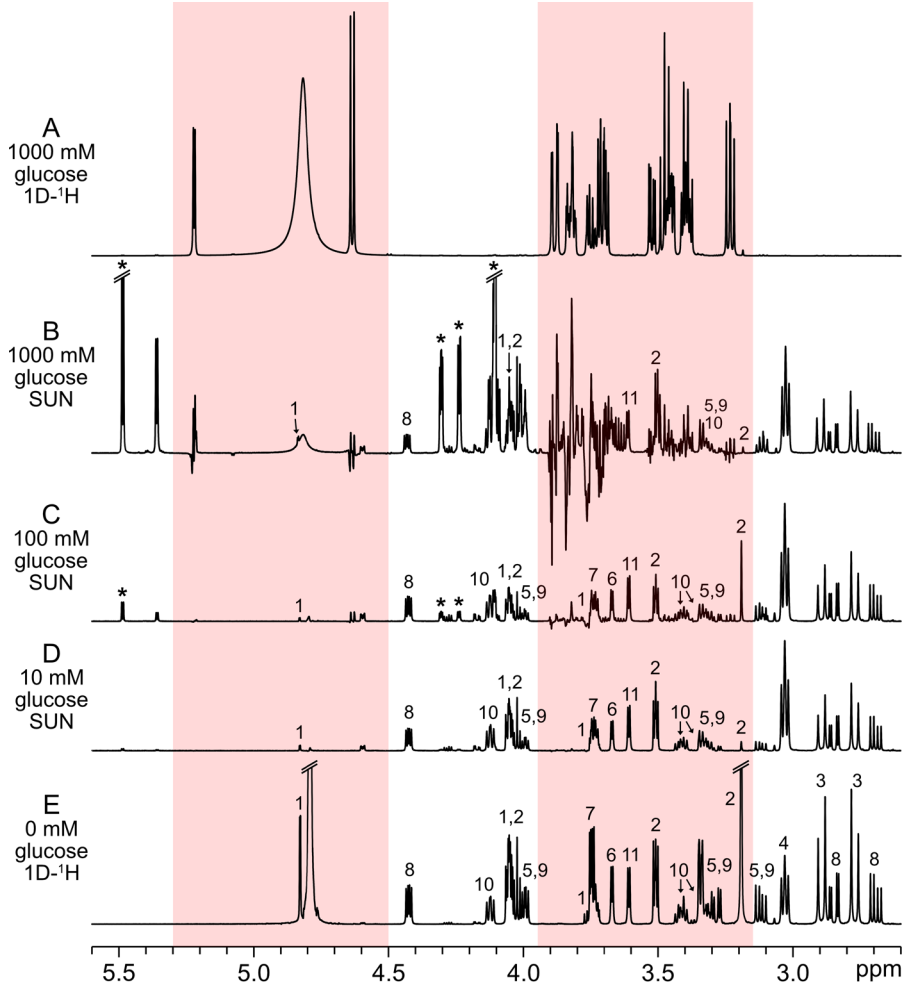


Figure 14. 1D-^{1}H NMR spectra of artificial mixtures containing glucose in different concentrations and 1 mM each of a number of other compounds (1: ascorbic acid, 2: choline, 3: citric acid, 4: GABA, 5: histidine, 6: isoleucine, 7: leucine, 8: malic acid, 9: phenylalanine, 10: proline, 11: valine, and DSS-d_6). Asterisks indicate glucofuranose signals. Spectra A and E are non-selective whereas B, C, and D show SUN spectra acquired with band-selective suppression of the spectral regions 3.15-3.95 ppm and 4.50-5.30 ppm (highlighted in red) and a TOCSY mixing time of 80 ms. Compared to the displayed intensity of spectrum A, the vertical scale has been magnified 500 times in spectra B-D and 250 times in spectrum E for better visualization.

The exact spectral region, or regions, to be targeted by the band-selective pulse is the most important factor affecting the suppression of the unwanted signals. Both the target bandwidth (i.e. the duration of the selective pulse and its RF field strength) and the target frequency need to be set appropriately. When two spectral regions were targeted, the best results were obtained when the regions were equally wide, both in the case of suppression and excitation experiments. In the SUN pulse sequences published in paper **I** there is therefore only one option to define the target bandwidth, even if several regions are selected.

In general, the biggest challenge when performing the SUN experiments was the phase of residual unwanted signals. Especially when band-selective suppression was used, the spectra often suffered from phase distortions that became more severe when two spectral regions were suppressed instead of one (see Figures 13 and 14). Distorted phase and line shapes of multiplet signals, so called *J*-modulation, is a common artefact in spin echo experiments and might be avoided by using "perfect echo" sequences in which a central 90° pulse refocuses homonuclear couplings (Takegoshi et al. 1989; Aguilar et al. 2012). Attempts were made to create a perfect echo version of SUN but without success. The phase problems could however be reduced by fine-tuning the precise phases of the band-selective pulses. Because zero-quantum coherences can also cause phase distortions, adjusting the *z*-filter gradient pulses might improve the phase properties. The SUN spectra recorded with band-selective excitation generally had less phase distortions and did not require adjustment of the pulse phases.

The length of the 90° pulse can also affect the quality of the SUN spectra; in fact, changing the pulse length with as little as 0.1 μs might influence the suppression efficiency as well as the phase of residual unwanted signals. However, the difference was only appreciable when the unwanted signals were much more intense than the wanted signals, such as in the artificial mixtures containing 100 mM and 1000 mM glucose.

Finally, the TOCSY mixing time affects what signals are reintroduced to the spectrum as well as their relative intensity. Thus, this parameter can be adjusted depending on what type of information is desired. When the mixing time is long, more signals are reintroduced although their intensity may decrease somewhat due to relaxation.

6.4 Applicability of SUN

When developing SUN, it was envisioned that the experiments would find their use primarily in metabolomics and other research fields that focus on complex biological samples. That there is a desire to suppress strong sugar signals in spectra of biological samples was confirmed in a recent publication by Singh et al. (Singh et al. 2024). However, the focus of the method presented there was not to uncover new signals in the sugar region, which is the purpose of SUN, but rather to increase the sensitivity in other spectral regions by decreasing the overall dynamic range. SUN also reduces the spectral dynamic range and improves S/N in non-suppressed spectral regions, but this is not its primary merit. Using SUN, previously obscured signals can be revealed that might be crucial for compound identification, especially when other signals are affected by severe interference. Although SUN was used to suppress signals from different sugars in paper **I** (as well as in paper **II**, which will be discussed in the next chapter), the approach is flexible and can be readily adapted to different sample types with varying types of interference issues. The major limitation of SUN is that *J*-couplings are required for signal retention, and therefore all singlets as well as certain other signals will disappear from the spectrum. This problem can be minimized by adequate selection of the spectral region to be suppressed, possibly by targeting several regions rather than one, but cannot be avoided completely unless another mechanism for magnetization transfer is used.

7. Spectral assignment of D-glucofuranose

When analysing the NMR spectra of the artificial mixtures, a few signals that could not be assigned either to glucose or to any of the other compounds added to the mixture were observed both in 1D-^{1}H spectra (Figure 14) and in ^{1}H,^{13}C-HSQC spectra. In the samples containing 10 mM and 100 mM glucose these signals were of low intensity, but in the sample with 1000 mM glucose they were comparable in size to the other metabolite signals. Upon reflection, there was two reasonable explanations for the presence of these signals: either the glucose used to prepare the artificial mixtures was not entirely pure, or the signals belonged to a minor form of glucose. A literature search revealed that some of these signals had in fact previously been assigned to the glucofuranose anomers. In paper **II**, all ^{1}H and ^{13}C NMR signals of these minor glucose forms were completely characterized, partly using SUN experiments.

7.1 The anomeric forms of D-glucose

In solution, most reducing sugars do not only have one possible structure but instead alternate between different ring and linear forms in an anomeric equilibrium. The relative amounts of the different forms vary between different sugars and can also shift depending on solvent, temperature, pH, and sample concentration. Glucose has six different anomeric forms: two six-membered rings (pyranoses), two five-membered rings (furanoses), and two linear forms (aldehyde and hydrate, respectively) (Figure 15). However, the non-pyranose forms of glucose are rarely considered due to their extremely low abundance. Over 99 % of glucose molecules assume pyranose forms in aqueous solution whereas the relative concentrations of the individual furanose and linear forms are estimated to merely be about 0.3 % and

0.005 %, respectively (Williams & Allerhand 1977; Maple & Allerhand 1987; Zhu et al. 2001). The relative concentrations of the minor forms of glucose increase with increasing temperature but none of them exceed 1 % even at 82-87 °C (Maple & Allerhand 1987; Kaufmann et al. 2018).

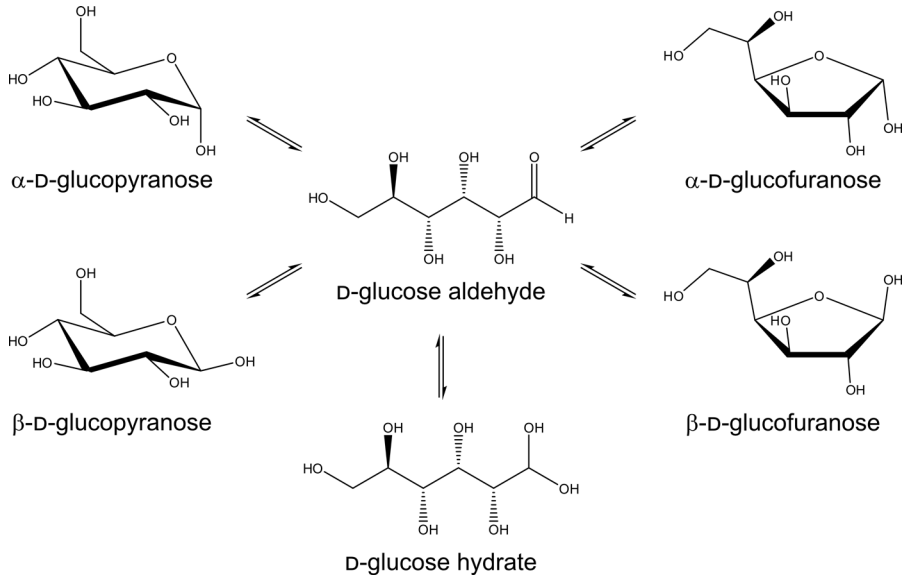


Figure 15. The anomeric equilibrium of D-glucose in solution.

7.2 Assignment of the D-glucofuranose ^{1}H and ^{13}C NMR signals

Prior to paper **II**, a few studies presenting NMR data of the furanose forms of D-glucose had been published. Interestingly, only one study contains ^{1}H NMR data (Kaufmann et al. 2018) whereas the rest were conducted using ^{13}C NMR spectroscopy (Williams & Allerhand 1977; Maple & Allerhand 1987; Zhu et al. 2001). The previous literature data is summarized in Table 1.

Table 1. Literature NMR data on the furanose forms of D-glucose.

Reference	Observed atoms	Conc. (M)	Temp. (°C)	pH/pD	^{1}H field strength (MHz)
Williams & Allerhand 1977	β-C1, β-C2, β-C4	2-4	41-43	4.3-4.7	60
Maple & Allerhand 1987[a]	α-C1, β-C1	1.4	27-82	4.8, 6.0	200
Zhu et al. 2001[a]	α-C1, β-C1	2	30	n.d.[b]	600
Kaufmann et al. 2018	α-H1, β-H1[c]	0.2	27-87	2.5	600

[a] [1-^{13}C]-glucose was used

[b] Not determined

[c] Observed when the temperature was 37 °C or more

The sparse NMR data available might seem to suggest that the glucofuranose signals are very difficult to detect in most NMR spectra. This is actually not the case; at least with a magnetic field strength of 600 MHz and a reasonably concentrated glucose sample (here ≥ 100 mM) they are clearly observable in both 1D-^{1}H and HSQC NMR spectra (Figure 16). The lower sensitivity of ^{13}C NMR spectroscopy makes the glucofuranose signals more difficult to detect in 1D-^{13}C spectra, but even then they become visible when enough transients are recorded (Figure 16). It was therefore rather surprising to learn that the majority of the signals have never been assigned. However, the furanose forms of glucose are extremely rare as building blocks in biomolecules, which together with their very low abundance might explain the limited previous interest in studying them by NMR spectroscopy. Furthermore, the much more intense pyranose signals is a severe source of interference in non-selective spectra, especially ^{1}H NMR spectra.

To study the putative glucofuranose NMR signals closer, different NMR experiments were performed. First, a 1D-^{13}C NMR spectrum was recorded on [1-^{13}C]-labelled glucose, which confirmed that the signals denoted α1 and β1 in Figure 16 indeed belong to two different glucose anomers, in agreement with earlier studies (Williams & Allerhand 1977; Maple & Allerhand 1987; Zhu et al. 2001). The intensity of these signals compared to the pyranose anomeric signals is within the previously reported glucose furanose/pyranose ratios. This, together with the experimental results described below, meant that it could be concluded beyond doubt that all the minor signals indicated in Figure 16 belong to the furanose forms of glucose.

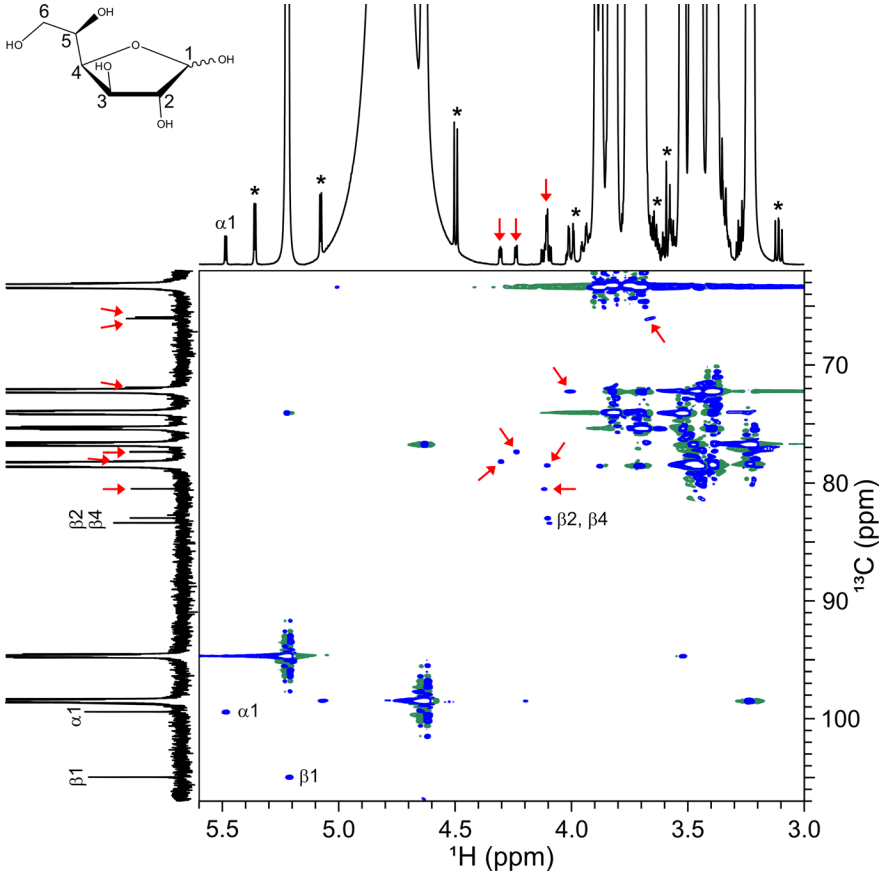


Figure 16. 1D-^{1}H, 1D-^{13}C, and ^{1}H,^{13}C-HSQC NMR spectra of D-glucose (1 M in D_2O, pD 7.0) at 25 °C. Previously identified glucofuranose signals (Williams & Allerhand 1977; Maple & Allerhand 1987; Kaufmann et al. 2018) are assigned whereas the arrows indicate other probable glucofuranose signals. ^{13}C-satellites in the 1D-^{1}H spectrum are marked with asterisks.

The three isolated glucofuranose signals in the 1D-^{1}H NMR spectrum, i.e. the ones at 5.49 ppm (previously identified as α-furanose H1 (Kaufmann et al. 2018)), 4.31 ppm, and 4.24 ppm, respectively, were excited in separate selective 1D-TOCSY experiments. The pulse sequence used was the excitation version of SUN but with a Gaussian selective pulse, instead of the I-BURP-2 normally used for band-selective excitation, to allow excitation of a single signal. The experiments revealed two different spin systems that each contained seven signals (Figure 17A and B).

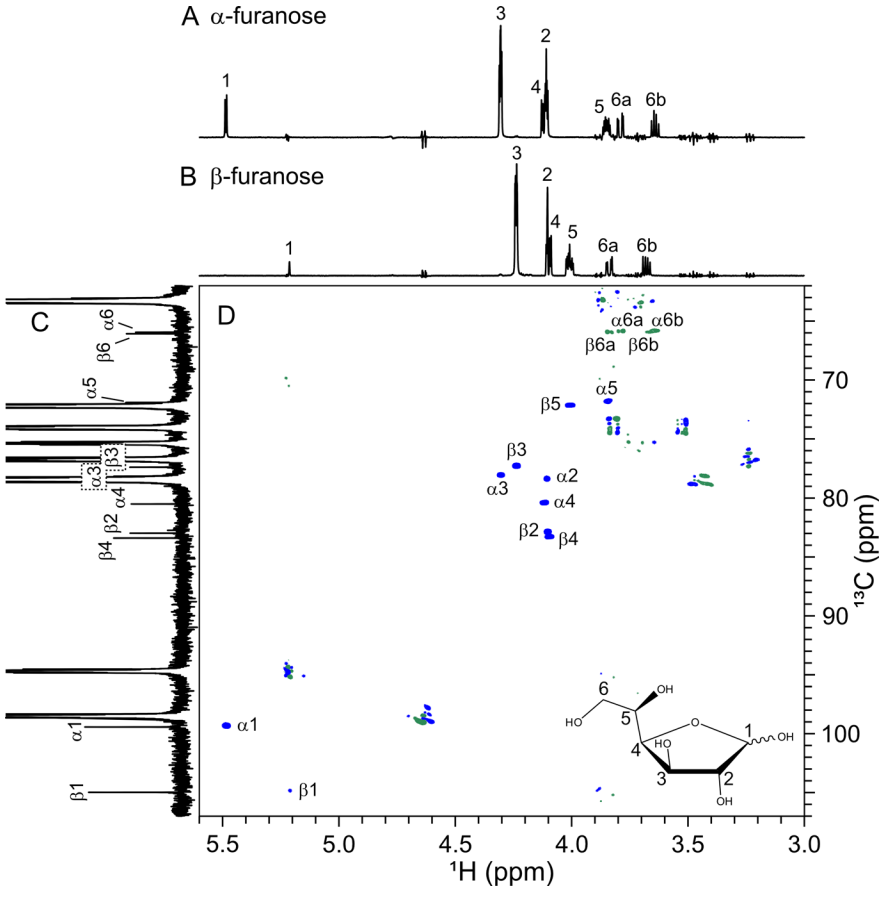


Figure 17. Assignment of all glucofuranose ^{1}H and ^{13}C NMR signals. A) 1D-^{1}H NMR spectrum of α-glucofuranose, recorded using a 1D-TOCSY experiment where the signal at 4.31 ppm (H3) was selectively excited. B) 1D-^{1}H NMR spectrum of β-glucofuranose, recorded using a 1D-TOCSY experiment where the signal at 4.24 ppm (H3) was selectively excited. C) Non-selective 1D-^{13}C NMR spectrum of glucose with the visible furanose signals assigned. The cropped unassigned signals are from the pyranose anomers. D) Multiplicity-edited SUN-HSQC spectrum of glucose obtained by band-selective excitation of the ^{1}H region 4.10-4.40 ppm. Blue signals indicate CH carbons whereas green indicate CH_2 carbons. The glucofuranose signals are assigned. Unassigned cross-peaks are mainly remnants or artefacts from the pyranose signals. The TOCSY mixing time was 100 ms in A), B), and D). The sample used was a 1 M solution of D-glucose in D_2O (pD 7.0) and the temperature was 25 °C in all experiments.

The relative order of the signals in the glucofuranose spin systems was established by performing series of selective 1D-TOCSY experiments with different mixing times (20-120 ms). The coupling constants of the anomeric

^{1}H signals were used to assign one spin system to the α-anomer and the other to the β-anomer of glucofuranose, based on previous observations on similar molecules (Angyal & Pickles 1972; Kiely & Benzing Nguyen 1975; Hayward & Angyal 1977; Lubineau & Fischer 1991; Kaufmann et al. 2018). The complete ^{1}H NMR signal assignment is shown in Figure 17A and B.

The ^{13}C NMR signals of both glucofuranose spin systems were visualized in a SUN-HSQC spectrum recorded with band-selective excitation of the ^{1}H region 4.10-4.40 ppm, i.e. signals H2, H3, and H4 from each furanose anomer. In the SUN-HSQC spectrum, all signals from the two spin systems are clearly visible without interference from the pyranose signals (Figure 17D). The ^{13}C NMR signals were then assigned based on the ^{1}H signal assignment (Figure 17C and D). It is worth pointing out that the signal previously ascribed to β-furanose C2 (Williams & Allerhand 1977) actually corresponds to β-C4, and vice versa.

7.3 Relative quantification of glucopyranose and glucofuranose

Relative quantification of the pyranose and furanose glucose anomers at 25 °C was performed using 1D-^{1}H and 1D-^{13}C NMR spectra recorded under quantitative conditions. The glucofuranose T_1 values were not measured but were assumed to be similar to those of glucopyranose, i.e. 1-2 s for both ^{1}H and ^{13}C (Williams & Allerhand 1977; Mulder et al. 2019). In both experiments, the magnetization flip angle was 30° and long relaxation delays (20 s for ^{1}H and 40 s for ^{13}C) were used to ensure complete relaxation.

In the ^{13}C NMR experiment, [1-^{13}C]-glucose was used and all glucose forms were quantified based on their respective anomeric signals. In the ^{1}H NMR spectrum, the anomeric signals of both β-pyranose and β-furanose are affected by signal interference and the quantification was therefore based on the following signals: α-pyranose H1 or H2, α-furanose H1 or H3, β-pyranose H2 or H6a, and β-furanose H3. It was found that the α-pyranose anomer accounted for 37.5 % and the β-pyranose for 62.2-62.3 % of the total glucose, which agrees well with previous findings (Williams & Allerhand 1977; Maple & Allerhand 1987; Zhu et al. 2001; Roslund et al. 2008; Maebayashi et al. 2017). The concentrations obtained for the furanose forms were around 0.12 % for both α- and β-furanose. In the ^{1}H NMR spectrum, β-furanose appeared to be slightly more abundant (0.13 % compared to

0.12 % for α-furanose) but no difference was found in the ^{13}C NMR spectrum. The proportions are comparable with results from most of the previous studies (Williams & Allerhand 1977; Maple & Allerhand 1987; Kaufmann et al. 2018). It should be noted that the anomeric equilibrium proportions of glucose are affected by e.g. temperature (Maple & Allerhand 1987; Kaufmann et al. 2018), sample concentration (Williams & Allerhand 1977; Maebayashi et al. 2017), buffer concentration (Los et al. 1956), pH, and solvent (Hyvönen et al. 1977). Therefore, the results presented here (acquired at 25 °C on 1 M D-glucose dissolved in a D_2O-based phosphate buffer with pD 7.0) might not be directly transferrable to other studies where different conditions were used.

7.4 Determination of chemical shifts and coupling constants

The ^{13}C NMR chemical shifts of the glucofuranoses are reported in Table 2. If possible, the chemical shifts were collected from a 1D-^{13}C NMR spectrum. Due to interference from the pyranose signals, the C2 signal of α-furanose and the C5 signal of β-furanose are obscured in the 1D spectrum and their chemical shifts were instead read from an HSQC spectrum. Due to the lower resolution of the HSQC spectrum compared to the 1D-^{13}C spectrum, the uncertainty of the reported α-C2 and β-C5 shifts is higher than for the other signals. Because previous studies used TMS as a chemical shift reference, the shifts were referenced both to DSS and to TMS. The obtained chemical shifts agree well with previous findings for both glucofuranose (Williams & Allerhand 1977; Zhu et al. 2001) and the corresponding methyl glucofuranosides (Ritchie et al. 1975). Table 2 also lists one-bond ^{13}C-^{1}H coupling constants for the glucofuranoses, determined by extracting 1D traces from a high-resolution SUN-HSQC spectrum recorded without carbon decoupling. The sensitivity, i.e. S/N, of this SUN-HSQC spectrum was lower than in the spectrum shown in Figure 17D and it also, naturally, decreased with increasing signal splitting. Furthermore, as previously discussed (section 6.2.2), the intensity of signals reintroduced via TOCSY transfer is lower than that of the other signals. As a result, $J_{\alpha\text{-}C5,\alpha\text{-}H5}$ could not be determined and the uncertainty of all $J_{C6,H6}$ values was relatively high due to the splitting pattern and spectral location of these signals.

Table 2. ^{13}C NMR chemical shifts (δ, ppm)[a] and $^{1}J_{C,H}$ coupling constants (Hz)[b] of α- and β-glucofuranose in D_2O (25 °C, pD 7.0).

		C1	C2	C3	C4	C5	C6
α	δ_{DSS}	99.43	78.49[c]	78.16	80.51	71.92	65.95
	δ_{TMS}	97.64	76.69[c]	76.36	78.71	70.12	64.15
	$(^{1}J_{C,H})$	(172.6)	(152.6)	(153.1)	(146.3)	(n.d.)[d]	(6a: 143.3) (6b: 142.7)
β	δ_{DSS}	104.98	82.98	77.38	83.40	72.27[c]	66.08
	δ_{TMS}	103.18	81.18	75.58	81.60	70.47[c]	64.28
	$(^{1}J_{C,H})$	(172.7)	(154.5)	(153.1)	(147.8)	(146.8)	(6a: 143.3) (6b: 142.7)

[a] Relative to DSS-d_6 or TMS at 0 ppm, as indicated in the table

[b] Standard error ≈ 0.1 Hz, except for the C6/H6 couplings where the error was ≈ 0.5 Hz

[c] Determined from an HSQC spectrum and the uncertainty is therefore around 0.01 ppm

[d] Not determined

Determination of the ^{1}H NMR chemical shifts and coupling constants was a little less straightforward. Although approximate values could be read directly from the 1D-spectra, spin simulations were needed to get more exact estimates, especially of the coupling constants. The approximate values from the experimental spectra were used as starting point for the simulations, after which the chemical shifts, *J*-values, and signal line widths were optimized in an iterative manner until the simulated spectra highly resembled the experimental spectra, as judged by visual evaluation. The determined chemical shifts and coupling constants are listed in Table 3 and the simulated spectra are shown in Figures 18 and 19.

Table 3. ^{1}H NMR chemical shifts (δ, ppm)[a] and $J_{H,H}$ coupling constants (Hz)[b] of α- and β-glucofuranose in D_2O (25 °C, pD 7.0).

	δ_{H1} $(^{3}J_{1,2})$	δ_{H2} $(^{3}J_{2,3})$	δ_{H3} $(^{3}J_{3,4})$	δ_{H4} $(^{3}J_{4,5})$	δ_{H5} $(^{3}J_{5,6a}, {}^{3}J_{5,6b})$	δ_{H6a} $(^{2}J_{6a,6b})$	δ_{H6b}
α	5.486 (3.96)	4.108 (2.42)	4.306 (3.85)	4.119 (8.50)	3.850 (2.87[c], 6.22)	3.790 ((-)12.00)	3.642
β	5.213 (< 1)[d]	4.104 (≈ 1.2)[d]	4.238 (4.19)	4.098 (9.00)	4.009 (2.75, 6.05)	3.838 ((-)11.99)	3.679

[a] Relative to DSS-d_6 at 0 ppm

[b] Standard error ≤ 0.05 Hz if not stated otherwise

[c] Standard error ≈ 0.1 Hz

[d] Standard error > 0.1 Hz

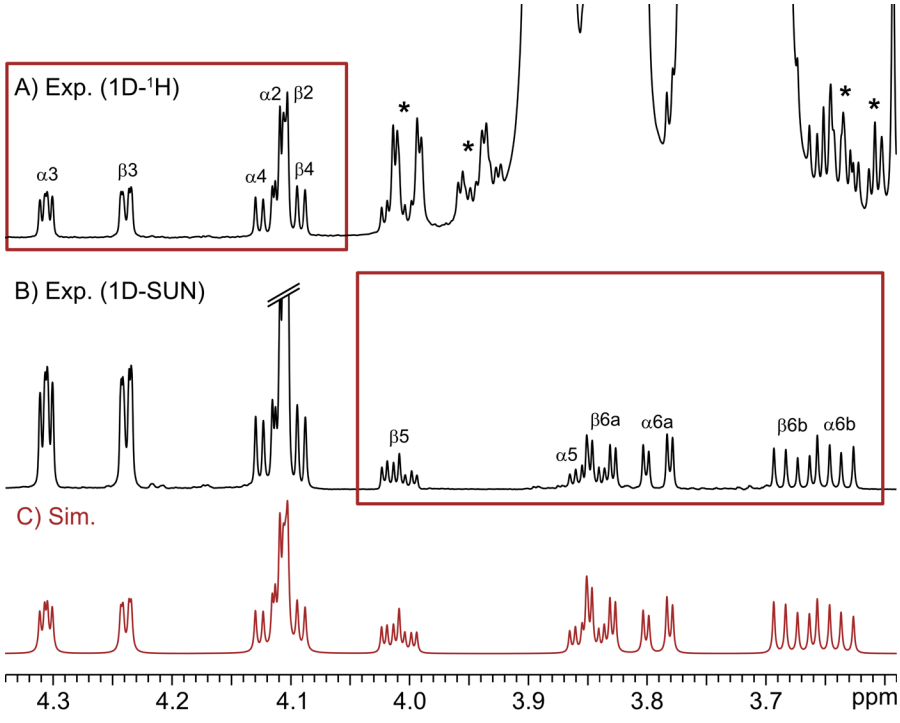


Figure 18. Experimental (A and B) and calculated (C) ^{1}H NMR spectra of the α- and β-glucofuranose ring proton signals. A non-selective experimental 1D spectrum (A) was used as input to simulate the H2, H3, and H4 signals, whereas remaining signals were simulated based on a 1D-SUN spectrum recorded with band-selective excitation of the region 4.10-4.40 ppm and a TOCSY mixing time of 100 ms (B). The regions of the experimental spectra that were used for the simulations are marked in red. Asterisks in A) denote glucopyranose ^{13}C satellites.

Because both signal line shapes and relative signal intensities may be distorted in selective spectra, the simulations were based on a non-selective spectrum whenever it was possible (i.e. for α-furanose H1-H4 and β-furanose H2-H4) (Figure 18A). For remaining signals, selective spectra had to be used as input due to overlap with the pyranose signals in the non-selective spectrum. Primarily, a band-selective SUN spectrum displaying both α- and β-furanose signals was used (Figure 18B) because the intensities and line shapes in this spectrum were generally better preserved compared to more selective spectra. However, selective spectra containing only signals from α-furanose or β-furanose were also used when determining the coupling constants (Figure 19). As can be seen in Figures 18 and 19, the simulated spectra agree well with the experimental spectra, hence the estimated

chemical shifts and coupling constants reported here are not far from their actual values. In an analogous investigation of the glucopyranose tautomers, long-range ^{1}H-^{1}H couplings were identified in addition to two- and three-bond couplings (Roslund et al. 2008). No glucofuranose long-range couplings were resolved in the spectra analysed here and so they are probably small in size if present at all. The variation of the ^{1}H NMR chemical shifts with temperature has previously been investigated for the anomeric signals of glucofuranose and was found to be around 0.01 ppm, relative to β-pyranose, in the temperature range 37-87 °C (Kaufmann et al. 2018). The chemical shifts reported here are based on spectra recorded at 25 °C; some deviation from these values is thus to be expected in spectra recorded at different temperatures.

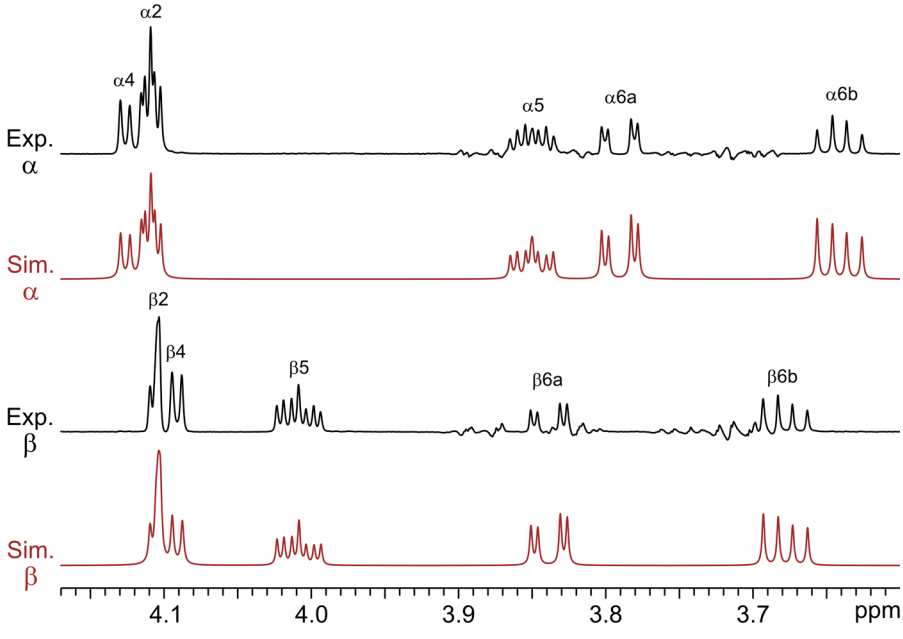


Figure 19. Experimental (exp.) and calculated (sim.) ^{1}H NMR spectra of the signals H2, H4, H5, H6a, and H6b from α-glucofuranose and β-glucofuranose, respectively. The experimental spectra are 1D-TOCSY spectra obtained by selective excitation of the respective H3 signals and a mixing time of 100 ms.

7.5 Relevance

The furanose forms of D-glucose are neglected in most studies due to their low abundance and the fact that they are practically absent in nature as biomolecular building blocks. Nevertheless, values of the glucofuranose chemical shifts and coupling constants can be useful when studying complex mixtures that contain large amounts of glucose, where the glucofuranose signals might be comparable in intensity to other minor compounds. The data can also be relevant when molecules that are similar to glucofuranose are studied. In fact, the data reported in paper **II** has already been used to characterize products from organic synthesis (Porter et al. 2023) and a fungal polysaccharide (Lamon et al. 2023). Furthermore, the furanose forms have been shown to be very important for the total reactivity of glucose (Kaufmann et al. 2018), something which could be further explored using the data presented here.

Another significant feature of paper **II** is that it again demonstrates how powerful selective and band-selective NMR experiments are for studying minor components of complex NMR spectra. The pyranose forms of glucose are 300-500 times more abundant than the furanose forms and their signals obscure the majority of the furanose signals in non-selective ^{1}H NMR spectra. In the SUN spectra, the pyranose signals were almost entirely suppressed which allowed the furanose chemical shifts and coupling constants to be determined with high precision.

8. Automated removal of broad background signals for accurate metabolite quantification

The majority of targeted NMR-based metabolomics studies are performed on the biofluids blood plasma, blood serum, or urine. Consequently, most quantification methods to date have been developed specifically for these sample types and practically all signals in their NMR spectra can be assigned based on literature data (Psychogios et al. 2011; Bouatra et al. 2013; Nagana Gowda et al. 2022). Analysis of many other types of samples remains more challenging due to their higher complexity and the limited reference data available. This was certainly the case for the system under investigation in paper **III**, namely plant root exudates. The root exudate 1D-^{1}H NMR spectra contained numerous metabolite signals and were further complicated by the presence of broad background signals. To enable metabolite quantification, a fully automated workflow was developed that first removes the interference from broad signals and then uses AQuA (Röhnisch et al. 2018) to compute accurate metabolite concentrations. The developed workflow, called extended AQuA, is a fast, accurate, and general approach that is applied to NMR spectra post-acquisition.

8.1 Plant root exudates studied by NMR spectroscopy

Plant roots continuously release a wide variety of compounds into the rhizosphere. This process is called exudation and has several functions, including mobilization of soil nutrients and mediation of plant-plant and plant-microbe interactions. The resulting root exudates are highly complex mixtures that contain primary metabolites such as sugars, amino acids, and

organic acids, as well as secondary metabolites and other compounds (Vives-Peris et al. 2020).

Although root exudates are important indicators of plant status, they are not well studied by NMR spectroscopy. The few existing studies have been performed using as diverse plant species as barley (Fan et al. 1997, 2001), wheat (Fan et al. 2001), rice (Fan et al. 2001), *Gladiolus* (Taddei et al. 2002), tomato (Escudero et al. 2014), cucumber (Zhao et al. 2016), pea (Fortier et al. 2023), and faba bean (Fortier et al. 2023). In paper **III**, root exudates from hydroponically grown oilseed rape (*Brassica napus* subsp. *napus*) seedlings were investigated. Figure 20 shows a 1D-^{1}H NMR spectrum of a rapeseed root exudate sample, freeze-dried and dissolved in D_2O (pD 7.0) without further treatment. The spectra were very crowded and contained no signal-free regions at all between ~0.5 ppm and ~8 ppm (although many signals were of very low intensity), which made it difficult to correct the spectral baseline. Furthermore, the spectra contained several broad signals, especially around 4 ppm (Figure 20B) and around 1 ppm (Figure 20C).

Metabolites were identified in the rapeseed root exudate NMR spectra based on literature data (Vives-Peris et al. 2020) and by manual spectral fitting using the Chenomx software and its inbuilt spectral library (version 8.6, Chenomx Inc., Edmonton, Canada). All assignments were verified by 2D NMR experiments and the ^{13}C NMR chemical shifts from HSQC spectra were compared with reference values from the Biological Magnetic Resonance Data Bank (Ulrich et al. 2008). Chenomx is primarily intended for biofluids and therefore it was not possible to assign all signals using this software. However, one aim of paper **III** was to investigate the accuracy of quantifying a subset of the metabolites present in a spectrum, and therefore it was decided to only target primary metabolites. All targeted metabolites are assigned in Figure 20. Figure 20A also shows the sum of the fitted reference spectra. One signal per metabolite was selected for quantification with AQuA, as indicated in Figure 20. However, the broad background signals had to be taken care of before AQuA could be executed.

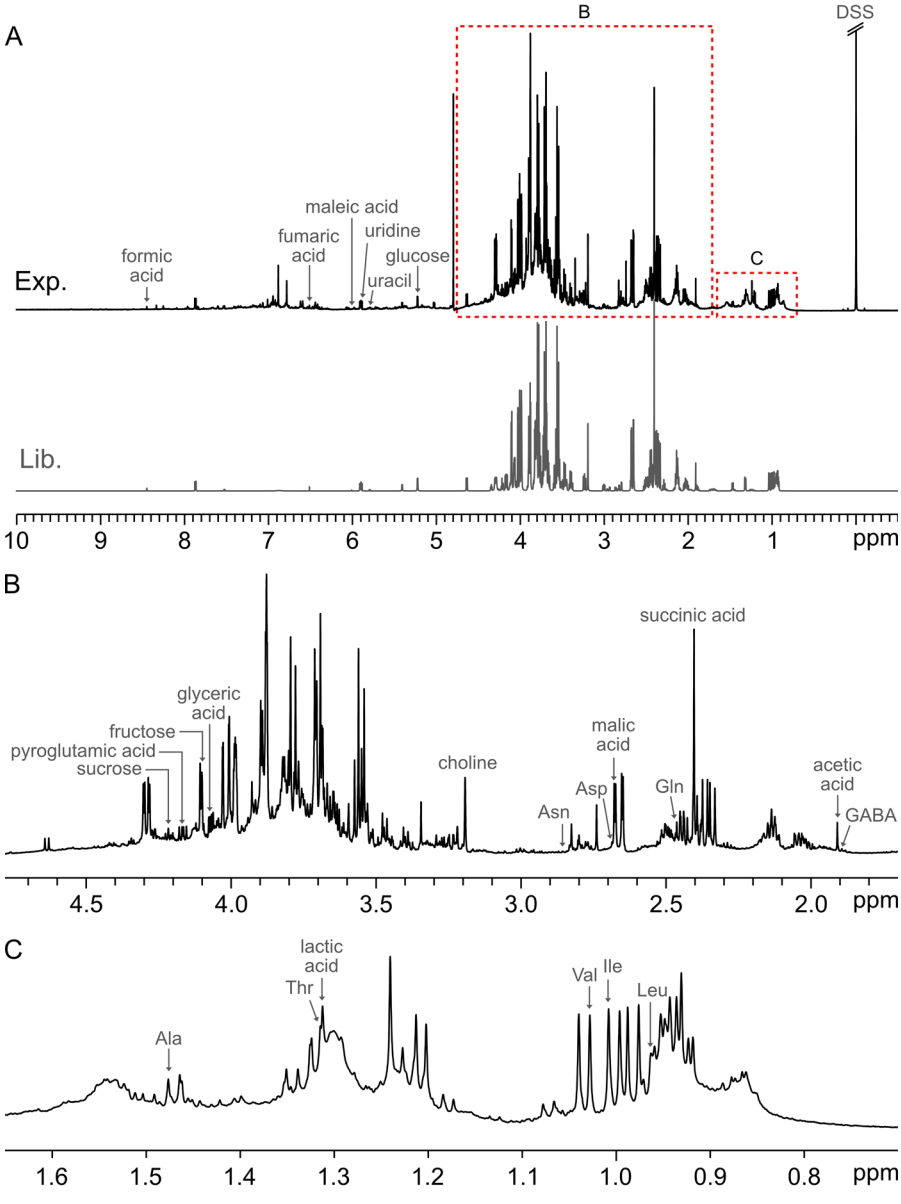


Figure 20. 1D-^{1}H NMR spectrum of a typical oilseed rape root exudate sample studied in paper **III**, with the signals apices that were used for quantification with AQuA assigned. Note that some targeted metabolites are below the limit of quantification in this particular sample. A) Experimental (Exp.) and fitted spectrum (Lib.). B) Region 1.7-4.8 ppm of the experimental spectrum. C) Region 0.70-1.65 ppm of the experimental spectrum.

8.2 Suppression of broad signals in root exudate NMR spectra

As described in chapters 3 and 4, it is common in NMR-based metabolomics to use different techniques for prevention or suppression of broad signals in the NMR spectra. However, such methods have not previously been used when preparing root exudates for NMR analysis and it appears that the broad signals have not been considered problematic in earlier studies (Fan et al. 1997, 2001; Taddei et al. 2002; Escudero et al. 2014; Zhao et al. 2016; Fortier et al. 2023). In the previous studies, the sample preparation has typically only involved freeze-drying of the collected root exudate, dissolution in water, and possibly centrifugation or filtration to remove microorganisms and root residues. This type of workflow, hereafter referred to as minimal sample preparation, was used when preparing the sample shown in Figure 20. In two publications, the samples were additionally treated with a cation-exchange resin to remove paramagnetic ions (Fan et al. 1997, 2001) but this was not considered necessary for the samples analysed in paper **III** as the signal line widths were sufficiently narrow even without this procedure (the full width at half maximum of the DSS signal was ≤ 1.20 Hz).

Different procedures for suppression or modelling of the broad signals in root exudate 1D-^{1}H NMR spectra were evaluated. The broad signals were not identified, although they are likely caused by lipids (Taddei et al. 2002). Because AQuA has previously only been applied to ultrafiltered blood plasma, the root exudate samples were first passed through a 3 kDa filter. However, the broad signals were not affected by this treatment (Figure 21). Solid phase extraction (SPE) was more efficient but not all broad signals were eliminated and the procedure affected certain metabolite signals as well (Figure 21). Since neither ultrafiltration nor SPE could remove the broad signals entirely, it was concluded that they originate from molecules that are smaller than 3 kDa and have limited hydrophobic properties.

Two NMR spectroscopic methods for suppression of broad signals – CPMG and 1D-diffusion-edited NMR – were also evaluated. Different settings for the T_2 filter and the diffusion filter, respectively, were tested but it was not possible to find parameters that selectively targeted the broad signals. Instead, both experiments caused intensity modulations amongst the metabolite signals and the removal of broad signals was incomplete (Figure 21). Thus, there was no clear difference in T_2 or diffusion coefficients between the metabolites and the compounds giving rise to the broad signals.

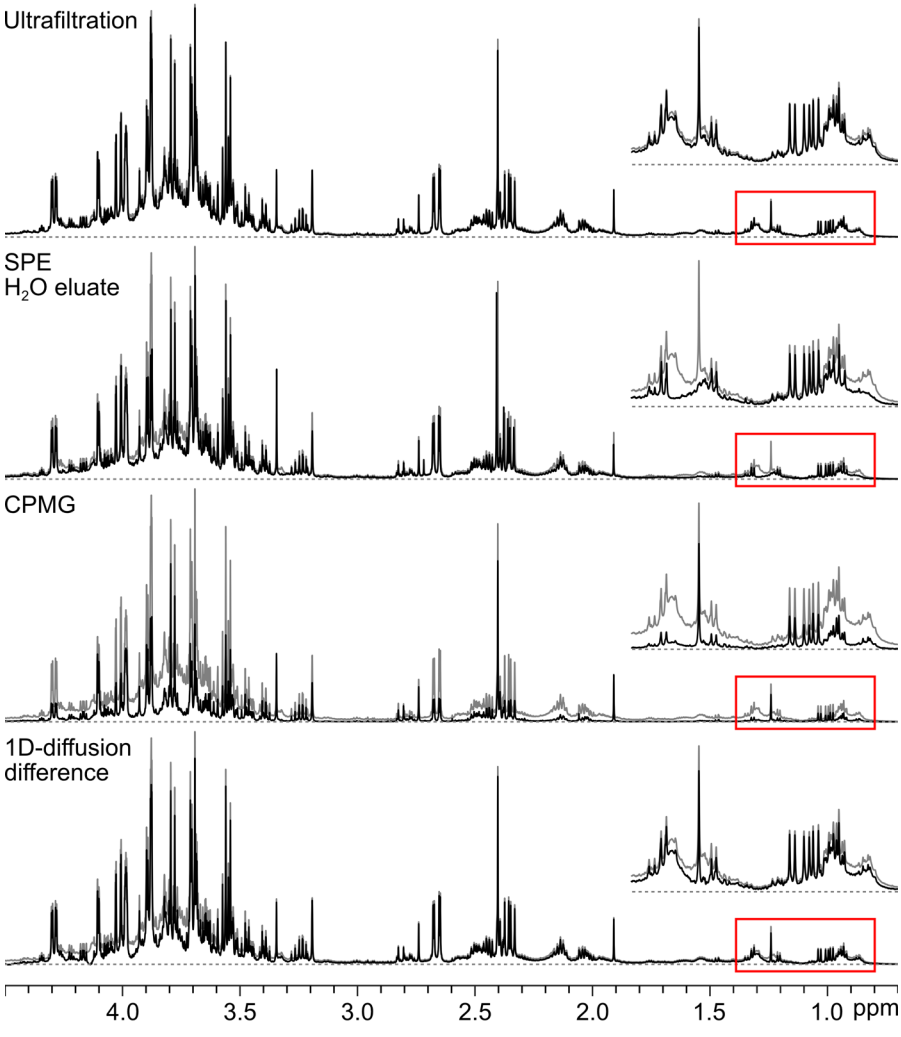


Figure 21. Evaluation of two experimental methods (ultrafiltration and SPE) and two NMR methods (CPMG and 1D-diffusion-edited NMR) for suppression of broad signals in the 1D-^{1}H NMR spectrum of an oilseed rape root exudate sample. All spectra (black) are displayed overlaid with a 1D-NOESY presaturation spectrum recorded on the intact sample (grey) and a baseline at zero intensity (dashed). The 1D-diffusion difference spectrum is the result of subtracting a 1D-diffusion-edited spectrum recorded with high gradient strength (to defocus fast-diffusing compounds) from the 1D-NOESY spectrum.

Two computational approaches were also evaluated. SMolESY (Takis et al. 2020, 2021) is an interesting strategy that mathematically suppresses broad signals whilst preserving the quantitative quality of the remaining spectrum.

After applying SMolESY to the NMR spectrum of a root exudate, the broad signals were completely suppressed and the other signals were significantly narrower than in the original spectrum (Figure 22). However, despite the improved resolution, there were still signal interferences in the SMolESY spectrum and due to the unusual shape of the SMolESY signals, there is no spectral library available that can model these interferences. Hence, it would be challenging to apply AQuA to SMolESY spectra with the purpose of obtaining absolute concentrations.

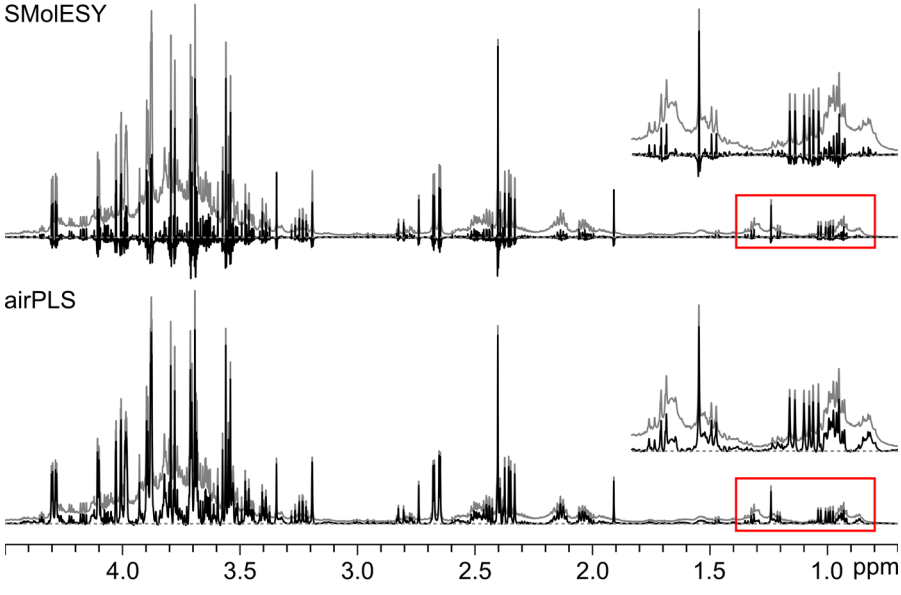


Figure 22. Evaluation of two computational methods (SMolESY and the baseline correction method airPLS, respectively) for suppression of broad signals in an oilseed rape root exudate 1D-^{1}H NMR spectrum. The computationally treated spectra (black) are displayed overlaid with the original 1D-NOESY presaturation spectrum (grey) and a baseline at zero intensity (dashed). The airPLS spectrum was obtained by subtracting the airPLS-generated baseline from the experimental spectrum.

Several studies have used a different approach, where broad signals are modelled using baseline correction functions (Provencher 1993; Zheng et al. 2011; Jacob et al. 2017). When one such method, the adaptive iteratively reweighted penalized least squares (airPLS) algorithm (Zhang et al. 2010), was applied to a root exudate spectrum, the broad signals could be efficiently suppressed without any pronounced visible impact on the metabolite signals (Figure 22). This strategy thus emerged as the most suitable option for

implementation in an AQuA workflow, especially since the method is fast, non-destructive, and automatable.

8.2.1 The airPLS algorithm

The airPLS algorithm (Zhang et al. 2010) is an automatic baseline correction function that is centred around the Whittaker smoother (Whittaker 1922; Eilers 2003) and approximates a baseline by iterated fitting. This particular algorithm was selected in paper **III** both because it is one of the most widely used methods for automatic baseline correction and because the code is available in MATLAB format, which made it straightforward to combine it with AQuA.

The smoothness of the baseline obtained with the airPLS algorithm is determined by the smoothing factor λ that can be set to any value between 1 and 1×10^9 (Figure 23). Although optimization of λ can be automated to some extent, manual supervision is usually necessary (Eilers 2003; Zhang et al. 2010). In the spectral processing tool NMRProcFlow, which uses the airPLS algorithm for local baseline corrections in regions with broad background signals, the number of λ values has been reduced to just six options where 1 results in "soft" correction and 6 leads to "high" correction (Jacob et al. 2017). In paper **III** a similar approach was adopted since it was found that it did not make much difference to the fitted baseline when λ was changed with less than a factor of 10. Optimizing the algorithm by visual inspection was thereby relatively fast and straightforward.

8.3 The extended AQuA workflow

The airPLS algorithm was combined with AQuA into an automated MATLAB-based workflow for calculation of absolute metabolite concentrations from 1D-^{1}H NMR spectra with broad background signals and baseline distortions. The workflow is illustrated in Figure 24. First, the airPLS algorithm is used to model broad signals and other baseline irregularities. The fitted baseline generated by the airPLS algorithm is then subtracted from the experimental spectrum to eliminate the broad signals. The resulting calculated spectrum is subjected to AQuA to compute metabolite concentrations. For the spectra analysed in paper **III**, the entire workflow targeting 24 metabolites could be performed in less than 1 second per spectrum.

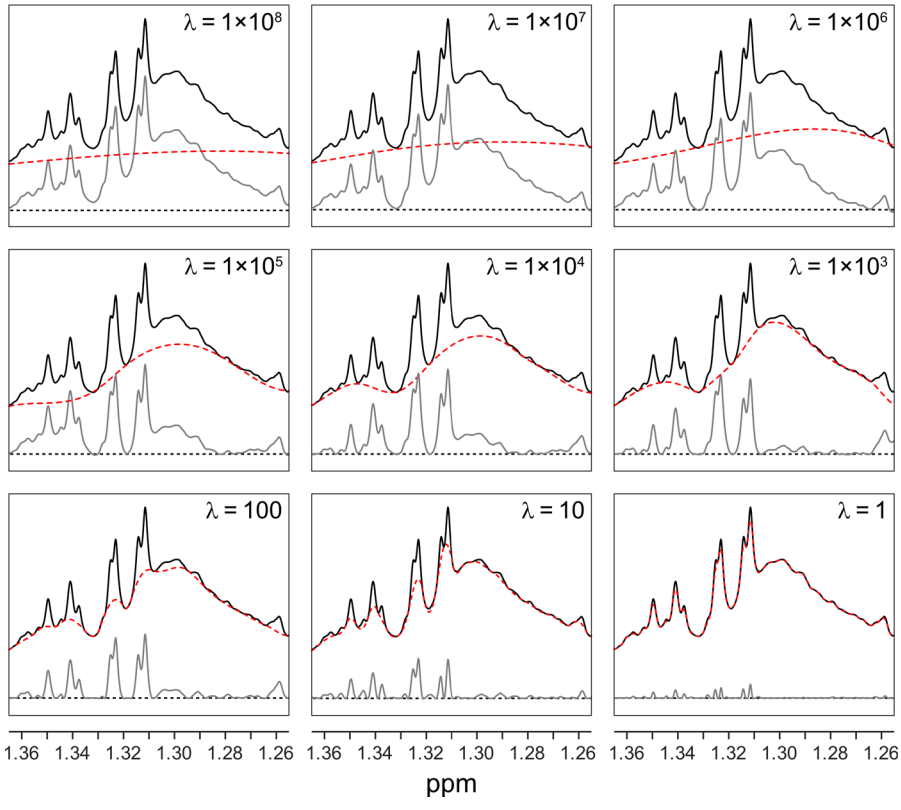


Figure 23. Illustration of the effect of the λ value on the baseline correction obtained using the airPLS algorithm. The figures include the region 1.255-1.365 ppm of an experimental root exudate spectrum (black), the fitted baselines (dashed red), the experimental spectrum after subtracting the fitted baselines (grey), and a reference baseline at zero intensity (dashed black).

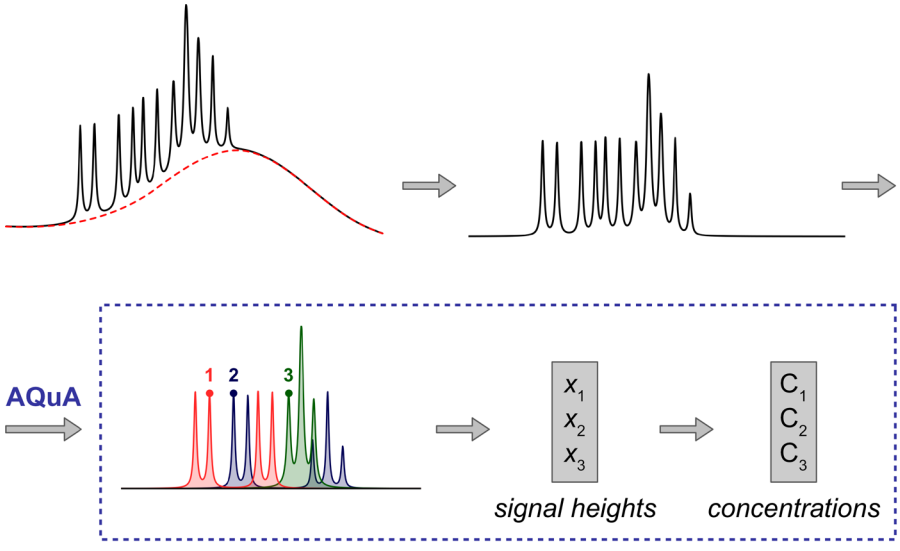


Figure 24. The extended AQuA workflow. An automatic baseline correction function (the airPLS algorithm) is used to model broad background signals. The fitted baseline is then subtracted from the experimental spectrum to generate a spectrum without broad signals. AQuA resolves remaining signal interferences and calculates absolute concentrations based on the height of one selected signal apex per metabolite.

8.4 Evaluation of the extended AQuA

Since AQuA has already been validated (Röhnisch et al. 2018, 2021), evaluation of the extended AQuA focused on the airPLS step. The idea of modelling broad signals as a baseline is not new, however the approach is often not evaluated separately other than by visual inspection. More common is that the model as a whole (i.e. the sum of all metabolite signals and broad signals) is evaluated based on its fit with the experimental spectrum (Provencher 1993; Zheng et al. 2011). In paper **III**, the absolute accuracy of the approach was investigated using simulations and spectra from a spike-in experiment. In both data sets, the rationale was to create pairs of matching spectra that have equal intensities of narrow metabolite signals, but where one of the spectra also contains broad background signals to be removed with the airPLS algorithm. The accuracy of the baseline correction could then be assessed by comparing the absolute signal intensities in the two spectra. If the airPLS algorithm suppressed the broad signals completely, without affecting the metabolite signals, the intensities in the spectra would be exactly the same. Higher signal intensities in the airPLS-corrected spectrum

than in the matching spectrum would be the result of incomplete broad signal suppression, whereas lower intensities would indicate that the algorithm suppressed both the broad signals and parts of the metabolite signals, i.e. that the fitted baseline was too flexible (see e.g. $\lambda = 10$ in Figure 23).

8.4.1 Simulations

To construct the simulated data set, reference NMR spectra of 24 different metabolites (acetic acid, alanine, asparagine, aspartic acid, choline, formic acid, fructose, fumaric acid, GABA, glucose, glyceric acid, isoleucine, lactic acid, leucine, maleic acid, malic acid, succinic acid, sucrose, tartaric acid, threonine, uracil, uridine, valine, and xylose) were summed together into one narrow signal spectrum. This spectrum was scaled into seven intensity levels, without changing the intensity ratio between the metabolites, and merged with three simulated backgrounds of varying smoothness. In that way, 21 simulated NMR spectra were generated (Figure 25).

The data set was evaluated in a high throughput manner. After analysing a few spectra to identify suitable λ values, the airPLS algorithm was applied three times to each spectrum, using three different λ values (1×10^6, 1×10^7, and 1×10^8). No further parameter optimization was performed. The airPLS-generated baselines were subtracted from the spectra, after which signal intensities in the corrected spectra were compared with the intensities of the same signals in the corresponding narrow signal spectra using linear regression and by calculating percentage differences. The results are summarized in Table 4. If the baseline correction worked perfectly, the signal intensities in the corrected spectra would be the same as in the narrow signal spectra, i.e. the slopes and R^2 coefficients would both be close to one whereas the intercepts and percentage differences would be close to zero. On average, this was indeed the case (Table 4). The roughest background (A in Figure 25) was the most difficult to model correctly, hence the relatively large intercepts and percentage differences. The choice of λ value typically became more important the more complex the spectral background was (Figure 26).

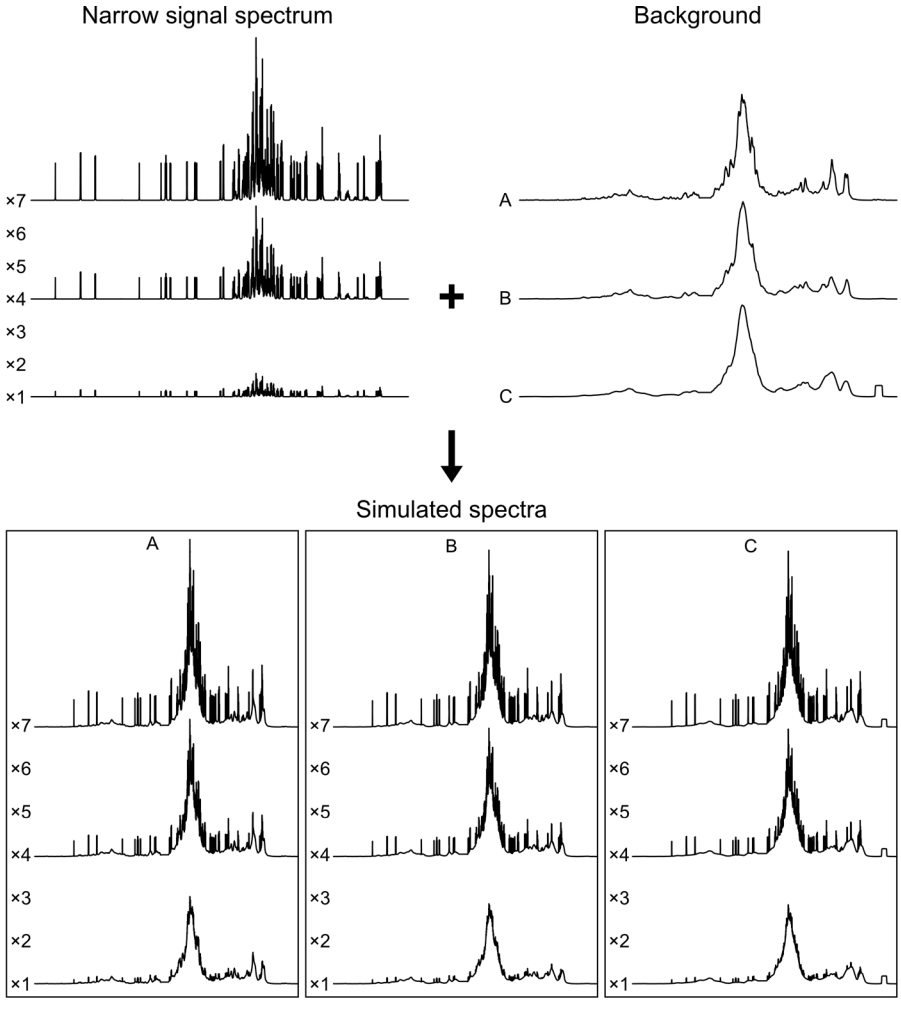


Figure 25. Illustration of the procedure for constructing the simulated NMR spectra. A simulated spectrum with narrow metabolite signals was scaled into seven different intensity levels and added to three spectral background models (A, B, and C), thus generating 21 simulated spectra. For clarity, only the spectra of intensity levels 1, 4, and 7 are shown.

Table 4. Median and mean (the latter in parentheses) percentage differences (Diff %)[a] and linear regression outcomes (slope, intercept, and R^2 value)[b] comparing signal intensities in airPLS-corrected simulated NMR spectra with intensities in the matching narrow signal spectra. The different spectral background models (A, B, and C) were evaluated separately using three different λ values.

Background	λ	Diff %	Slope	Intercept[c]	R^2
A	10^6	5.2 (7.0)	0.9541 (0.9496)	0.0287 (0.1055)	0.9999 (0.9982)
	10^7	2.0 (8.0)	0.9806 (0.9584)	0.1014 (0.2722)	0.9999 (0.9964)
	10^8	5.9 (20.1)	0.9882 (0.9723)	0.2342 (0.6513)	0.9999 (0.9966)
B	10^6	5.7 (7.2)	0.9473 (0.9366)	0.0181 (0.0316)	1.0000 (1.0000)
	10^7	3.0 (4.3)	0.9848 (0.9726)	0.0437 (0.0572)	1.0000 (0.9998)
	10^8	2.6 (6.7)	0.9918 (0.9818)	0.0746 (0.1856)	1.0000 (0.9993)
C	10^6	5.3 (6.8)	0.9475 (0.9367)	0.0133 (0.0240)	1.0000 (0.9999)
	10^7	2.1 (3.6)	0.9817 (0.9700)	0.0143 (0.0340)	0.9999 (0.9998)
	10^8	1.3 (3.2)	0.9894 (0.9818)	0.0230 (0.0597)	0.9999 (0.9994)

[a] Calculated separately for 24 metabolite signals in each pair of matching spectra using the following formula: $100 \times |\text{Intensity}_{\text{narrow spectrum}} - \text{Intensity}_{\text{simulated spectrum}}| / \text{Intensity}_{\text{narrow spectrum}}$

[b] Linear regression was performed comparing the intensities of 24 signals in the simulated narrow signal spectra (summed reference spectra) (*x*-axis) with the intensities of the same signals in the corresponding airPLS-corrected simulated spectra (*y*-axis). This was done separately for each metabolite across seven scaling levels (1, 2, 3, 4, 5, 6, and 7 arbitrary intensity units, see Figure 25).

[c] Absolute values (in arbitrary intensity units)

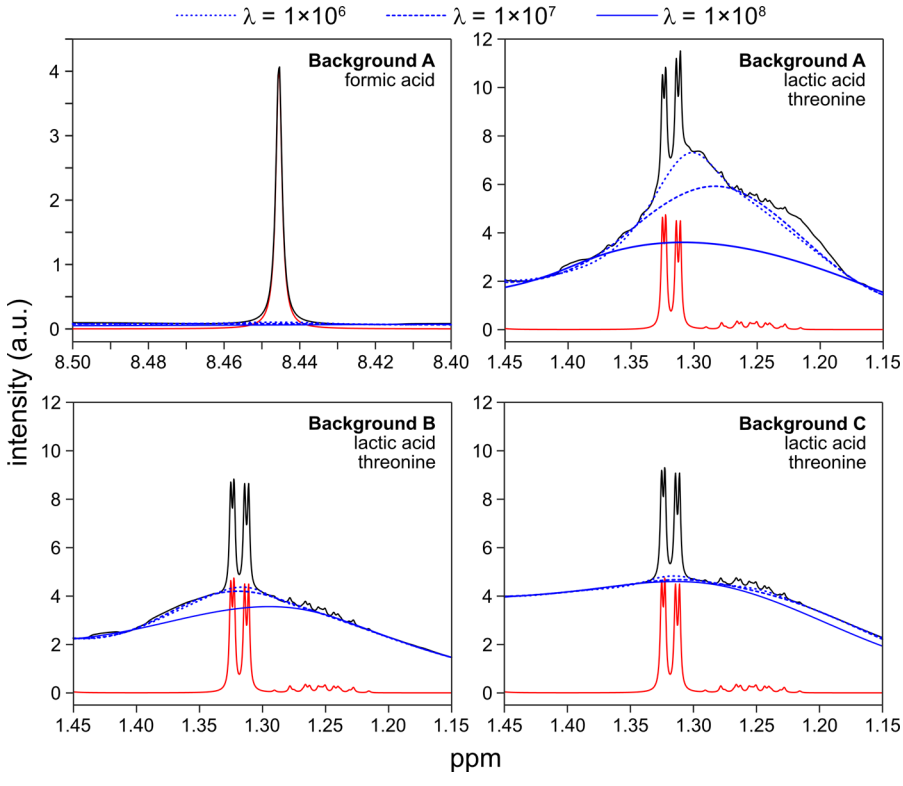


Figure 26. The effect of different λ values on the baseline correction of the formic acid region and the lactic acid/threonine region, respectively, of the simulated spectra (black). The spectra were created from background A, B, or C, as indicated in the figure, and a narrow signal spectrum of medium intensity (level 4 in Figure 25, shown in red for reference).

8.4.2 Spike-in experiment

Using the simulated NMR spectra, the performance of the airPLS algorithm could be evaluated across a wide range of metabolite signal intensities and spectral background features. However, in many of the simulated spectra the intensity of the background relative to the metabolite signals was much higher than what is commonly encountered in metabolomics samples (compare the spectra in Figure 25 with the spectrum in Figure 20). A spike-in experiment was therefore conducted as a complement to the simulations, with more representative intensity ratios between metabolite signals and the spectral background.

Five metabolites (asparagine, GABA, tartaric acid, threonine, and xylose) were added to aliquots of a pooled root exudate sample and to aliquots of a blank sample, respectively (Figure 27). These metabolites were selected because they have signals of different multiplicities that are located in different spectral regions, with varying degree of interference from other metabolite signals or the spectral background. In contrast to the simulations, the concentrations of the spiked-in metabolites were varied in a Latin square fashion so that each blank-sample pair had a unique concentration ratio between the different metabolites. Furthermore, the actual concentration ranges varied between the different metabolites to reflect biological variation. Importantly, none of the spiked-in metabolites was originally present in the root exudate sample, which again made it possible to evaluate the absolute accuracy of the broad signal suppression.

The spiked root exudate spectra contained both broad signals and signals from other metabolites and were therefore subjected to the airPLS algorithm. Here, the λ value was optimized individually for all metabolite signals (1-5 in Figure 27) in each spectrum, based on visual inspection. It was found that the same λ value (1×10^7) could be used for the regions of asparagine, tartaric acid, and xylose, regardless of concentration. Because the signal of GABA is relatively broad, it was necessary to use a higher λ value in this spectral region to avoid that the algorithm subtracted a part of the signal. Conversely, the threonine region contained a broad signal for which a lower λ value was needed. Remaining parts of the spectra were not considered in the optimization procedure but were corrected using $\lambda = 1 \times 10^7$.

AQuA was used to compute concentrations in both the blank spectra and the corrected root exudate spectra based on the signals indicated in Figure 27. The calculated concentrations are reported in Table 5. Here, the blank spectra were used as the references for assessing the accuracy of the baseline correction performed on the root exudate spectra. Although the spectra were recorded on different samples, the calculated concentrations were overall very similar (Table 5). As could be expected, the broad signal in the threonine region was the most difficult to suppress, but even then the difference between the sample and the blank was less than 10 % for four out of five samples. Tartaric acid has been excluded from the table due to an experimental error that resulted in too much compound being added to the root exudate samples.

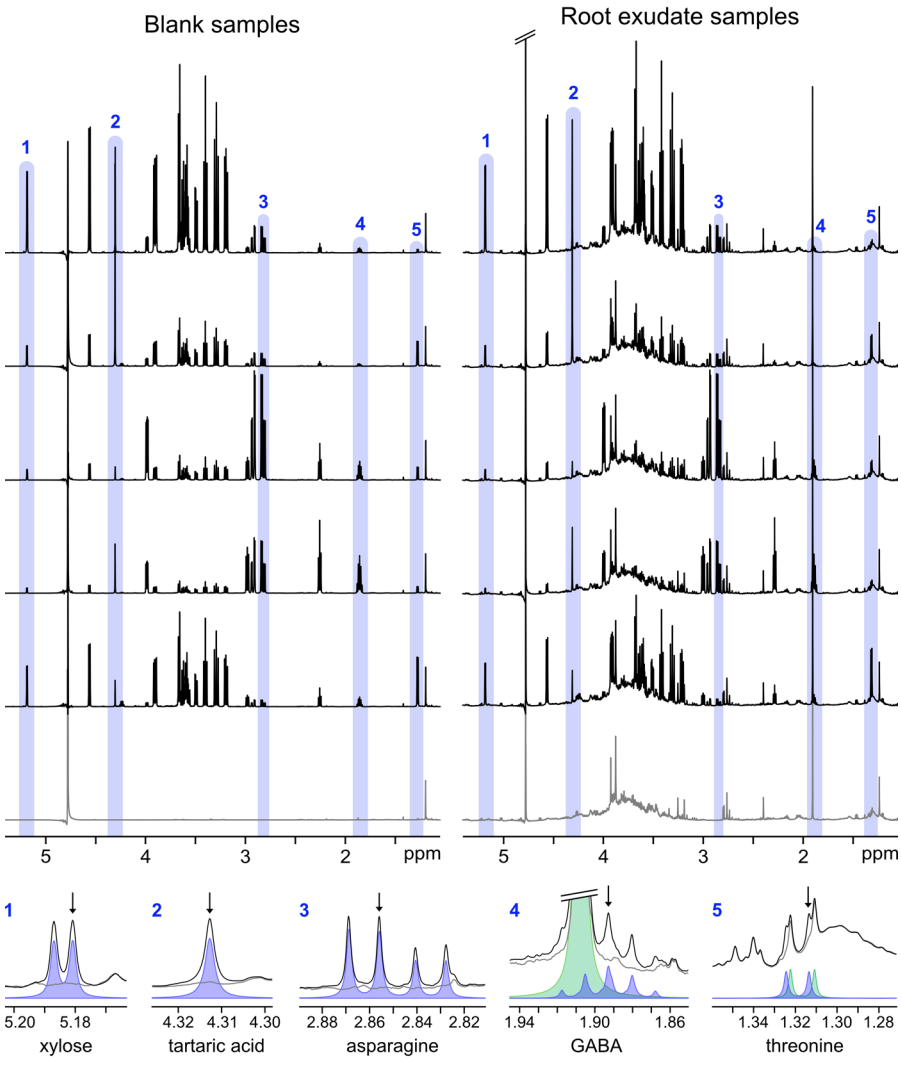


Figure 27. 1D-^{1}H NMR spectra from the spike-in experiment. Upper panel: the pooled blank sample and matched pooled root exudate sample before (grey) and after (black) addition of the spike-in metabolites. The target signals (one per metabolite) used by AQuA to compute metabolite concentrations are marked in blue. Lower panel: Magnifications of the target signal regions, with the target signals indicated (arrows). Shown are the pooled root exudate sample without added metabolites (grey), the spiked root exudate sample with the smallest amount of the respective metabolites added (black), and the respective library signals (blue). Additionally, the library signals of acetic acid and lactic acid are shown (green) because of their overlap with the signals of GABA and threonine, respectively.

Table 5. Calculated concentrations in the spiked blank samples[a] and in the spiked root exudate samples[b].

Metabolite	λ sample	Conc. sample (µM)	Conc. blank (µM)	Diff % blank-sample[c]	Mean diff %
Asparagine	1×10^7	128	133	3.8	2.7
	1×10^7	251	257	2.4	
	1×10^7	507	522	2.8	
	1×10^7	1021	1042	2.1	
	1×10^7	2057	2110	2.5	
GABA	1×10^8	25	27	6.4	4.0
	1×10^8	51	55	8.0	
	1×10^8	104	107	3.3	
	1×10^8	206	209	1.7	
	1×10^8	415	418	0.7	
Threonine	1×10^5	15	13	15.4	6.2
	1×10^5	27	26	2.2	
	5×10^5	56	51	9.0	
	1×10^6	102	99	3.2	
	1×10^6	200	198	1.4	
Xylose	1×10^7	207	197	5.2	1.2
	1×10^7	398	398	0.1	
	1×10^7	775	774	0.1	
	1×10^7	1539	1535	0.2	
	1×10^7	3048	3046	0.1	

[a] Calculated using an AQuA that included asparagine, GABA, lactic acid, tartaric acid, threonine, and xylose

[b] Calculated using an airPLS-extended AQuA that included 24 metabolites (acetic acid, alanine, asparagine, aspartic acid, choline, formic acid, fructose, fumaric acid, GABA, glucose, glyceric acid, isoleucine, lactic acid, leucine, maleic acid, malic acid, succinic acid, sucrose, tartaric acid, threonine, uracil, uridine, valine, and xylose). The λ values that were used are indicated in the table. For remaining spectral regions, $\lambda = 1 \times 10^7$ was used.

[c] Calculated using the following formula: $100 \times |C_{blank} - C_{sample}| / C_{blank}$

8.5 Concluding remarks

The extended AQuA was developed as a fast and quantitative alternative to sample preparation procedures and NMR experiments that are used to eliminate broad background signals in metabolomics 1D-^{1}H NMR spectra. The approach was found to be both linear and accurate and could be applied after minimal sample preparation. After a quick and straightforward optimization process, the extended AQuA automatically removed interferences from broad background signals and quantified absolute metabolite concentrations in less than 1 second per spectrum, regardless of the number of metabolites that were targeted.

Evaluation of the extended AQuA was performed by comparing spectra that had identical contribution from narrow metabolite signals but differed in their broad signal background. This set-up, although somewhat unconventional, allowed for a precise evaluation of the method performance. New methods are often evaluated primarily based on their performance relative to other methods, but that allows only for a relative estimation of the accuracy. The outcome of spike-in experiments is usually assessed by comparing the calculated concentrations with the actual (weighed) amounts. When the spike-in experiment of paper **III** was evaluated, the actual metabolite concentrations were deliberately not considered because that would have introduced another possible source of error, namely differences in signal shapes or relative intensities between the database and the experimental spectra that can affect the accuracy of calculated concentrations. Here, any deviations between the corrected spectra and the corresponding narrow signal spectra were due to the method performance or, in the case of the spike-in experiment, experimental errors, but not by discrepancies between the database and the analysed spectra.

Smoothing algorithms such as the airPLS algorithm are general in nature and can be executed regardless of the origin of the broad signals. Therefore, the extended AQuA should be readily applicable to NMR spectra from many different kinds of samples, both within the metabolomics field and beyond, as long as the quantified compounds are unaffected by macromolecular interactions. Furthermore, both the airPLS algorithm and AQuA can be replaced by alternative methods for baseline correction and quantification, respectively, as long as the selected methods are compatible with each other. However, this was not explored in paper **III**.

9. Conclusions and future perspectives

The work in this thesis focused on the development, evaluation, and implementation of two distinct strategies for reduction of interference in NMR spectra of complex mixtures of small molecules.

In paper **I**, a family of 1D and 2D NMR experiments called SUN was devised and optimized. SUN is an efficient and straightforward strategy for reduction of signal interference and dynamic range issues, and is especially suitable for NMR spectra with intense, unwanted signals that are localized to one or a few spectral regions. As demonstrated in both paper **I** and paper **II**, SUN and other selective NMR experiments are very useful for identifying and characterizing minor components of complex mixtures. In paper **II**, the SUN experiments enabled the complete spectral assignment of the minor furanose forms of D-glucose.

Future applications of SUN could include the study of trace compounds in biological samples that contain large amounts of sugars, such as milk, blood, and fruit extracts. However, SUN is not limited to the suppression of sugar signals but can be used to reduce interference in spectra from different sample types. The 1D SUN experiments could also be used as faster alternatives to conventional 2D-^{1}H,^{1}H NMR experiments such as TOCSY, especially when the primary aim is to increase signal resolution. Apart from finding new applications, future work should investigate whether the performance of SUN could be improved by modifying the pulse sequence, for instance by using different selective pulses or by incorporating a perfect echo element. The catalogue of SUN experiments could also be further expanded, for example by creating NOESY versions of the pulse sequences.

Paper **III** tackled another type of signal interference, namely that of broad background signals and other baseline distortions whose presence prevent accurate metabolite quantification. An automated, high-throughput

workflow – the extended AQuA – was designed that first removes these interferences computationally and then calculates absolute metabolite concentrations. The extended AQuA was successfully used to quantify metabolites in highly complex 1D-^{1}H NMR spectra from plant root exudates. Because the method is purely computational it can potentially be applied to quantitative spectra of different origins, even from other analytical techniques than solution-state NMR spectroscopy.

References

Aguilar, J.A., Nilsson, M., Bodenhausen, G. & Morris, G.A. (2012). Spin echo NMR spectra without J modulation. *Chemical Communications*, 48 (6), 811–813. https://doi.org/10.1039/C1CC16699A

Akoka, S., Barantin, L. & Trierweiler, M. (1999). Concentration measurement by proton NMR using the ERETIC method. *Analytical Chemistry*, 71 (13), 2554–2557. https://doi.org/10.1021/ac981422i

Alum, M.F., Shaw, P.A., Sweatman, B.C., Ubhi, B.K., Haselden, J.N. & Connor, S.C. (2008). 4,4-Dimethyl-4-silapentane-1-ammonium trifluoroacetate (DSA), a promising universal internal standard for NMR-based metabolic profiling studies of biofluids, including blood plasma and serum. *Metabolomics*, 4 (2), 122–127. https://doi.org/10.1007/s11306-008-0103-9

Angyal, S.J. & Pickles, V.A. (1972). Equilibria between pyranoses and furanoses. II. Aldoses. *Australian Journal of Chemistry*, 25 (8), 1695–1710. https://doi.org/10.1071/ch9721695

Bao, Q., Feng, J., Chen, F., Mao, W., Liu, Z., Liu, K. & Liu, C. (2012). A new automatic baseline correction method based on iterative method. *Journal of Magnetic Resonance*, 218, 35–43. https://doi.org/10.1016/j.jmr.2012.03.010

Barantin, L., Le Pape, A. & Akoka, S. (1997). A new method for absolute quantitation MRS metabolites. *Magnetic Resonance in Medicine*, 38 (2), 179–182. https://doi.org/10.1002/mrm.1910380203

Bauer, C., Freeman, R., Frenkiel, T., Keeler, J. & Shaka, A.J. (1984). Gaussian pulses. *Journal of Magnetic Resonance (1969)*, 58 (3), 442–457. https://doi.org/10.1016/0022-2364(84)90148-3

Bax, A. & Davis, D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance (1969)*, 65 (2), 355–360. https://doi.org/10.1016/0022-2364(85)90018-6

Bax, A. & Summers, M.F. (1986). Proton and carbon-13 assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR. *Journal of the American Chemical Society*, 108 (8), 2093–2094. https://doi.org/10.1021/ja00268a061

Beckonert, O., Keun, H.C., Ebbels, T.M.D., Bundy, J., Holmes, E., Lindon, J.C. & Nicholson, J.K. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols*, 2 (11), 2692–2703. https://doi.org/10.1038/nprot.2007.376

Bell, J.D., Brown, J.C.C. & Sadler, P.J. (1989). NMR studies of body fluids. *NMR in Biomedicine*, 2 (5–6), 246–256. https://doi.org/10.1002/nbm.1940020513

Bharti, S.K., Sinha, N., Joshi, B.S., Mandal, S.K., Roy, R. & Khetrapal, C.L. (2008). Improved quantification from 1H-NMR spectra using reduced repetition times. *Metabolomics*, 4 (4), 367–376. https://doi.org/10.1007/s11306-008-0130-6

Bhinderwala, F., Roth, H.E., Noel, H., Feng, D. & Powers, R. (2022). Chemical shift variations in common metabolites. *Journal of Magnetic Resonance*, 345, 107335. https://doi.org/10.1016/j.jmr.2022.107335

Bliziotis, N.G., Engelke, U.F.H., Aspers, R.L.E.G., Engel, J., Deinum, J., Timmers, H.J.L.M., Wevers, R.A. & Kluijtmans, L.A.J. (2020). A comparison of high-throughput plasma NMR protocols for comparative untargeted metabolomics. *Metabolomics*, 16 (5), 64. https://doi.org/10.1007/s11306-020-01686-y

Bodenhausen, G. & Ruben, D.J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69 (1), 185–189. https://doi.org/10.1016/0009-2614(80)80041-8

Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C., Bjorndahl, T.C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z.T., Poelzer, J., Huynh, J., Yallou, F.S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C. & Wishart, D.S. (2013). The human urine metabolome. *PLOS ONE*, 8 (9), e73076. https://doi.org/10.1371/journal.pone.0073076

Braunschweiler, L. & Ernst, R.R. (1983). Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *Journal of Magnetic Resonance (1969)*, 53 (3), 521–528. https://doi.org/10.1016/0022-2364(83)90226-3

Burton, I.W., Quilliam, M.A. & Walter, J.A. (2005). Quantitative 1H NMR with external standards: Use in preparation of calibration solutions for algal toxins and other natural products. *Analytical Chemistry*, 77 (10), 3123–3131. https://doi.org/10.1021/ac048385h

Campbell, I.D., Dobson, C.M., Williams, R.J.P. & Wright, P.E. (1975). Pulse methods for the simplification of protein NMR spectra. *FEBS Letters*, 57 (1), 96–99. https://doi.org/10.1016/0014-5793(75)80160-8

Canlet, C., Deborde, C., Cahoreau, E., Da Costa, G., Gautier, R., Jacob, D., Jousse, C., Lacaze, M., Le Mao, I., Martineau, E., Peyriga, L., Richard, T., Silvestre, V., Traïkia, M., Moing, A. & Giraudeau, P. (2023). NMR metabolite

quantification of a synthetic urine sample: an inter-laboratory comparison of processing workflows. *Metabolomics*, 19 (7), 65. https://doi.org/10.1007/s11306-023-02028-4

Carr, H.Y. & Purcell, E.M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94 (3), 630–638. https://doi.org/10.1103/PhysRev.94.630

Cobas, J.C., Bernstein, M.A., Martín-Pastor, M. & García Tahoces, P. (2006). A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. *Journal of Magnetic Resonance*, 183 (1), 145–151. https://doi.org/10.1016/j.jmr.2006.07.013

Davis, D.G. & Bax, A. (1985). Simplification of proton NMR spectra by selective excitation of experimental subspectra. *Journal of the American Chemical Society*, 107 (24), 7197–7198. https://doi.org/10.1021/ja00310a085

Daykin, C.A., Foxall, P.J.D., Connor, S.C., Lindon, J.C. & Nicholson, J.K. (2002). The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by 1H nuclear magnetic resonance spectroscopy. *Analytical Biochemistry*, 304 (2), 220–230. https://doi.org/10.1006/abio.2002.5637

De Marco, A. (1977). pH dependence of internal references. *Journal of Magnetic Resonance (1969)*, 26 (3), 527–528. https://doi.org/10.1016/0022-2364(77)90104-4

Deborde, C., Fontaine, J.-X., Jacob, D., Botana, A., Nicaise, V., Richard-Forget, F., Lecomte, S., Decourtil, C., Hamade, K., Mesnard, F., Moing, A. & Molinié, R. (2019). Optimizing 1D 1H-NMR profiling of plant samples for high throughput analysis: extract preparation, standardization, automation and spectra processing. *Metabolomics*, 15 (3), 28. https://doi.org/10.1007/s11306-019-1488-3

Deborde, C., Moing, A., Roch, L., Jacob, D., Rolin, D. & Giraudeau, P. (2017). Plant metabolism as studied by NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 102–103, 61–97. https://doi.org/10.1016/j.pnmrs.2017.05.001

Eilers, P.H.C. (2003). A perfect smoother. *Analytical Chemistry*, 75 (14), 3631–3636. https://doi.org/10.1021/ac034173t

Emwas, A.-H., Roy, R., McKay, R.T., Ryan, D., Brennan, L., Tenori, L., Luchinat, C., Gao, X., Zeri, A.C., Nagana Gowda, G.A., Raftery, D., Steinbeck, C., Salek, R.M. & Wishart, D.S. (2016). Recommendations and standardization of biomarker quantification using NMR-based metabolomics with particular focus on urinary analysis. *Journal of Proteome Research*, 15 (2), 360–373. https://doi.org/10.1021/acs.jproteome.5b00885

Emwas, A.-H., Saccenti, E., Gao, X., McKay, R.T., Martins dos Santos, V.A.P., Roy, R. & Wishart, D.S. (2018). Recommended strategies for spectral

processing and post-processing of 1D 1H-NMR data of biofluids with a particular focus on urine. *Metabolomics*, 14 (3), 31. https://doi.org/10.1007/s11306-018-1321-4

Escudero, N., Marhuenda-Egea, F.C., Ibanco-Cañete, R., Zavala-Gonzalez, E.A. & Lopez-Llorca, L.V. (2014). A metabolomic approach to study the rhizodeposition in the tritrophic interaction: tomato, Pochonia chlamydosporia and Meloidogyne javanica. *Metabolomics*, 10 (5), 788–804. https://doi.org/10.1007/s11306-014-0632-3

Fan, T.W.-M., Lane, A.N., Pedler, J., Crowley, D. & Higashi, R.M. (1997). Comprehensive analysis of organic ligands in whole root exudates using nuclear magnetic resonance and gas chromatography–mass spectrometry. *Analytical Biochemistry*, 251 (1), 57–68. https://doi.org/10.1006/abio.1997.2235

Fan, T.W.-M., Lane, A.N., Shenker, M., Bartley, J.P., Crowley, D. & Higashi, R.M. (2001). Comprehensive chemical profiling of gramineous plant root exudates using high-resolution NMR and MS. *Phytochemistry*, 57 (2), 209–221. https://doi.org/10.1016/S0031-9422(01)00007-3

Foroozandeh, M., Adams, R.W., Meharry, N.J., Jeannerat, D., Nilsson, M. & Morris, G.A. (2014). Ultrahigh-resolution NMR spectroscopy. *Angewandte Chemie International Edition*, 53 (27), 6990–6992. https://doi.org/10.1002/anie.201404111

Fortier, M., Lemyre, J., Ancelin, E., Oulyadi, H., Driouich, A., Vicré, M., Follet-Gueye, M.-L. & Guilhaudis, L. (2023). Development of a root exudate collection protocol for metabolomics analysis using nuclear magnetic resonance. *Plant Science*, 331, 111694. https://doi.org/10.1016/j.plantsci.2023.111694

Geen, H. & Freeman, R. (1991). Band-selective radiofrequency pulses. *Journal of Magnetic Resonance (1969)*, 93 (1), 93–141. https://doi.org/10.1016/0022-2364(91)90034-Q

Golotvin, S. & Williams, A. (2000). Improved baseline recognition and modeling of FT NMR spectra. *Journal of Magnetic Resonance*, 146 (1), 122–125. https://doi.org/10.1006/jmre.2000.2121

de Graaf, R.A. & Behar, K.L. (2003). Quantitative 1H NMR spectroscopy of blood plasma metabolites. *Analytical Chemistry*, 75 (9), 2100–2104. https://doi.org/10.1021/ac020782+

de Graaf, R.A., Prinsen, H., Giannini, C., Caprio, S. & Herzog, R.I. (2015). Quantification of 1H NMR spectra from human plasma. *Metabolomics*, 11 (6), 1702–1707. https://doi.org/10.1007/s11306-015-0828-1

Häckl, M., Tauber, P., Schweda, F., Zacharias, H.U., Altenbuchinger, M., Oefner, P.J. & Gronwald, W. (2021). An R-package for the deconvolution and integration of 1D NMR data: MetaboDecon1D. *Metabolites*, 11 (7), 452. https://doi.org/10.3390/metabo11070452

Hall, L.D. & Norwood, T.J. (1988). A chemical-shift-selective filter. *Journal of Magnetic Resonance (1969)*, 76 (3), 548–554. https://doi.org/10.1016/0022-2364(88)90358-7

Hand, E.S. & Cohen, T. (1965). Hydrophobic bonding. Its detection by nuclear magnetic resonance spectroscopy and its effect on the chemical shifts of internal standards. *Journal of the American Chemical Society*, 87 (1), 133–134. https://doi.org/10.1021/ja01079a033

Hao, J., Astle, W., De Iorio, M. & Ebbels, T.M.D. (2012). BATMAN—an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics*, 28 (15), 2088–2090. https://doi.org/10.1093/bioinformatics/bts308

Harris, R.K., Becker, E.D., Cabral De Menezes, S.M., Goodfellow, R. & Granger, P. (2001). NMR nomenclature: Nuclear spin properties and conventions for chemical shifts (IUPAC recommendations 2001). *Pure and Applied Chemistry*, 73 (11), 1795–1818. https://doi.org/10.1351/pac200173111795

Hayward, L.D. & Angyal, S.J. (1977). A symmetry rule for the circular dichroism of reducing sugars, and the proportion of carbonyl forms in aqueous solutions thereof. *Carbohydrate Research*, 53 (1), 13–20. https://doi.org/10.1016/S0008-6215(00)85450-6

Heikkinen, S., Toikka, M.M., Karhunen, P.T. & Kilpeläinen, I.A. (2003). Quantitative 2D HSQC (Q-HSQC) via suppression of J-dependence of polarization transfer in NMR spectroscopy: Application to wood lignin. *Journal of the American Chemical Society*, 125 (14), 4362–4367. https://doi.org/10.1021/ja029035k

Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K.S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A., Daviglus, M.L., Kesteloot, H., Ueshima, H., Zhao, L., Nicholson, J.K. & Elliott, P. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, 453 (7193), 396–400. https://doi.org/10.1038/nature06882

Hu, K., Westler, W.M. & Markley, J.L. (2011). Simultaneous quantification and identification of individual chemicals in metabolite mixtures by two-dimensional extrapolated time-zero 1H−13C HSQC (HSQC0). *Journal of the American Chemical Society*, 133 (6), 1662–1665. https://doi.org/10.1021/ja1095304

Hwang, T.L. & Shaka, A.J. (1995). Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A*, 112 (2), 275–279. https://doi.org/10.1006/jmra.1995.1047

Hyvönen, L., Varo, P. & Koivistoinen, P. (1977). Tautomeric equilibria of D-glucose and D-fructose: Gas-liquid chromatographic measurements. *Journal of*

Food Science, 42 (3), 654–656. https://doi.org/10.1111/j.1365-2621.1977.tb12571.x

Jacob, D., Deborde, C., Lefebvre, M., Maucourt, M. & Moing, A. (2017). NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. *Metabolomics*, 13 (4), 36. https://doi.org/10.1007/s11306-017-1178-y

Joesten, W.C. & Kennedy, M.A. (2019). RANCM: a new ranking scheme for assigning confidence levels to metabolite assignments in NMR-based metabolomics studies. *Metabolomics*, 15 (1), 5. https://doi.org/10.1007/s11306-018-1465-2

Kaufmann, M., Mügge, C. & Kroh, L.W. (2018). NMR analyses of complex D-glucose anomerization. *Food Chemistry*, 265, 222–226. https://doi.org/10.1016/j.foodchem.2018.05.100

Kiely, D.E. & Benzing Nguyen, L. (1975). Delta-dicarbonyl sugars. IV. Oxidation of carbohydrates with chromic acid. Synthesis of 6-acetamido-6-deoxy-D-xylo-hexos-5-ulose. *The Journal of Organic Chemistry*, 40 (18), 2630–2634. https://doi.org/10.1021/jo00906a012

Kim, H.K., Choi, Y.H. & Verpoorte, R. (2010). NMR-based metabolomic analysis of plants. *Nature Protocols*, 5 (3), 536–549. https://doi.org/10.1038/nprot.2009.237

Kiraly, P., Kern, N., Plesniak, M.P., Nilsson, M., Procter, D.J., Morris, G.A. & Adams, R.W. (2021). Single-scan selective excitation of individual NMR signals in overlapping multiplets. *Angewandte Chemie International Edition*, 60 (2), 666–669. https://doi.org/10.1002/anie.202011642

Koda, M., Furihata, K., Wei, F., Miyakawa, T. & Tanokura, M. (2011). F2-selective two-dimensional NMR spectroscopy for the analysis of minor components in foods. *Magnetic Resonance in Chemistry*, 49 (11), 710–716. https://doi.org/10.1002/mrc.2813

Kövér, K.E., Fehér, K., Szilágyi, L., Borbás, A., Herczegh, P. & Lipták, A. (2000). 2D NMR spectra of oligosaccharides enhanced by band-selective suppression of unwanted signals. *Tetrahedron Letters*, 41 (3), 393–396. https://doi.org/10.1016/S0040-4039(99)02068-7

Kriat, M., Confort-Gouny, S., Vion-Dury, J., Sciaky, M., Viout, P. & Cozzone, P.J. (1992). Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards. *NMR in Biomedicine*, 5 (4), 179–184. https://doi.org/10.1002/nbm.1940050404

Lam, Y.-F. & Kotowycz, G. (1977). Caution concerning the use of sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as a reference for proton NMR chemical shift studies. *FEBS Letters*, 78 (2), 181–183. https://doi.org/10.1016/0014-5793(77)80301-3

Lamon, G., Lends, A., Valsecchi, I., Wong, S.S.W., Duprès, V., Lafont, F., Tolchard, J., Schmitt, C., Mallet, A., Grélard, A., Morvan, E., Dufourc, E.J., Habenstein, B., Guijarro, J.I., Aimanianda, V. & Loquet, A. (2023). Solid-state NMR molecular snapshots of Aspergillus fumigatus cell wall architecture during a conidial morphotype transition. *Proceedings of the National Academy of Sciences*, 120 (6), e2212003120. https://doi.org/10.1073/pnas.2212003120

Lefort, G., Liaubet, L., Canlet, C., Tardivel, P., Père, M.-C., Quesnel, H., Paris, A., Iannuccelli, N., Vialaneix, N. & Servien, R. (2019). ASICS: an R package for a whole analysis workflow of 1D 1H NMR spectra. *Bioinformatics*, 35 (21), 4356–4363. https://doi.org/10.1093/bioinformatics/btz248

Li, D.-W., Bruschweiler-Li, L., Hansen, A.L. & Brüschweiler, R. (2023). DEEP Picker1D and Voigt Fitter1D: a versatile tool set for the automated quantitative spectral deconvolution of complex 1D-NMR spectra. *Magnetic Resonance*, 4 (1), 19–26. https://doi.org/10.5194/mr-4-19-2023

Liu, M., Nicholson, J.K. & Lindon, J.C. (1996). High-resolution diffusion and relaxation edited one- and two-dimensional 1H NMR spectroscopy of biological fluids. *Analytical Chemistry*, 68 (19), 3370–3376. https://doi.org/10.1021/ac960426p

Liu, M., Tang, H., Nicholson, J.K. & Lindon, J.C. (2001). Recovery of underwater resonances by magnetization transferred NMR spectroscopy (RECUR-NMR). *Journal of Magnetic Resonance*, 153 (1), 133–137. https://doi.org/10.1006/jmre.2001.2424

Liu, X., Zhang, Z., Liang, Y., Sousa, P.F.M., Yun, Y. & Yu, L. (2014). Baseline correction of high resolution spectral profile data based on exponential smoothing. *Chemometrics and Intelligent Laboratory Systems*, 139, 97–108. https://doi.org/10.1016/j.chemolab.2014.09.018

Los, J.M., Simpson, L.B. & Wiesner, K. (1956). The kinetics of mutarotation of D-glucose with consideration of an intermediate free-aldehyde form. *Journal of the American Chemical Society*, 78 (8), 1564–1568. https://doi.org/10.1021/ja01589a017

Lubineau, A. & Fischer, J.-C. (1991). High-yielding one-step conversion of D-glucose and D-galactose to the corresponding α and β methyl-D-glucofuranosides and galactofuranosides. *Synthetic Communications*, 21 (6), 815–818. https://doi.org/10.1080/00397919108019762

Madrid-Gambin, F., Oller, S., Marco, S., Pozo, Ó.J., Andres-Lacueva, C. & Llorach, R. (2023). Quantitative plasma profiling by 1H NMR-based metabolomics: impact of sample treatment. *Frontiers in Molecular Biosciences*, 10. https://doi.org/10.3389/fmolb.2023.1125582

Maebayashi, M., Ohba, M. & Takeuchi, T. (2017). Anomeric proportions of D-glucopyranose at the equilibrium determined from 1H-NMR spectra I. Investigation of experimental conditions and concentration dependence at

25.0°C. *Journal of Molecular Liquids*, 232, 408–415. https://doi.org/10.1016/j.molliq.2017.02.062

Maple, S.R. & Allerhand, A. (1987). Detailed tautomeric equilibrium of aqueous D-glucose. Observation of six tautomers by ultrahigh resolution carbon-13 NMR. *Journal of the American Chemical Society*, 109 (10), 3168–3169. https://doi.org/10.1021/ja00244a063

Meiboom, S. & Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29 (8), 688–691. https://doi.org/10.1063/1.1716296

Morris, K.F. & Johnson, C.S.Jr. (1992). Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society*, 114 (8), 3139–3141. https://doi.org/10.1021/ja00034a071

Mulder, F.A.A., Tenori, L. & Luchinat, C. (2019). Fast and quantitative NMR metabolite analysis afforded by a paramagnetic co-solute. *Angewandte Chemie International Edition*, 58 (43), 15283–15286. https://doi.org/10.1002/anie.201908006

Nagana Gowda, G.A., Hong, N.N. & Raftery, D. (2021). Evaluation of fumaric acid and maleic acid as internal standards for NMR analysis of protein precipitated plasma, serum, and whole blood. *Analytical Chemistry*, 93 (6), 3233–3240. https://doi.org/10.1021/acs.analchem.0c04766

Nagana Gowda, G.A., Pascua, V. & Raftery, D. (2022). A new limit for blood metabolite analysis using 1H NMR spectroscopy. *Journal of Magnetic Resonance Open*, 12–13, 100082. https://doi.org/10.1016/j.jmro.2022.100082

Nagana Gowda, G.A. & Raftery, D. (2014). Quantitating metabolites in protein precipitated serum using NMR spectroscopy. *Analytical Chemistry*, 86 (11), 5433–5440. https://doi.org/10.1021/ac5005103

Nicholson, J.K., Lindon, J.C. & Holmes, E. (1999). 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29 (11), 1181–1189. https://doi.org/10.1080/004982599238047

Nowick, J.S., Khakshoor, O., Hashemzadeh, M. & Brower, J.O. (2003). DSA: A new internal standard for NMR studies in aqueous solution. *Organic Letters*, 5 (19), 3511–3513. https://doi.org/10.1021/ol035347w

Oliver, S.G., Winson, M.K., Kell, D.B. & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, 16 (9), 373–378. https://doi.org/10.1016/S0167-7799(98)01214-1

Peterson, D.J. & Loening, N.M. (2007). QQ-HSQC: a quick, quantitative heteronuclear correlation experiment for NMR spectroscopy. *Magnetic Resonance in Chemistry*, 45 (11), 937–941. https://doi.org/10.1002/mrc.2073

Porter, J., Lima, M.A., Pongener, I. & Miller, G.J. (2023). Synthesis of 4-thio-d-glucopyranose and interconversion to 4-thio-d-glucofuranose. *Carbohydrate Research*, 524, 108759. https://doi.org/10.1016/j.carres.2023.108759

Prosa, N., Scherrmann, M.-C., Merlet, D. & Farjon, J. (2013). SENSASS NMR: New NMR techniques for enhancing the sensitivity and the spectral resolution of polymer supported chemicals. *Journal of Magnetic Resonance*, 237, 63–72. https://doi.org/10.1016/j.jmr.2013.09.011

Provencher, S.W. (1993). Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magnetic Resonance in Medicine*, 30 (6), 672–679. https://doi.org/10.1002/mrm.1910300604

Psychogios, N., Hau, D.D., Peng, J., Guo, A.C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T.L., Smith, S.R., Bamforth, F., Greiner, R., McManus, B., Newman, J.W., Goodfriend, T. & Wishart, D.S. (2011). The human serum metabolome. *PLOS ONE*, 6 (2), e16957. https://doi.org/10.1371/journal.pone.0016957

Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., van Dam, K. & Oliver, S.G. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*, 19 (1), 45–50. https://doi.org/10.1038/83496

Ravanbakhsh, S., Liu, P., Bjordahl, T.C., Mandal, R., Grant, J.R., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R. & Wishart, D.S. (2015). Accurate, fully-automated NMR spectral profiling for metabolomics. *PLOS ONE*, 10 (5), e0124219. https://doi.org/10.1371/journal.pone.0124219

Razzaq, A., Wishart, D.S., Wani, S.H., Hameed, M.K., Mubin, M. & Saleem, F. (2022). Advances in metabolomics-driven diagnostic breeding and crop improvement. *Metabolites*, 12 (6), 511. https://doi.org/10.3390/metabo12060511

Ritchie, R.G.S., Cyr, N., Korsch, B., Koch, H.J. & Perlin, A.S. (1975). Carbon-13 chemical shifts of furanosides and cyclopentanols. Configurational and conformational influences. *Canadian Journal of Chemistry*, 53 (10), 1424–1433. https://doi.org/10.1139/v75-197

Röhnisch, H.E., Eriksson, J., Müllner, E., Agback, P., Sandström, C. & Moazzami, A.A. (2018). AQuA: An automated quantification algorithm for high-throughput NMR-based metabolomics and its application in human plasma. *Analytical Chemistry*, 90 (3), 2095–2102. https://doi.org/10.1021/acs.analchem.7b04324

Röhnisch, H.E., Eriksson, J., Tran, L.V., Müllner, E., Sandström, C. & Moazzami, A.A. (2021). Improved automated quantification algorithm (AQuA) and its

application to NMR-based metabolomics of EDTA-containing plasma. *Analytical Chemistry*, 93 (25), 8729–8738. https://doi.org/10.1021/acs.analchem.0c04233

Roslund, M.U., Tähtinen, P., Niemitz, M. & Sjöholm, R. (2008). Complete assignments of the (1)H and (13)C chemical shifts and J(H,H) coupling constants in NMR spectra of D-glucopyranose and all D-glucopyranosyl-D-glucopyranosides. *Carbohydrate Research*, 343 (1), 101–112. https://doi.org/10.1016/j.carres.2007.10.008

Ross, I.L., Beardslee, J.A., Steil, M.M., Chihanga, T. & Kennedy, M.A. (2023). Statistical considerations and database limitations in NMR-based metabolic profiling studies. *Metabolomics*, 19 (7), 64. https://doi.org/10.1007/s11306-023-02027-5

Rout, M., Lipfert, M., Lee, B.L., Berjanskii, M., Assempour, N., Vazquez Fresno, R., Serra Cayuela, A., Dong, Y., Johnson, M., Shahin, H., Gautam, V., Sajed, T., Oler, E., Peters, H., Mandal, R. & Wishart, D.S. (2023). MagMet: A fully automated web server for targeted nuclear magnetic resonance metabolomics of plasma and serum. *Magnetic Resonance in Chemistry*, 61 (12), 681–704. https://doi.org/10.1002/mrc.5371

Rucker, S.P. & Shaka, A.J. (1989). Broadband homonuclear cross polarization in 2D N.M.R. using DIPSI-2. *Molecular Physics*, 68 (2), 509–517. https://doi.org/10.1080/00268978900102331

Rundlöf, T., Mathiasson, M., Bekiroglu, S., Hakkarainen, B., Bowden, T. & Arvidsson, T. (2010). Survey and qualification of internal standards for quantification by 1H NMR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, 52 (5), 645–651. https://doi.org/10.1016/j.jpba.2010.02.007

Rutherford, T.J., Kartha, K.P.R., Readman, S.K., Cura, P. & Field, R.A. (1999). Adaptation of an NMR signal suppression pulse sequence for the selective removal of benzylic methylene signals of benzyl ether-protected carbohydrates. *Tetrahedron Letters*, 40 (10), 2025–2028. https://doi.org/10.1016/S0040-4039(99)00108-2

Sandusky, P. & Raftery, D. (2005). Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: Application to the metabonomics of amino acids in honey. *Analytical Chemistry*, 77 (8), 2455–2463. https://doi.org/10.1021/ac0484979

Saude, E.J., Slupsky, C.M. & Sykes, B.D. (2006). Optimization of NMR analysis of biological fluids for quantitative accuracy. *Metabolomics*, 2 (3), 113–123. https://doi.org/10.1007/s11306-006-0023-5

Schievano, E., Tonoli, M. & Rastrelli, F. (2017). NMR quantification of carbohydrates in complex mixtures. A challenge on honey. *Analytical Chemistry*, 89 (24), 13405–13414. https://doi.org/10.1021/acs.analchem.7b03656

Schmid, N., Bruderer, S., Paruzzo, F., Fischetti, G., Toscano, G., Graf, D., Fey, M., Henrici, A., Ziebart, V., Heitmann, B., Grabner, H., Wegner, J.D., Sigel, R.K.O. & Wilhelm, D. (2023). Deconvolution of 1D NMR spectra: A deep learning-based approach. *Journal of Magnetic Resonance*, 347, 107357. https://doi.org/10.1016/j.jmr.2022.107357

Shimizu, A., Ikeguchi, M. & Sugai, S. (1994). Appropriateness of DSS and TSP as internal references for 1H NMR studies of molten globule proteins in aqueous media. *Journal of Biomolecular NMR*, 4 (6), 859–862. https://doi.org/10.1007/BF00398414

Singh, U., Al-Nemi, R., Alahmari, F., Emwas, A.-H. & Jaremko, M. (2024). Improving quality of analysis by suppression of unwanted signals through band-selective excitation in NMR spectroscopy for metabolomics studies. *Metabolomics*, 20 (1), 7. https://doi.org/10.1007/s11306-023-02069-9

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J. & Viant, M.R. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 3 (3), 211–221. https://doi.org/10.1007/s11306-007-0082-2

Taddei, P., Tugnoli, V., Bottura, G., Dallavalle, E. & D'Aulerio, A.Z. (2002). Vibrational, 1H-NMR spectroscopic, and thermal characterization of gladiolus root exudates in relation to Fusarium oxysporum f. sp. gladioli resistance. *Biopolymers*, 67 (6), 428–439. https://doi.org/10.1002/bip.10170

Takegoshi, K., Ogura, K. & Hikichi, K. (1989). A perfect spin echo in a weakly homonuclear J-coupled two spin-12 system. *Journal of Magnetic Resonance (1969)*, 84 (3), 611–615. https://doi.org/10.1016/0022-2364(89)90127-3

Takis, P.G., Jiménez, B., Al-Saffar, N.M.S., Harvey, N., Chekmeneva, E., Misra, S. & Lewis, M.R. (2021). A computationally lightweight algorithm for deriving reliable metabolite panel measurements from 1D 1H NMR. *Analytical Chemistry*, 93 (12), 4995–5000. https://doi.org/10.1021/acs.analchem.1c00113

Takis, P.G., Jiménez, B., Sands, C.J., Chekmeneva, E. & Lewis, M.R. (2020). SMolESY: an efficient and quantitative alternative to on-instrument macromolecular 1H-NMR signal suppression. *Chemical Science*, 11 (23), 6000–6011. https://doi.org/10.1039/D0SC01421D

Tardivel, P.J.C., Canlet, C., Lefort, G., Tremblay-Franco, M., Debrauwer, L., Concordet, D. & Servien, R. (2017). ASICS: an automatic method for identification and quantification of metabolites in complex 1D 1H NMR spectra. *Metabolomics*, 13 (10), 109. https://doi.org/10.1007/s11306-017-1244-5

Thrippleton, M.J. & Keeler, J. (2003). Elimination of zero-quantum interference in two-dimensional NMR spectra. *Angewandte Chemie International Edition*, 42 (33), 3938–3941. https://doi.org/10.1002/anie.200351947

Tiziani, S., Emwas, A.-H., Lodi, A., Ludwig, C., Bunce, C.M., Viant, M.R. & Günther, U.L. (2008). Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. *Analytical Biochemistry*, 377 (1), 16–23. https://doi.org/10.1016/j.ab.2008.01.037

Tredwell, G.D., Behrends, V., Geier, F.M., Liebeke, M. & Bundy, J.G. (2011). Between-person comparison of metabolite fitting for NMR-based quantitative metabolomics. *Analytical Chemistry*, 83 (22), 8683–8687. https://doi.org/10.1021/ac202123k

Tweeddale, H., Notley-McRobb, L. & Ferenci, T. (1998). Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis. *Journal of Bacteriology*, 180 (19), 5109–5116. https://doi.org/10.1128/jb.180.19.5109-5116.1998

Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C.F., Tolmie, D.E., Wenger, R.K., Yao, H. & Markley, J.L. (2008). BioMagResBank. *Nucleic Acids Research*, 36 (suppl_1), D402–D408. https://doi.org/10.1093/nar/gkm957

Vives-Peris, V., de Ollas, C., Gómez-Cadenas, A. & Pérez-Clemente, R.M. (2020). Root exudates: from plant to rhizosphere and beyond. *Plant Cell Reports*, 39 (1), 3–17. https://doi.org/10.1007/s00299-019-02447-5

Wang, K.-C., Wang, S.-Y., Kuo, C. & Tseng, Y.J. (2013). Distribution-based classification method for baseline correction of metabolomic 1D proton nuclear magnetic resonance spectra. *Analytical Chemistry*, 85 (2), 1231–1239. https://doi.org/10.1021/ac303233c

Weljie, A.M., Newton, J., Mercier, P., Carlson, E. & Slupsky, C.M. (2006). Targeted profiling: Quantitative analysis of 1H NMR metabolomics data. *Analytical Chemistry*, 78 (13), 4430–4442. https://doi.org/10.1021/ac060209g

Whittaker, E.T. (1922). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41, 63–75. https://doi.org/10.1017/S0013091500077853

Wider, G. & Dreier, L. (2006). Measuring protein concentrations by NMR spectroscopy. *Journal of the American Chemical Society*, 128 (8), 2571–2576. https://doi.org/10.1021/ja055336t

Williams, C. & Allerhand, A. (1977). Detection of β-D-glucofuranose in aqueous solutions of D-glucose. Application of carbon-13 fourier-transform n.m.r. spectroscopy. *Carbohydrate Research*, 56 (1), 173–179. https://doi.org/10.1016/S0008-6215(00)84250-0

Wishart, D.S., Cheng, L.L., Copié, V., Edison, A.S., Eghbalnia, H.R., Hoch, J.C., Gouveia, G.J., Pathmasiri, W., Powers, R., Schock, T.B., Sumner, L.W. &

Uchimiya, M. (2022). NMR and metabolomics—A roadmap for the future. *Metabolites*, 12 (8), 678. https://doi.org/10.3390/metabo12080678

Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, David.D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G.E., MacInnis, G.D., Weljie, A.M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B.D., Vogel, H.J. & Querengesser, L. (2007). HMDB: the Human Metabolome Database. *Nucleic Acids Research*, 35 (suppl_1), D521–D526. https://doi.org/10.1093/nar/gkl923

Xu, G. & Evans, J.S. (1996). The application of "excitation sculpting" in the construction of selective one-dimensional homonuclear coherence-transfer experiments. *Journal of Magnetic Resonance, Series B*, 111 (2), 183–185. https://doi.org/10.1006/jmrb.1996.0079

Yuan, J., Zhang, B., Wang, C. & Brüschweiler, R. (2018). Carbohydrate background removal in metabolomics samples. *Analytical Chemistry*, 90 (24), 14100–14104. https://doi.org/10.1021/acs.analchem.8b04482

Zangger, K. & Sterk, H. (1997). Homonuclear broadband-decoupled NMR spectra. *Journal of Magnetic Resonance*, 124 (2), 486–489. https://doi.org/10.1006/jmre.1996.1063

Zhang, Z.-M., Chen, S. & Liang, Y.-Z. (2010). Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135 (5), 1138–1146. https://doi.org/10.1039/B922045C

Zhao, L., Huang, Y., Hu, J., Zhou, H., Adeleye, A.S. & Keller, A.A. (2016). 1H NMR and GC-MS based metabolomics reveal defense and detoxification mechanism of cucumber plant under nano-Cu stress. *Environmental Science & Technology*, 50 (4), 2000–2010. https://doi.org/10.1021/acs.est.5b05011

Zheng, C., Zhang, S., Ragg, S., Raftery, D. & Vitek, O. (2011). Identification and quantification of metabolites in 1H NMR spectra by Bayesian model selection. *Bioinformatics*, 27 (12), 1637–1644. https://doi.org/10.1093/bioinformatics/btr118

Zhu, Y., Zajicek, J. & Serianni, A.S. (2001). Acyclic Forms of [1-13C]Aldohexoses in Aqueous Solution: Quantitation by 13C NMR and Deuterium Isotope Effects on Tautomeric Equilibria. *The Journal of Organic Chemistry*, 66 (19), 6244–6251. https://doi.org/10.1021/jo010541m

Popular science summary

Nuclear magnetic resonance (NMR) spectroscopy is an important analytical technique in several different research fields. The principle behind NMR spectroscopy is that certain types of atomic nuclei generate small magnetic fields that make them "NMR active". When these nuclei are placed in a strong, external magnetic field, they begin to rotate about their own axis at a frequency that is dependent on their local electrochemical environment. For example, a hydrogen atom attached to an oxygen atom will rotate at a slightly higher frequency than a hydrogen atom attached to a carbon atom. These two hydrogen atoms then give rise to two different signals in an NMR spectrum, while atomic nuclei located in exactly the same environment are observed as a single signal with an intensity proportional to the number of nuclei. This means that NMR spectroscopy can be used both to study the structure of molecules in detail and to quantify different components in a mixture.

A common problem in NMR spectroscopy is signal interference, also called spectral overlap. Different atomic nuclei that are situated in a similar electrochemical environment rotate at approximately the same frequency in the external magnetic field and their signals will therefore overlap with each other in the resulting NMR spectrum. Interference can make it difficult to distinguish individual signals, especially when strong signals overlap with weak signals, and it can thus become difficult or even impossible to identify and quantify certain molecules. The higher the complexity of a certain mixture, i.e. the larger the number of different molecules in the mixture and the greater the differences in concentration between the molecules, the greater the risk of signal interference. Most biological samples, e.g. plant extracts, blood and tissues, are highly complex and their NMR spectra consequently display a lot of signal interference.

Problematic signal interference can be prevented by physically separating the different components of a sample before analysis, for example by filtering the sample or by using some type of chromatography, but this is often both expensive and time consuming and might cause misleading results as the sample changes in the process. Another way to reduce signal interference is to use certain NMR experiments or mathematical methods. These options are considerably faster than a physical separation, especially as they can be applied to intact samples. This thesis describes two such methods, an NMR experiment and a mathematical workflow, both of which are designed to reduce signal interference by removing unwanted signals from NMR spectra.

The NMR experiment described in this thesis is called SUN (suppression of unwanted signals) and is used to reduce specific unwanted signals so that other, low-intensity signals can be studied. To evaluate and optimize the SUN experiment, it was applied to various samples, including orange juice and three artificial mixtures containing high amounts of sugar. SUN, in combination with other selective NMR experiments, was also used to describe and characterize all NMR signals from the minor furanose forms of glucose, even though these forms only make up about 0.2% of all glucose in water solution.

The second method presented in this thesis is an extension of the quantification method AQuA (automated quantification algorithm) and enables absolute quantification of metabolites from complicated NMR spectra with interference from both narrow and broad signals. The new, fully automated, workflow first performs an automatic baseline correction to specifically remove broad signals from NMR spectra and then uses AQuA to calculate accurate concentrations. The extended AQuA method was evaluated using simulations and by adding known amounts of a few different substances to a plant sample. The entire workflow took less than one second per spectrum when 24 metabolites were quantified.

Populärvetenskaplig sammanfattning

Kärnmagnetisk resonansspektroskopi, förkortat NMR (från engelskans nuclear magnetic resonance), är en viktig analytisk teknik inom flera olika forskningsfält. Principen bakom NMR-spektroskopi är att vissa typer av atomkärnor genererar små magnetfält som gör dem ”NMR-aktiva”. När dessa atomkärnor placeras i ett starkt, yttre magnetfält börjar de att rotera runt sin egen axel med en frekvens som är beroende av deras lokala elektrokemiska miljö. En väteatom som är bunden till en syreatom kommer till exempel att rotera med en lite högre frekvens än en väteatom som är bunden till en kolatom. Dessa två väteatomer ger sedan upphov till två olika signaler i ett NMR-spektrum, medan atomkärnor som befinner sig i exakt samma miljö observeras som en enda signal med en intensitet som är proportionell till antalet kärnor. Detta gör att NMR-spektroskopi kan användas både för att detaljstudera molekylers struktur och för att kvantifiera olika komponenter i en blandning.

Ett vanligt förekommande problem inom NMR-spektroskopi är signalinterferens, även kallat överlapp. Olika atomkärnor som befinner sig i liknande elektrokemisk miljö roterar med ungefär samma frekvens i det yttre magnetfältet och deras signaler kommer därför att överlappa med varandra i det resulterande NMR-spektrumet. Interferens gör att det kan vara svårt att urskilja enskilda signaler, speciellt när starka signaler överlappar med svaga signaler, och det kan därför bli svårt eller rentav omöjligt att identifiera och kvantifiera vissa molekyler. Ju komplexare en blandning är, d.v.s. ju fler olika molekyler blandningen innehåller och ju större skillnader i koncentration mellan molekylerna, desto större är risken för signal-interferens. De flesta biologiska prover, t.ex. växtextrakt, blod och vävnader, är väldigt komplexa och deras NMR-spektra uppvisar följaktligen mycket signalinterferens.

Problem med signalinterferens kan förhindras genom att fysiskt dela upp ett provs olika komponenter innan analys, till exempel genom att filtrera provet eller genom att använda någon typ av kromatografi, men detta är ofta både dyrt och tidskrävande och riskerar dessutom att ge missvisande resultat eftersom provet förändras i processen. Ett annat sätt att minska signalinterferens är att använda särskilda NMR-experiment eller matematiska metoder. Dessa alternativ är betydligt snabbare än en fysisk separation, särskilt som de kan appliceras på intakta prover. Denna avhandling handlar om två sådana metoder, ett NMR-experiment och ett matematiskt arbetsflöde, som båda är framtagna för att reducera signalinterferens genom att ta bort oönskade signaler från NMR-spektra.

NMR-experimentet som beskrivs i denna avhandling kallas SUN (suppression of unwanted signals) och används för att reducera specifika oönskade signaler så att andra signaler med lägre intensitet blir synliga och kan studeras. För att utvärdera och optimera SUN-experimentet applicerades det på olika prover, bland annat apelsinjuice och tre konstgjorda blandningar som innehöll stora mängder socker. SUN, i kombination med andra selektiva NMR-experiment, användes även för att beskriva och karakterisera alla NMR-signaler från de ovanliga furanosformerna av glukos, trots att dessa former bara utgör ca 0,2 % av all glukos i vattenlösning.

Den andra metoden som presenteras i denna avhandling är en utvidgning av kvantifieringsmetoden AQuA (automated quantification algorithm) och möjliggör absolutkvantifiering av metaboliter från komplicerade NMR-spektra med interferens från både smala och breda signaler. Det nya, helt automatiserade, arbetsflödet utför först korrigering av baslinjen för att specifikt ta bort breda signaler från NMR-spektra och använder därefter AQuA för att beräkna korrekta koncentrationer. Den utökade AQuA-metoden utvärderades med hjälp av simuleringar och genom att spetsa ett växtprov med kända halter av några olika ämnen. Hela arbetsflödet tog mindre än en sekund per spektrum när 24 metaboliter kvantifierades.

Acknowledgements

In this last chapter, I will try my best to thank and acknowledge everyone who has contributed to my time as a PhD student, in one way or another.

First, I would like to thank my supervisors. When I was offered this PhD position, I looked up the main supervisor, Corine Sandström, and found that she had been awarded a prize for her supervising skills. That piece of information was an important reason why I dared to become a PhD student and, needless to say, I think the prize was well deserved. I am very grateful to have had you as my supervisor, Corine. Not only are you an excellent scientist but your kind, generous and considerate manners make you a true role model in the competitive academic world. I feel a little sad about being your last PhD student but I am sure you will enjoy your upcoming retirement life in France.

Gustav Nestor, together we uncovered the glucofuranose NMR data no one knew they needed. Your relentless enthusiasm about NMR and carbohydrates is inspiring! Also, thank you for your witty sarcasm and all the advice you have given me over the years, about science, career, and babies.

Anders Broberg, thank you for your patience with my LC-MS confusion and for carefully proofreading my manuscripts. Although the peptide project didn't proceed as planned, I still learned a lot.

Johan Meijer, thank you for your curiosity about my projects and for taking your time whenever I had questions, even after your retirement. I always found your comments and suggestions regarding my manuscripts very helpful.

Moving on to other (current and former) members of the organic chemistry group:

Hanna Eriksson Röhnisch, my friend, collaborator and unofficial co-supervisor, thank you for all your help and support. I enjoyed working together on the extended AQuA project, but not nearly as much as I have enjoyed the countless chats, laughs and discussions we have had over the years. Your ability to always try to see things from a positive point of view has been a useful, although oftentimes challenging, lesson for me.

Yan Xue, I miss you at SLU. I really appreciated having you by my side from our very first day as PhD students, not only at SLU but also during the 15(!) credit course in separation and mass spectrometry at Uppsala University, our trips to different conferences (which would have been more numerous if not for covid and the US Visa application process), and the times we went to the "beach" in Gamla Uppsala.

Tarja Wiegel and Luying Wang, thank you for bringing new energy into the group! I wish both of you the best of luck with your remaining PhD journeys.

Anders Sandström, I enjoyed our chats and teaching-related discussions and that you were always open to my various course improvement ideas. Also, thank you for your meticulous feedback regarding the Swedish sections (and colour choices…) of this thesis.

Jan Eriksson, thank you for helping me with the Orbitrap, your feedback and enthusiasm about the extended AQuA, and for sharing your anecdotes.

Peter Agback, thank you for helping me with the NMR and for enlightening us all about the academic traditions of Uppsala.

Tatiana Agback, I'm impressed by your deep NMR knowledge as well as your storytelling skills.

My long-lasting office mate, the MS oracle Suresh Gohil, thank you for your kindness. I hope you will not have to exceed your 4 hour working week too often in the future.

Johnny Östman and Christina Nord, the conversation topics in the fika room certainly changed after you left. Thank you for all the fun, including the times when you introduced us to various strange beverages (I especially remember the mustard and chili vodkas...).

Thanks to my other former and current office mates for keeping me company in D254. Frida Wende, although it's been some years since you left, your

guardian angel is still watching me from the ceiling lamp above my desk. Mathilde Brunel, I hope you're happy in Bonn. Thank you for many nice conversations – I enjoyed having a fellow cat lover in the office! Johanna Östlund, thank you for the life-saving baby carrier and the other things you gave us. I look forward to trying your yoghurt one day. Nazila Dardmeh, it was nice having you here. Good luck with your thesis and defence!

Thank you Anja Herneke, Klara Nilsson, Solja Pietiäinen, Mathias Johansson, Ani Vardanyan, Jaqueline Auer, and Sara Targonska for great company in the fika room and various social activities outside of work.

Thanks to all other current residents of the D2 corridor for a good working environment: Alejandra Fernandez Castaneda, Björn Greijer, Filip Kozlowski, Galia Zamaratskaia, Gulaim Seisenbaeva, Henrik Hansson, Ingmar Persson, Jing Lu, Maud Langton (thanks for agreeing to chair my defence!), Oksana Dudarko, Saeid Karkehabadi, Troy Breijaert, and Vadim Kessler.

I would also like to acknowledge some previous colleagues who contributed to a nice atmosphere in the corridor and the fika room over the years: Ali Moazzami, Daniel Johansson, Daniel Lundberg, Eva Bajnoczi, Fredric Svensson, Gunnar Almkvist, Ievgen Pylypchuk, Lena Lundqvist, Marijana Lakić, Rasmus Björk, and Tahereh Jafari. Fredric, I appreciated our small gossiping/complaining sessions and our infrequent workouts at Friskis & Svettis.

There are of course also people outside of the D2 corridor that I need to thank. Janicka Nilsson and Jolanta Levenfors, thank you for helping me with various laboratory tasks in the peptide project. Monika Johansson, thank you for being such an engaged director of studies and always supporting the PhD students. Gunilla Mårsäter, thank you for making the course lab exercises run so smoothly. Nils Mikkelsen and Mikołaj Chmielarz, thank you for the technical support. Tack till institutionens administrativa personal, särskilt Anna Weinheimer och Erica Häggström, för att ni hjälpte mig med fakturor och hanterade mina undermåligt ifyllda reseräkningar. Finally, thanks to all my other colleagues at the department, especially my fellow PhD students, for making Biocentrum such a nice workplace.

Tack till min ”gamla” familj – mamma, pappa och Anna – för ert stöd genom åren. Jag hoppas att vi får en härlig sommar i Gamla Köpstad!

Tack till min nya lilla familj: David och Alvar (samt Kalle och Magda!). Jag är otroligt tacksam för allt ditt stöd och de uppoffringar du har gjort för min skull, David. Det är ingenting jag tar för givet. Nu ser jag fram emot att bygga upp en mer permanent tillvaro tillsammans. Alvar, om du läser det här någon gång i framtiden så ska du veta att även om du inte direkt gjorde det lättare att skriva avhandling så gjorde du det desto roligare att komma hem på kvällarna. Nu kommer vi äntligen att få mer tid tillsammans!

I

RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2020, **10**, 32511

Received 7th August 2020
Accepted 24th August 2020

DOI: 10.1039/d0ra06828d

rsc.li/rsc-advances

Band-selective NMR experiments for suppression of unwanted signals in complex mixtures†

Elin Alexandersson, Corine Sandström, Lena C. E. Lundqvist and Gustav Nestor *

Band-selective NMR experiments are presented that allow selective suppression of unwanted signals (SUN) from the spectra of complex metabolite mixtures. As a result, spectral overlap and dynamic range problems are substantially reduced and low-intensity signals normally covered by dominant signals can be observed. The usefulness of the experiments is exemplified with selective suppression of sugar signals from the NMR spectra of fruit juice and a plant sample. Other possible applications include blood, milk, and wine samples.

Introduction

Biological samples typically contain numerous different metabolites with large variations in concentration, where certain compounds are present in considerably higher concentration than others. This complexity causes severe spectral overlap and dynamic range problems when the samples are analysed by NMR spectroscopy. For instance, NMR spectra of plant extracts, fruit juices, blood, milk, and tissues are typically dominated by sugars such as glucose, fructose, sucrose, and lactose, meaning that other metabolite signals in the spectral region ~3–5 ppm are obscured by the sugar signals. Sample pre-treatment, *e.g.* chromatography, is therefore often needed before the NMR analysis to enable low-abundant metabolites to be studied. Alternative ways to study complex mixtures by NMR have been developed. One method is to record a second spectrum of the sample where an extra amount of the abundant metabolite has been added and then calculate the difference between the two spectra.[1] However, this requires sample manipulation, highly stable conditions, and extensive spectral fitting. Another way to remove carbohydrates is to chemically degrade them by adding an oxidative agent to the sample before analysis.[2] Although allegedly efficient, this strategy is irreversible and might break down certain non-carbohydrate molecules as well. Other approaches include methods based on band-selective excitation of the spectral area of interest[3,4] (often combined with statistical analyses[5]), computational methods,[6] and selective experiments utilizing differences in relaxation, diffusion, and *J*-coupling, combined with mathematical modelling.[7,8]

Most of the cited strategies require statistical and computational analyses and/or that the sample is altered in some way. Our aim was to develop an NMR-based approach requiring minimal sample manipulation and computational work that still gives information about as many compounds in a mixture as possible. In 1999, Rutherford *et al.* suggested an NMR experiment for selective removal of benzylic methylene signals from the spectra of benzyl ether-protected carbohydrates.[9] The method is based on the excitation sculpting pulse sequence[10] and uses band-selective pulses to defocus all signals in a selected region of the spectrum. This first step is followed by a TOCSY spin-lock that restores any signal in the selected region that is *J*-coupled to a signal not affected by the excitation sculpting sequence. 2D extensions of this method were later developed for the same purpose[11] and the principle, albeit with a different pulse sequence, has also been employed to selectively suppress signals from water[12] and polyethylene glycol.[13]

We show here that modified versions of the original pulse sequences[9,11] can efficiently be used to selectively remove signals from dominant compounds, *e.g.* sugars, in the spectra of complex metabolite mixtures. Thereby, signals that are otherwise hidden by the dominant signals can be observed. We call the approach SUN, Suppression of UNwanted signals. The experiments can be applied to any type of sample where certain compounds are present in excess and cause spectral overlap and/or dynamic range problems.

Results and discussion

An overview of the pulse sequences used in this work is shown in Fig. 1. The first one (Fig. 1A) is a modified version of the methods described above.[9,11] The second pulse sequence (Fig. 1B) utilizes the opposite strategy, *i.e.* band-selective excitation of the regions of the spectrum that do not contain the metabolite(s) to be suppressed. The double-pulsed field gradient spin-echo (DPFGSE) suppression or excitation is followed by a TOCSY spin-lock, here DIPSI-2 [14] with zero-quantum coherence suppression.[15] As indicated in the figure, the pulse sequences can be used in conjunction with *e.g.* TOCSY, HSQC,

Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden. E-mail: gustav.nestor@slu.se

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0ra06828d

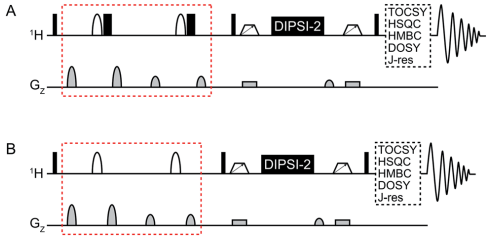


Fig. 1 General overview of the SUN pulse sequences used for (A) band-selective suppression and (B) band-selective excitation. The difference between the two versions is marked in red. Narrow black rectangles denote 90° hard pulses while wide black rectangles denote 180° hard pulses. Shaped white bars represent band-selective inversion pulses. White trapezoids with arrows denote chirp pulses aimed for suppression of zero-quantum coherences.[15] DIPSI-2 is used for isotropic mixing.[14]

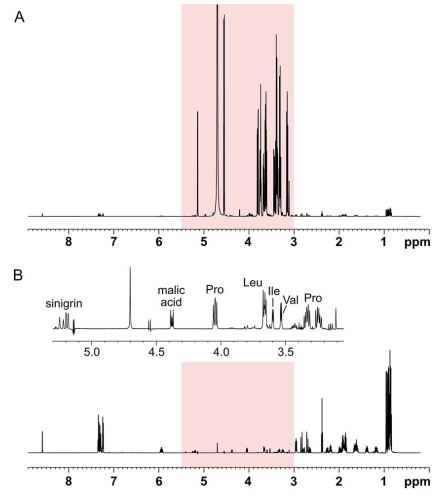


Fig. 2 SUN applied to an artificial mixture containing 100 mM D-glucose and 1 mM of proline, leucine, isoleucine, valine, histidine, phenylalanine, γ-aminobutyric acid, choline, malic acid, citric acid, ascorbic acid, and sinigrin. (A) 1D-^{1}H spectrum where the region targeted for band-selective suppression is highlighted in red, (B) 1D-SUN with suppression of the region 3.0–5.5 ppm. The inset shows an expansion of the region targeted by the band-selective pulse.

or HMBC to obtain various 2D experiments with band-selective suppression or excitation. If desired, the band-selective inversion pulse can be designed to target several regions of the spectrum at the same time. It is crucial that all signals from the dominant compound(s) are suppressed in the DPFGSE step; otherwise they will also be restored during the spin-lock.

The performance of the SUN approach was evaluated using an artificial mixture containing twelve common plant metabolites (proline, leucine, isoleucine, valine, histidine, phenylalanine, γ-aminobutyric acid, choline, malic acid, citric acid, ascorbic acid, and sinigrin) in equal concentrations and glucose in higher concentration. The ratio between glucose and the other metabolites was varied between 10 : 1 and 1000 : 1. Using SUN with band-selective suppression, it was possible to suppress the glucose signals almost entirely (≥98%) (Fig. 2). Several other signals in the area were retained, including the alpha protons of proline, valine, leucine, and isoleucine, of which the latter two were previously buried under the glucose signals (Fig. 2B).

When band-selective excitation was used, almost identical results were obtained in terms of glucose suppression, resolution, and signal-to-noise ratio (Fig. 3A and B) (note that Fig. 2 shows a mixture with the proportion 100 : 1 whereas it is 1000 : 1 in Fig. 3). However, the performance of this pulse sequence appeared to be more sensitive to the exact positioning of the selective pulse than the version with band-selective suppression. Thus, both the width of the spectral regions chosen for excitation and their location relative to each other highly influenced the suppression efficiency. The best suppression of glucose signals was obtained when the two spectral regions excited simultaneously were of the same width.

Although spectral overlap is less of a problem in two-dimensional spectra, abundant compounds can still prevent identification of other compounds present in lower concentration. Therefore, 2D analogues of the SUN pulse sequences were developed, including TOCSY, HSQC, HMBC, DOSY, and *J*-resolved spectroscopy with band-selective suppression or excitation. The performance of the 2D experiments is illustrated with 2D-SUN-TOCSY applied to the 1000 : 1 artificial mixture, carried out using band-selective excitation to selectively remove glucose (Fig. 3C and D). An HSQC version was evaluated as well (see Fig. S2†). It is apparent that the glucose suppression obtained in the 2D experiments highly resembles that of their 1D counterparts. Apart from the possibility of detecting analytes in the glucose region, the improved receiver gain achieved both in the 1D and 2D experiments when suppressing the glucose signals significantly improves identification of low-concentration analytes throughout the spectra.

As expected, the length of the TOCSY spin-lock and the size of the *J*-coupling between the suppressed signals and their non-suppressed neighbours determine to what extent signals are recovered in the TOCSY step. Since signal intensity is also influenced by the relaxation rate of the individual spins, a compromise might be needed so that the mixing time is long enough for TOCSY transfer to take place, but sufficiently short to avoid signal attenuation due to relaxation. In the examples presented here, the intensity of the recovered signals did not accurately reflect the actual concentration of the compounds, meaning that quantitative analyses may require calibration curves to determine the correlation between concentration and signal intensity.

Since only signals that are *J*-coupled to another signal located outside the targeted area can be reintroduced by the TOCSY step, some non-glucose signals are missing from the band-selective spectra presented in Fig. 2 and 3. For instance,

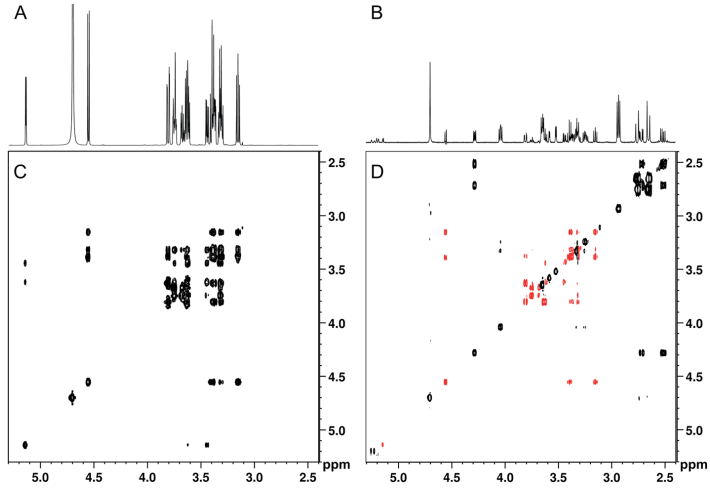


Fig. 3 SUN applied to an artificial mixture containing 100 mM D-glucose and 0.1 mM of proline, leucine, isoleucine, valine, histidine, phenylalanine, γ-aminobutyric acid, choline, malic acid, citric acid, ascorbic acid, and sinigrin. (A) 1D-^{1}H spectrum, (B) 1D-SUN with excitation of the spectral regions 5.8–8.6 ppm and −0.2-2.6 ppm, (C) 2D-TOCSY, (D) 2D-SUN-TOCSY with excitation of the same regions as in (B) and remaining glucose signals coloured red.

the non-aromatic protons of histidine and phenylalanine were not restored during the spin-lock. To obtain a more specific suppression of glucose with minimal impact on the other compounds, a band-selective inversion pulse targeting two separate areas of the spectrum – 3.1–3.9 ppm (the glucose ring protons) and 4.5–5.3 ppm (the anomeric protons) – was used. Thereby, additional signals were retained, both in the area in-between the glucose regions and directly underneath the

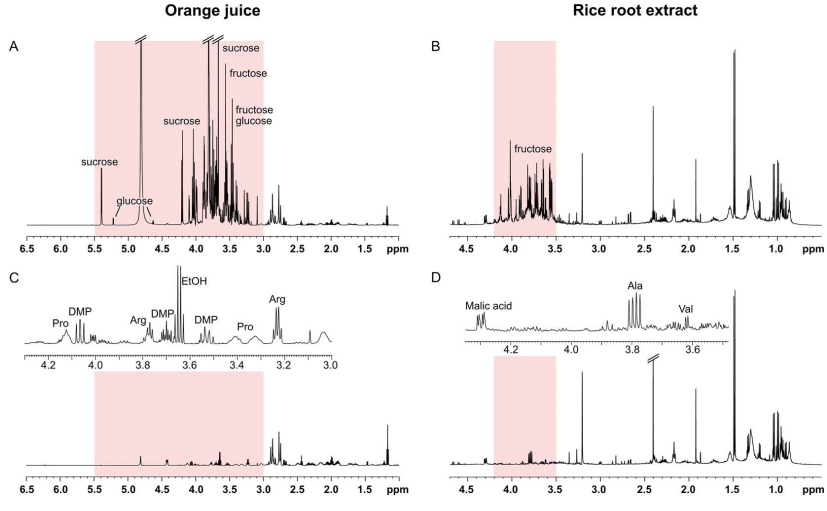


Fig. 4 SUN applied to orange juice (left) and a rice root extract (right). (A) 1D-^{1}H spectrum of orange juice recorded using excitation sculpting to suppress the water signal. (B) 1D-^{1}H spectrum of a rice root extract recorded using water presaturation. (C) 1D-SUN spectrum of orange juice with band-selective suppression of the region 3.0–5.5 ppm (highlighted in red). The inset shows an expansion of the sugar ring proton region 3.0–4.3 ppm. DMP = dimethylproline. (D) 1D-SUN spectrum of the rice root extract with band-selective suppression of the region 3.5–4.2 ppm (highlighted in red). The inset shows an expansion of the fructose region 3.45–4.35 ppm.

glucose signals (see Fig. S3†). Non-aromatic protons belonging to histidine and phenylalanine could then be identified, as well as the methylene protons from choline. Unfortunately, suppressing two separate areas produced more severe phase distortions, mainly affecting glucose, than with only one area targeted. Both the suppression efficiency and the phase distortions were affected by the width and position of the selective pulse, as well as the pulse phase. Similar to when band-selective excitation was used, the best results were obtained when the two targeted regions were of equal size.

The SUN experiments were also applied to two authentic samples: orange juice and a root extract from rice (Fig. 4). Fruit juices are complex, sugar-rich mixtures that are often analysed by NMR for quality control reasons.[16,17] As can be seen in Fig. 4A, the orange juice spectrum was dominated by sucrose, fructose, and glucose. The sugar signals could be completely suppressed using the SUN pulse sequences which enabled identification of other metabolites in the sugar region, including ethanol, arginine, and dimethylproline (Fig. 4C). These metabolites were identified based on 1D and 2D experiments (see Fig. S4†) as well as already published orange juice signal assignments.[17–19] It is also worth noting that the water signal was highly reduced in the band-selective experiment, even though no additional water suppression was used.

The other sample had been collected from the soil surrounding a rice plant root. Plant roots excrete a wide array of different compounds into the soil, including sugars, organic acids, and amino acids.[20] In the spectra of the sample used here, the sugar region was dominated by fructose (Fig. 4B). Although not present in large excess, the fructose signals obstructed identification of other metabolites in the region 3.4–4.0 ppm. Since the fructose signals are all concentrated to a narrow part of the spectrum, a selective pulse targeting only this region was used in the SUN experiment. The water signal is residing outside of this region, and therefore a presaturation step was added to the pulse sequence for water suppression. The resulting spectrum is shown in Fig. 4D. With fructose removed, the signals belonging to the alanine and valine alpha protons became clearly visible (Fig. 4D). Two-dimensional TOCSY was performed for signal assignment, with similar fructose suppression as in the 1D experiment.

Conclusions

Here we have presented two NMR experiments that are promising for suppression of unwanted signals in the spectra of complex mixtures. Using the SUN approach, the intensity of dominant signals can be dramatically reduced or completely suppressed while other signals in the targeted spectral regions are retained *via J*-coupling. Thus, problems caused by spectral overlap and a too large concentration range can be solved without physically altering the sample. In the most favourable cases, the approach can be used to identify compounds whose signals outside of the suppressed area are highly overlapped. The experiments are fast and easy to use and can be readily applied to practically any sample that contains high-intensity signals localized to just one or a few regions of the NMR spectrum. Samples where this approach would be advantageous include blood plasma, fruit juices, and different plant extracts. Both SUN versions work well to suppress sugar signals and the choice of which one to use depends on the experimental aim and sample type. We believe that both versions can become valuable tools in the study of complex mixtures.

Experimental

Sample preparation

Artificial mixtures. The mixtures were prepared with the following compounds (all purchased from Merck): D-glucose, DL-proline, L-leucine, L-isoleucine, L-valine, L-phenylalanine, L-histidine, γ-aminobutyric acid, choline chloride, malic acid, citric acid, ascorbic acid, and sinigrin hydrate. The concentration of D-glucose was either 100 mM or 1000 mM while the concentration of the other compounds was 10 mM, 1 mM, or 0.1 mM. All samples were prepared in D_2O with a final volume of 600 μl.

Orange juice. Orange juice (1 ml) bought at a local supermarket was centrifuged at 13 500 rpm for 5 minutes. 540 μl of the supernatant was then mixed with 60 μl D_2O in an NMR tube.

Rice root sample. The root from an eight-week old rice plant was removed from the soil and vortexed for 1 min in 30 ml MilliQ water, after which the solution was freeze-dried. 100 mg of the freeze-dried material was ultra-sonicated together with 8 ml methanol for 10 min and then centrifuged for 10 min at 5000 rpm. The supernatant was freeze-dried, after which 380 μl MilliQ water was added. After vortexing, the sample was ultra-sonicated for 10 min and then centrifuged for 10 min at 5000 rpm. 350 μl of the supernatant was then mixed with 50 μl D_2O, 20 μl MilliQ water, 150 μl 0.4 M phosphate buffer (pH 7), and 30 μl TSP internal standard (5.8 mM). To concentrate the sample further, two identical samples prepared as described above were pooled together, freeze-dried, and then dissolved in 150 μl D_2O for NMR analysis.

NMR experiments

All spectra were acquired on a Bruker Avance III 600 MHz spectrometer with a 5 mm $^{1}H/^{13}C/^{15}N/^{31}P$ inverse detection cryoprobe equipped with a *z* gradient with a maximum nominal gradient strength of 48.1 G cm^{-1}. Spectra were recorded at 25 °C and were processed with TopSpin 4.0.6. The ^{1}H spectral window was set to 9 ppm or 12 ppm and the ^{13}C spectral window was set to 110–122 ppm (HSQC). The carrier frequency was placed on the water signal (4.70 ppm). Further details about the NMR experiments are provided in the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Authors acknowledge Dr Yunkai Jin for preparing the root sample and Dr Hanna Eriksson Röhnisch for assisting the

analysis of this sample. Dr Helena Kovacs at Bruker BioSpin is acknowledged for fruitful discussions.

Notes and references

1 T. Ye, C. Zheng, S. Zhang, G. A. N. Gowda, O. Vitek and D. Raftery, *Anal. Chem.*, 2012, **84**, 994–1002.
2 J. Yuan, B. Zhang, C. Wang and R. Brüschweiler, *Anal. Chem.*, 2018, **90**, 14100–14104.
3 F. Rastrelli, E. Schievano, A. Bagno and S. Mammi, *Magn. Reson. Chem.*, 2009, **47**, 868–872.
4 M. Koda, K. Furihata, F. Wei, T. Miyakawa and M. Tanokura, *Magn. Reson. Chem.*, 2011, **49**, 710–716.
5 P. Sandusky and D. Raftery, *Anal. Chem.*, 2005, **77**, 2455–2463.
6 K. Bingol and R. Brüschweiler, *Anal. Chem.*, 2011, **83**, 7412–7417.
7 G. Dal Poggetto, L. Castañar, R. W. Adams, G. A. Morris and M. Nilsson, *Chem. Commun.*, 2017, **53**, 7461–7464.
8 G. Dal Poggetto, L. Castañar, R. W. Adams, G. A. Morris and M. Nilsson, *J. Am. Chem. Soc.*, 2019, **141**, 5766–5771.
9 T. J. Rutherford, K. P. R. Kartha, S. K. Readman, P. Cura and R. A. Field, *Tetrahedron Lett.*, 1999, **40**, 2025–2028.
10 T. L. Hwang and A. J. Shaka, *J. Magn. Reson., Ser. A*, 1995, **112**, 275–279.
11 K. E. Kövér, K. Fehér, L. Szilágyi, A. Borbás, P. Herczegh and A. Lipták, *Tetrahedron Lett.*, 2000, **41**, 393–396.
12 M. Liu, H. Tang, J. K. Nicholson and J. C. Lindon, *J. Magn. Reson.*, 2001, **153**, 133–137.
13 N. Prosa, M.-C. Scherrmann, D. Merlet and J. Farjon, *J. Magn. Reson.*, 2013, **237**, 63–72.
14 S. P. Rucker and A. J. Shaka, *Mol. Phys.*, 1989, **68**, 509–517.
15 M. J. Thrippleton and J. Keeler, *Angew. Chem., Int. Ed.*, 2003, **42**, 3938–3941.
16 M. Spraul, B. Schütz, P. Rinke, S. Koswig, E. Humpfer, H. Schäfer, M. Mörtter, F. Fang, U. C. Marx and A. Minoja, *Nutrients*, 2009, **1**, 148–155.
17 G. Le Gall, M. Puaud and I. J. Colquhoun, *J. Agric. Food Chem.*, 2001, **49**, 580–588.
18 C. R. de Oliveira, R. L. Carneiro and A. G. Ferreira, *Food Chem.*, 2014, **164**, 446–453.
19 E. G. Alves Filho, F. D. L. Almeida, R. S. Cavalcante, E. S. de Brito, P. J. Cullen, J. M. Frias, P. Bourke, F. A. N. Fernandes and S. Rodrigues, *Food Chem.*, 2016, **204**, 102–107.
20 F. D. Dakora and D. A. Phillips, *Plant Soil*, 2002, **245**, 35–47.

Electronic Supplementary Material (ESI) for RSC Advances.

Band-selective NMR experiments for suppression of unwanted signals in complex mixtures

Elin Alexandersson, Corine Sandström, Lena C.E. Lundqvist and Gustav Nestor*

Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

Electronic supplementary information

1. Pulse sequences

The pulse sequences for the **1D SUN** experiments are shown in Fig. S1a and S1b. Narrow black rectangles denote 90° hard pulses while wide black rectangles denote 180° hard pulses. Shaped white bars represent band-selective inversion or excitation 180° pulses. IBURP-2 shapes were used for both suppression and excitation experiments. Typically, a 3.5 ms IBURP pulse centred in the middle of the glucose region yields a bandwidth of 1500 Hz, which is enough for suppression of glucose. The Bruker WaveMaker tool was used for construction of shaped pulses targeting more than one region of the spectrum. TOCSY transfer is achieved by using the DIPSI-2 mixing scheme with a mixing time of 10-150 ms depending on the spin system (here, 100 ms was used for the rice root sample whereas 50 ms was used for the orange juice and the artificial mixtures). White trapezoids with arrows denote low-power 180° chirp pulses of 20 kHz bandwidth aimed for suppression of zero-quantum coherences. Their durations were set to 20 and 15 ms before and after the DIPSI-2 mixing, respectively. All gradient pulses except for G_0 have a duration of 1 ms. G_1 and G_2 (amplitude of 14.9 and 5.3 G cm^{-1}, respectively) are used to eliminate magnetization of signals within the bandwidth of the selective pulse (band-selective suppression, Fig S1a) or outside of the bandwidth of the selective pulse (band-selective excitation, Fig. S1b). G_3 is a spoil gradient pulse with an amplitude of 3.4 G cm^{-1}. G_0 (amplitude of ca. 2.4 and 3.2 G cm^{-1} before and after the DIPSI-2 mixing, respectively) is a weak pulsed field gradient applied simultaneously with the chirp pulses to suppress zero quantum coherence. All gradient pulses are followed by a recovery delay of 200 μs. The phase cycle is given in Table S1.

To achieve optimal suppression of unwanted signals, it might be necessary to adjust the 90° pulse length and the precise phases of the band-selective pulses (by phase corrections). The TOCSY mixing time might also need some optimization in order to obtain the best recovery of remaining signals.

The 2D TOCSY and HSQC versions of the SUN experiments are shown in Fig. S1c and S1d. The **2D SUN-TOCSY** experiment has two TOCSY transfers, where the first is applied to restore signals in the sugar region of the spectrum and the second is applied to yield TOCSY correlations in the f1 dimension. The two TOCSY steps could be used with different mixing times, but for the examples herein, the same mixing time was used. The 2D TOCSY was run in the States-TPPI manner, with phases presented in Table S1.

The HSQC step of the **2D SUN-HSQC** experiment is almost identical to the Bruker pulse program *hsqcedetgpsisp.2*, which is a multiplicity-edited HSQC with PEP sensitivity enhancement and adiabatic inversion and refocusing pulses. Chirp pulses of 20 kHz bandwidth and 500 us duration were utilized for 180° ^{13}C inversion and composite chirp pulses of 2 ms duration for 180° ^{13}C refocusing. Delays were set to τ = 1.7 ms, Δ = 3.45 ms, Δ' = 862 μs, and δ = 1.2 ms. G_4 and G_5 amplitudes were set to 38.5 and 9.7 G cm^{-1}, respectively, with a duration of 1 ms, followed by a recovery delay of 200 μs. ^{13}C decoupling was obtained with the GARP-4 decoupling scheme. The experiment was run in the echo-antiecho mode, with gradient selection obtained by the reversal of the G_4 gradient pulse. The phase cycle is given in Table S1.

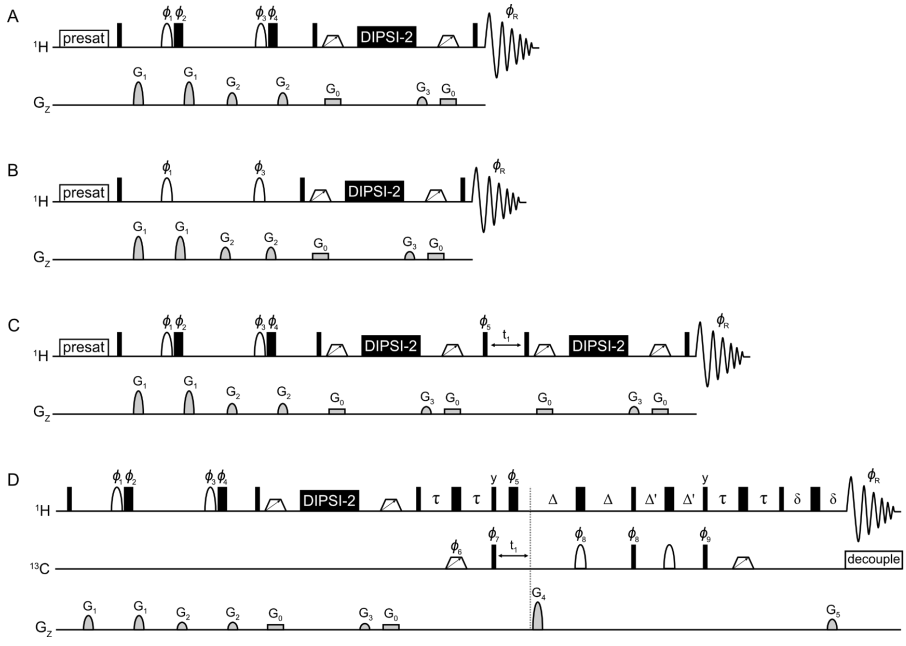


Fig. S1. Pulse sequences of (a) 1D SUN with band-selective suppression, (b) 1D SUN with band-selective excitation, (c) 2D SUN-TOCSY with band-selective suppression, (d) 2D SUN-HSQC with band-selective suppression. Phases are listed in Table S1. Unless indicated otherwise, pulses are applied along the x-axis. Band-selective excitation versions of the SUN-TOCSY and SUN-HSQC versions are obtained by eliminating the ϕ_2 and ϕ_4 180° ^{1}H pulses.

Table S1. Phase cycling of pulse sequences in Fig S1 (A-D).

	A	**B**	**C**[a]	**D**[b]
ϕ_1	x, y, -x, -y	x, y, -x, -y	x, y, -x, -y	x, y, -x, -y
ϕ_2	-x, -y, x, y	-	(-x, -y, x, y)[c]	(-x, -y, x, y)[c]
ϕ_3	x_4, y_4, $(-x)_4$, $(-y)_4$	x_4, y_4, $(-x)_4$, $(-y)_4$	x_4, y_4, $(-x)_4$, $(-y)_4$	x_4, y_4, $(-x)_4$, $(-y)_4$
ϕ_4	$(-x)_4$, $(-y)_4$, x_4, y_4	-	$((-x)_4$, $(-y)_4$, x_4, $y_4)$[c]	$((-x)_4$, $(-y)_4$, x_4, $y_4)$[c]
ϕ_5	-	-	x, -x	x_2, $(-x)_2$
ϕ_6	-	-	-	x
ϕ_7	-	-	-	x_8, $(-x)_8$
ϕ_8	-	-	-	x_2, $(-x)_2$
ϕ_9	-	-	-	y_2, $(-y)_2$
ϕ_R	$(x, -x)_2$, $(-x, x)_2$	$(x, -x)_2$, $(-x, x)_2$	x_4, $(-x)_4$	$(-x)$, x_2, $(-x)$, $(x, (-x)_2$, $x)_2$, $(-x)$, x_2, $(-x)$

[a] Quadrature in the t_1 dimension is obtained by incrementing ϕ_5 with 90°. [b] Quadrature in the t_1 dimension is obtained by incrementing ϕ_6, ϕ_7, ϕ_9, and ϕ_R with 180°. [c] For band-selective suppression.

2. Supplementary spectra

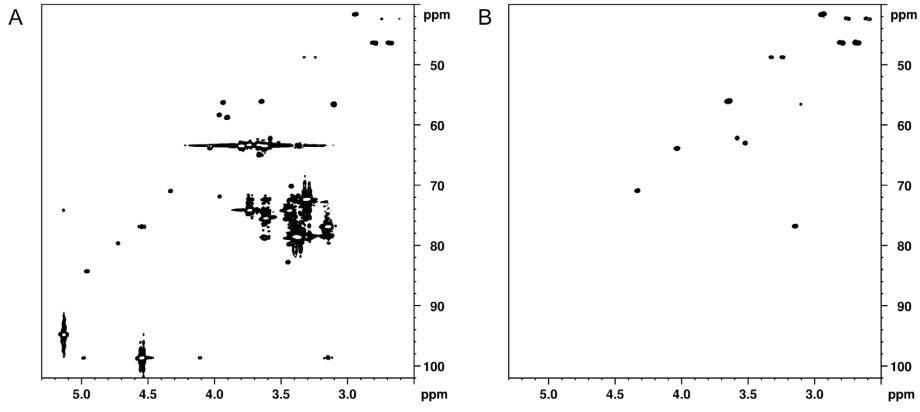


Fig. S2. (a) HSQC of the artificial mixture containing 100 mM glucose and 1 mM of the other metabolites, (b) SUN-HSQC with band-selective excitation of the spectral regions 5.8-8.6 ppm and -0.2-2.6 ppm. Data were recorded with 64 scans, 1024 data points in t_2 and 256 increments in t_1, with an experimental time of 7 hours in a) and 6.5 hours in b). The data were zero-filled before applying a $\pi/2$ shifted sine-squared bell function in both dimensions.

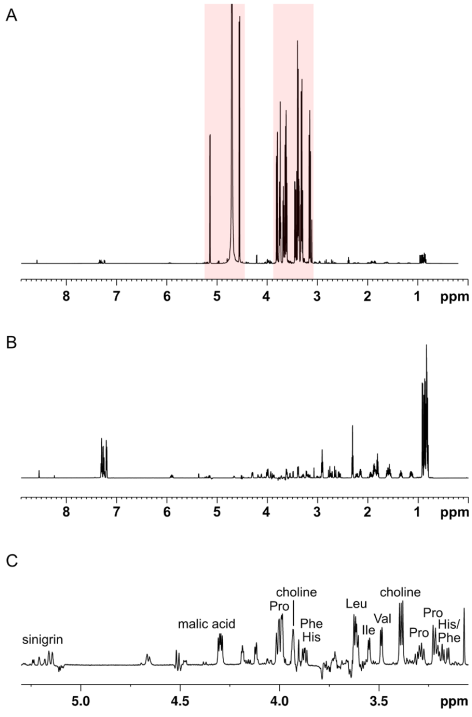


Fig. S3. 1D SUN spectra with band-selective suppression targeting two spectral regions, 4.45-5.25 ppm and 3.08-3.88 ppm. (a) 1D-^{1}H spectrum with the regions targeted for band-selective suppression highlighted in red, (b) 1D-SUN spectrum with band-selective suppression, (c) enlargement of the region 3.0-5.3 ppm in b). Data were recorded with 128 scans and 64k points, with an experimental time of 13 minutes. The data were zero-filled before applying an exponential function with 0.3 Hz line broadening.

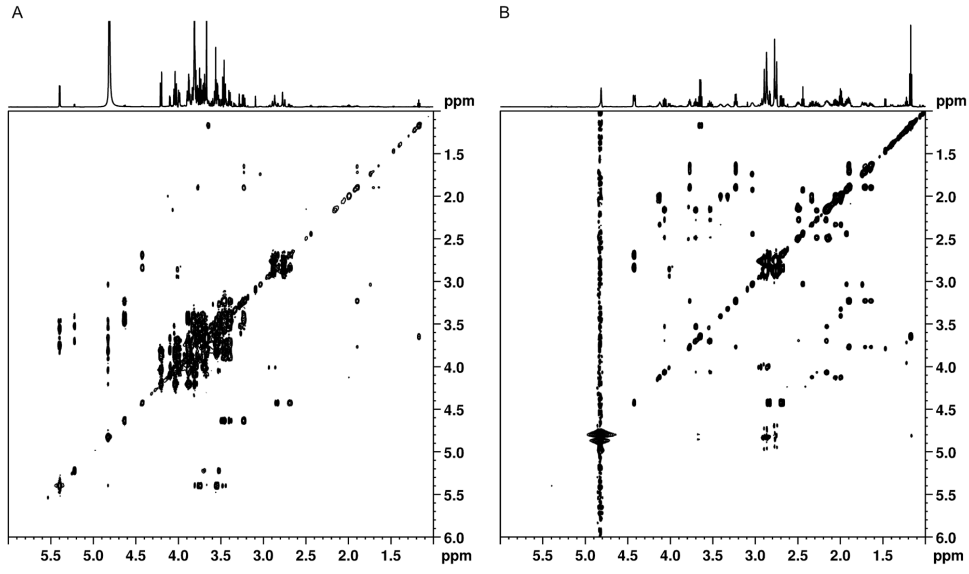


Fig. S4. 2D-TOCSY of orange juice. (a) 2D-TOCSY, (b) 2D-SUN-TOCSY with band-selective suppression of the spectral region 3.0-5.5 ppm. Data were recorded with 32 scans, 2048 data points in t_2 and 256 increments in t_1, with an experimental time of 3.5 hours. The data were zero-filled before applying a $\pi/2$ shifted sine-squared bell function in both dimensions.

3. Pulse Sequence (Bruker format)

Quick guide for starting-up

1. Copy the pulse program from this document to a text file, save it as "SUN1d" and put it in the folder named Bruker\Topspin(X.X)\exp\stan\nmr\lists\pp\user.
2. Run a ^{1}H NMR experiment and decide whether band-selective suppression or excitation will be used.
3. Define peak(s) in the middle of the desired region(s) for band-selective suppression or excitation.
4. Make a copy of the ^{1}H NMR experiment and change the pulse program to "SUN1d".
5. Set probe/solvent dependent parameters with the command "getprosol" and fill in acquisition parameters that are not automatically adjusted, such as gradient files and gradient strengths. The correct settings are provided in the pulse program.
6. Copy the peak list from the ^{1}H NMR experiment to the SUN experiment. This can be done with AU programs such as copyPL or getPL (can be found at the Bruker web library).
7. The default setting is band-selective suppression without water presaturation. To use band-selective excitation instead, add "-DEXCIT" in the ZGOPTNS field. For water presaturation, add "-DPRESAT" in the ZGOPTNS field. If both excitation and water presaturation are desired, add both ("-DEXCIT -DPRESAT") in the ZGOPTNS field.
8. Define the shape function as userA1, for instance iburp2. (For Topspin versions older than 3.6, the shape function must be defined in the pulse program. In this case, change "userA1" to the desired shape function on the line starting with "sp2:wvm".)
9. Define the bandwidth (in ppm) of the band-selective suppression or excitation region(s) as cnst18. The bandwidth is equal to the width of the selected region(s) and must be the same if more than one region is selected with the peak list.
10. Define the TOCSY mixing time (d9). Typically 50-100 ms gives efficient TOCSY transfer.
11. Run "wvm –a" to calculate the selective pulse and to add the result into the current experiment set-up.
12. Check the receiver gain with rga and start the experiment with zg.
13. If signals that should be suppressed still appear, the bandwidth for band-selective suppression (cnst18) may need to be increased to allow defocusing of the entire spin system, or the band-width for band-selective excitation (cnst18) may need to be decreased to avoid excitation of the unwanted spin system(s). In band-selective excitation, the region for excitation can also be moved away from the unwanted signals by adjusting the peak list. Adjusting p1 can also improve the suppression of unwanted signals.
14. To optimize the phase of signals in the suppressed region, if necessary, phase corrections of ph2 and ph4 may be added with the "phcor" command. Try out small changes from zero to obtain optimized phase and suppression. Also p1 can be changed to optimize the phase of the spectra. In addition, gpz0 can be adjusted to optimize the suppression of zero-quantum coherences.

For further instructions on how to use WaveMaker, we refer to the Bruker WaveMaker User Manual.

1D SUN with band-selective suppression/excitation

```
;sun1d
;
;1D suppression of unwanted signals (SUN)
;   using band-selective suppression or excitation
;   using DIPSI2 sequence for TOCSY transfer
;   with presaturation during relaxation delay (optional)
;
;Use WaveMaker for definition of shaped pulses
;Define offset(s) for shaped pulses with a peak list (PL)
;The peak list must be copied from another experiment with a macro (getPL or copyPL)
;
;Band-selective suppression without water presaturation is default
;For band-selective excitation: Use ZGOPTNS = -DEXCIT
;For presaturation during relaxation delay: Use ZGOPTNS = -DPRESAT
;
;Modified 24/06/2020
;
;               Elin Alexandersson, Corine Sandström, Lena Lundqvist and Gustav Nestor
;               Swedish University of Agricultural Sciences
;
;Avance III version
;
;$CLASS=HighRes
;$DIM=1D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
"d12=20u"

"spoff29=0"

"FACTOR1=(d9/(p6*115.112))/2"
"l1=FACTOR1*2"

"d2=cnst18*p1/4"

"acqt0=-p1*2/PI"


1 ze
2 30m
  d12 BLKGRAD
  d2

#   ifdef PRESAT
  d12 pl9:f1
  d1 cw:f1 ph1
  4u do:f1
#   else
  d1
#   endif /*PRESAT*/

  d12 pl1:f1
  50u UNBLKGRAD
  (p1 ph1):f1
  3u
  p16:gp1
```

```
  d16 pl0:f1
  p12:sp2:f1 ph2:r

#   ifdef EXCIT
  3u
  d12
#   else
  3u
  d12 pl1:f1
  p2 ph3
  3u
#   endif /*EXCIT*/

  p16:gp1
  d16
  3u
  p16:gp2
  d16 pl0:f1
  p12:sp2:f1 ph4:r

#   ifdef EXCIT
  3u
  d12 pl1:f1
#   else
  3u
  d12 pl1:f1
  p2 ph5
  3u
#   endif /*EXCIT*/

  p16:gp2
  d16
  p1 ph1
  3u
  3u pl0:f1
  10u gron0
  (p32:sp29 ph1):f1
  20u groff
  d16 pl10:f1

                                                    ;begin DIPSI2
3 p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
```

```
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23
  lo to 3 times l1
                                                        ;end DIPSI2

  p16:gp3
  d16 pl0:f1
  10u gron0*1.333
  (p32*0.75:sp29 ph1):f1
  20u groff

  d16 pl1:f1
  p1 ph1

  go=2 ph31
  30m mc #0 to 2 F0(zd)
  20u BLKGRAD
exit


ph1=0
ph2=0 1 2 3
ph3=2 3 0 1
ph4=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph5=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph23=3
ph25=1
ph31=0 2 0 2 2 0 2 0


;pl0 : 0W
;pl1 : f1 channel - power level for pulse (default)
;pl9 : f1 channel - power level for presaturation
;pl10: f1 channel - power level for TOCSY-spinlock
;sp2:wvm: userA1(cnst18 ppm, PL; PA=1.0) np=1000
;sp29: f1 channel - shaped pulse (adiabatic)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p6 : f1 channel -  90 degree low power pulse
;p12: f1 channel - 180 degree shaped pulse
;p16: homospoil/gradient pulse                  [1 msec]
;p32: f1 channel - 180 degree shaped pulse (adiabatic)    [20 msec]
;     smoothed chirp (sweepwidth, 20% smoothing, 10000 points)
;d1 : relaxation delay; 1-5 * T1
;d9 : TOCSY mixing time
;d12: delay for power switching                 [20 usec]
;d16: delay for homospoil/gradient recovery
;l1: loop for DIPSI cycle: ((p6*115.112) * l1) = mixing time
;NS: 8 * n
;DS: 4

;cnst18: effective bandwidth for shaped pulse (ppm)
;PL: peak list to define offset(s) for shaped pulse
;userA1: shape function (sp2)
```

```
;phcor 2 : phase difference between power levels sp2 and pl1
;phcor 4 : phase difference between power levels sp2 and pl1

;choose p12 and sp2 according to desired selectivity or use WaveMaker



;for z-only gradients:
;gpz0: ca. 5%
;gpz1: 31%
;gpz2: 11%
;gpz3: 7%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100

                                ;preprocessor-flags-start
;EXCIT: for band-selective excitation start experiment with
;           option -DEXCIT (eda: ZGOPTNS)
;PRESAT: for presaturation during relaxation delay start experiment with
;           option DPRESAT (eda: ZGOPTNS)
                                ;preprocessor-flags-end
```

2D SUN-TOCSY with band-selective suppression/excitation

```
;suntocsy
;
;2D suppression of unwanted signals (SUN-TOCSY)
;   using band-selective suppression or excitation
;   using DIPSI2 sequence for TOCSY transfer
;   with presaturation during relaxation delay (optional)
;
;Use WaveMaker for definition of shaped pulses
;Define offset(s) for shaped pulses with a peak list (PL)
;The peak list must be copied from another experiment with a macro (getPL or copyPL)
;
;Band-selective suppression without water presaturation is default
;For band-selective excitation: Use ZGOPTNS = -DEXCIT
;For presaturation during relaxation delay: Use ZGOPTNS = -DPRESAT
;
;Modified 18/08/2020
;
;               Elin Alexandersson, Corine Sandström, Lena Lundqvist and Gustav Nestor
;               Swedish University of Agricultural Sciences
;
;Avance III version
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
"spoff29=0"

"in0=inf1"

"d0=in0/2-p1*4/3.1416"
"d2=cnst18*p1/4"
"d11=30m"
"d12=20u"

"FACTOR1=(d9/(p6*115.112))/2"
"l1=FACTOR1*2"
"FACTOR2=(d10/(p6*115.112))/2"
"l2=FACTOR2*2"

"acqt0=-p1*2/3.1416"


1 ze
2 d11
  d2

#   ifdef PRESAT
  d12 pl9:f1
  d1 cw:f1 ph1
  4u do:f1
#   else
  d1
#   endif /*PRESAT*/

  d12 pl1:f1
```

```
 50u UNBLKGRAD
 (p1 ph1):f1
 3u
 p16:gp1
 d16 pl0:f1
 p12:sp2:f1 ph2:r

#  ifdef EXCIT
 3u
 d12
#  else
 3u
 d12 pl1:f1
 p2 ph3
 3u
#  endif /*EXCIT*/

 p16:gp1
 d16
 3u
 p16:gp2
 d16 pl0:f1
 p12:sp2:f1 ph4:r

#  ifdef EXCIT
 3u
 d12 pl1:f1
#  else
 3u
 d12 pl1:f1
 p2 ph5
 3u
#  endif /*EXCIT*/

 p16:gp2
 d16
 p1 ph1
 3u
 3u pl0:f1
 10u gron0
 (p32:sp29 ph1):f1
 20u groff
 d16 pl10:f1

                                                    ;begin DIPSI2
3 p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph25
  p6*4.556 ph23
```

```
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23
  lo to 3 times l1
                                                    ;end DIPSI2

  p16:gp3*-1
  d16 pl0:f1
  10u gron0*1.333
  (p32*0.75:sp29 ph1):f1
  20u groff
  d16 pl1:f1
  p1 ph6
  d0
  p1 ph1
  3u
  3u pl0:f1
  10u gron0*-1
  (p32:sp29 ph1):f1
  20u groff
  d16 pl10:f1

                                                    ;begin DIPSI2
4 p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25
```

```
  p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23
  lo to 4 times l2
                                                        ;end DIPSI2

  p16:gp3
  d16 pl0:f1
  10u gron0*-1.333
  (p32*0.75:sp29 ph1):f1
  20u groff
  d16 pl1:f1

  4u BLKGRAD
  p1 ph1
  go=2 ph31
  d11 mc #0 to 2 F1PH(calph(ph6, +90), caldel(d0, +in0))
exit


ph1=0
ph2=0 1 2 3
ph3=2 3 0 1
ph4=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph5=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph6=0 2
ph23=3
ph25=1
ph31=0 0 0 0 2 2 2 2


;pl0 : 0 W
;pl1 : f1 channel - power level for pulse (default)
;pl9 : f1 channel - power level for presaturation
;pl10: f1 channel - power level for TOCSY-spinlock
;sp2:wvm: userA1(cnst18 ppm, PL; PA=1.0) np=1000
;sp29: f1 channel - shaped pulse (adiabatic)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p6 : f1 channel -  90 degree low power pulse
;p12: f1 channel - 180 degree shaped pulse
;p16: homospoil/gradient pulse                      [1 msec]
;p32: f1 channel - 180 degree shaped pulse (adiabatic)    [20 msec]
;     smoothed chirp (sweepwidth, 20% smoothing, 10000 points)
;d0 : incremented delay (2D)                        [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d9 : first TOCSY mixing time (back transfer)
;d10: second TOCSY mixing time (2D)
;d16: delay for homospoil/gradient recovery
;l1: loop for first DIPSI cycle: ((p6*115.112) * l1) = mixing time
;l2: loop for second DIPSI cycle: ((p6*115.112) * l2) = mixing time
;inf1: 1/SW = 2 * DW
;in0: 1/(1 * SW) = 2 * DW
;nd0: 1
;NS: 8 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEQ

;cnst18: = effective bandwidth for shaped pulse (ppm)
;PL: peak list to define offset(s) for shaped pulse
```

```
;userA1: shape function (sp2)

;phcor 2 and 4: phase difference between power levels sp2 and pl1

;choose p12 and sp2 according to desired selectivity or use WaveMaker


;for z-only gradients:

;gpz0: ca. 5%
;gpz1: 31%
;gpz2: 11%
;gpz3: 7%

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100


                                ;preprocessor-flags-start
;EXCIT: for band-selective excitation start experiment with
;           option -DEXCIT (eda: ZGOPTNS)
;PRESAT: for presaturation during relaxation delay start experiment with
;           option DPRESAT (eda: ZGOPTNS)
                                ;preprocessor-flags-end
```

2D SUN-HSQC with band-selective suppression/excitation

```
;sunhsqc
;
;2D suppression of unwanted signals (SUN-HSQC)
;   using band-selective suppression or excitation
;   using DIPSI2 sequence for TOCSY transfer
;   2D H-1/X correlation via double inept transfer
;   using sensitivity improvement
;   phase sensitive using Echo/Antiecho-TPPI gradient selection
;   with decoupling during acquisition
;   with multiplicity editing during selection step
;   using shaped pulses for all 180 degree pulses on f2 - channel
;
;Use WaveMaker for definition of shaped pulses
;Define offset(s) for shaped pulses with a peak list (PL)
;The peak list must be copied from another experiment with a macro (getPL or copyPL)
;
;Band-selective suppression is default
;For band-selective excitation: Use ZGOPTNS = -DEXCIT
;
;Modified 19/08/2020
;
;           Elin Alexandersson, Corine Sandström, Lena Lundqvist and Gustav Nestor
;           Swedish University of Agricultural Sciences
;
;Avance III version
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
"d2=cnst18*p1/4"
"d4=1s/(cnst2*4)"
"d11=30m"
"d12=20u"

"spoff29=0"

"d0=3u"

"in0=inf1/2"

"DELTA=d21-cnst17*p24/2-p16-d16-p2-d0*2"
"DELTA1=p16+d16-p1*0.78+de+8u"
"DELTA2=d4-larger(p2,p14)/2"
"DELTA3=d21-cnst17*p24/2-4u"
"DELTA4=d24-cnst17*p24/2-4u"

"FACTOR1=(d9/(p6*115.112))/2"
"l1=FACTOR1*2"

"acqt0=0"
baseopt_echo


1 ze
  d2
```

```
  d11 pl12:f2
2 d1 do:f2

;dpfgse element

  d12 pl1:f1
  50u UNBLKGRAD
  (p1 ph1):f1
  3u
  p16:gp1
  d16 pl0:f1
  p12:sp2:f1 ph2:r

#   ifdef EXCIT
  3u
  d12
#   else
  3u
  d12 pl1:f1
  p2 ph3
  3u
#   endif /*EXCIT*/

  p16:gp1
  d16
  3u
  p16:gp2
  d16 pl0:f1
  p12:sp2:f1 ph4:r

#   ifdef EXCIT
  3u
  d12 pl1:f1
#   else
  3u
  d12 pl1:f1
  p2 ph5
  3u
#   endif /*EXCIT*/

  p16:gp2
  d16
  p1 ph1
  3u
  3u pl0:f1
  10u gron0
  (p32:sp29 ph1):f1
  20u groff
  d16 pl10:f1

;DIPSI2 element

3 p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
```

```
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph25
  p6*4.556 ph23
  p6*3.222 ph25
  p6*3.167 ph23
  p6*0.333 ph25
  p6*2.722 ph23
  p6*4.167 ph25
  p6*2.944 ph23
  p6*4.111 ph25

  p6*3.556 ph23
  p6*4.556 ph25
  p6*3.222 ph23
  p6*3.167 ph25
  p6*0.333 ph23
  p6*2.722 ph25
  p6*4.167 ph23
  p6*2.944 ph25
  p6*4.111 ph23
  lo to 3 times l1

  p16:gp3
  d16 pl0:f1
  10u gron0*1.333
  (p32*0.75:sp29 ph1):f1
  20u groff
  d16 pl1:f1
  p1 ph1

;hsqc element

4 DELTA2 pl3:f2
  (center (p2 ph1) (p14:sp3 ph8):f2 )
  4u
  DELTA2 pl2:f2
  (p1 ph6) (p3 ph9):f2
  d0
  (p2 ph7)
  d0
  p16:gp4*EA
  d16
  DELTA pl3:f2
  (center (p2 ph1) (p24:sp7 ph10):f2 )
  4u
  DELTA3 pl2:f2
  (center (p1 ph1) (p3 ph10):f2 )
  4u
  DELTA4 pl3:f2
  (center (p2 ph1) (p24:sp7 ph1):f2 )
  4u
  DELTA4 pl2:f2
  (center (p1 ph6) (p3 ph11):f2 )
  DELTA2 pl3:f2
  (center (p2 ph1) (p14:sp3 ph1):f2 )
  DELTA2
  (p1 ph1)
  DELTA1
  (p2 ph1)
  4u
  p16:gp5
  d16 pl12:f2
  4u BLKGRAD
```

```
  go=2 ph31 cpd2:f2
  d1 do:f2 mc #0 to 2
    F1EA(calgrad(EA) & calph(ph11, +180), caldel(d0, +in0) & calph(ph9, +180) & calph(ph8, +180) & calph(ph31, +180))
exit


ph1=0
ph2=0 1 2 3
ph3=2 3 0 1
ph4=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph5=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph6=1
ph7=0 0 2 2
ph8=0
ph9=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph10=0 0 2 2
ph11=1 1 3 3
ph23=1
ph25=3
ph31=2 0 0 2 0 2 2 0 0 2 2 0 2 0 0 2


;pl0 : f1 channel - 0W
;pl1 : f1 channel - power level for pulse (default)
;pl2 : f2 channel - power level for pulse (default)
;pl3 : f2 channel - 0W
;pl10: f1 channel - power level for TOCSY-spinlock
;pl12: f2 channel - power level for CPD/BB decoupling
;sp2:wvm: userA1(cnst18 ppm, PL; PA=1.0) np=1000
;sp3: f2 channel - shaped pulse (180 degree inversion)
;spnam3: Crp60,0.5,20.1
;sp7: f2 channel - shaped pulse (180 degree refocussing)
;spnam7: Crp60comp.4
;sp29: f1 channel - shaped pulse (adiabatic)
;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel -  90 degree high power pulse
;p6 : f1 channel -  90 degree low power pulse
;p12: f1 channel - 180 degree shaped pulse
;p14: f2 channel - 180 degree shaped pulse for inversion
;     = 500 usec for Crp60,0.5,20.1
;p16: homospoil/gradient pulse
;p24: f2 channel - 180 degree shaped pulse for refocussing
;     = 2 msec for Crp60comp.4
;p32: f1 channel - 180 degree shaped pulse (adiabatic)    [20 msec]
;     smoothed chirp (sweepwidth, 20% smoothing, 10000 points)
;d0 : incremented delay (2D)                 [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d9 : TOCSY mixing time
;d11: delay for disk I/O                             [30 msec]
;d12: delay for power switching                      [20 usec]
;d16: delay for homospoil/gradient recovery
;d21: set d21 according to multiplicity selection
;       1/(2J(XH))  XH, XH3 positive, XH2 negative
;d24: 1/(8J)XH for all multiplicities
;     1/(4J)XH for XH
;cnst2: = J(XH)
;cnst17: = -0.5 for Crp60comp.4
;inf1: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(X)) = DW(X)
;l1: loop for DIPSI cycle: ((p6*115.112) * l1) = mixing time
;nd0: 2
;NS: 16 * n
;DS: >= 16
;td1: number of experiments
;FnMODE: echo-antiecho
```

```
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

;cnst18: effective bandwidth for shaped pulse (ppm)
;PL: peak list to define offset(s) for shaped pulse
;userA1: shape function (sp2)

;phcor 2 : phase difference between power levels sp2 and pl1
;phcor 4 : phase difference between power levels sp2 and pl1

;choose p12 and sp2 according to desired selectivity or use WaveMaker


;for z-only gradients:
;gpz0: ca. 5%
;gpz1: 31%
;gpz2: 11%
;gpz3: 7%
;gpz4: 80%
;gpz5: 20.1% for C-13, 8.1% for N-15

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100
;gpnam5: SMSQ10.100


;cnst17: Factor to compensate for coupling evolution during a pulse
;       (usually +1). A positive factor indicates that coupling
;       evolution continues during the pulse, whereas a negative
;       factor is necessary if the coupling is (partially) refocussed.

                                ;preprocessor-flags-start
;EXCIT: for band-selective excitation start experiment with
;           option -DEXCIT (eda: ZGOPTNS)
                                ;preprocessor-flags-end
```

II

Carbohydrate Research 511 (2022) 108477

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Complete ^{1}H and ^{13}C NMR spectral assignment of D-glucofuranose

Elin Alexandersson *, Gustav Nestor *

Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

ARTICLE INFO

Keywords:
Glucose
Furanose
Glucofuranose
NMR spectroscopy
Selective NMR experiment
Coupling constants

ABSTRACT

Although D-glucose is the most common sugar in nature, only a few NMR studies have focused on its minor furanose forms, and they have been limited to the anomeric position. Here, complete ^{1}H and ^{13}C NMR spectral analysis of α- and β-D-glucofuranose was performed, including signal assignment, chemical shifts, and coupling constants. Selective and non-selective 1D and 2D NMR experiments were used for the analysis, complemented by spin simulations and iterative spectral analysis.

1. Introduction

D-glucose is a ubiquitous metabolite in biological systems, and is thereby commonly encountered and studied in fields such as carbohydrate chemistry, metabolomics, and food science. Like many other sugars, glucose exists in an equilibrium between different ring forms and open chain tautomers in aqueous solution (Fig. 1). Of these structures, the α- and β-pyranose forms together account for more than 99% of all glucose molecules at ambient temperature, whereas the two furanose forms constitute around 0.3–0.4% and the acyclic aldehyde and hydrate forms around 0.005% each [1,2]. Despite their low abundance, the furanose forms are responsible for about half of the total reactivity of D-glucose, at least at high temperature, due to their fast ring opening rate compared to the pyranose forms [3]. However, in contrast to galactofuranoses, which are frequently found in plant and bacterial polysaccharides [4,5], glucofuranoses are extremely rare as building blocks in biomolecules. A few studies have reported glucofuranose residues from e.g. a bacterial lipopolysaccharide [6] (although this structure might need revision [7]), a plant polysaccharide [8], and as a C-glycosyl derivative from *Aloe barbadensis* [9], but these and other studies with glucofuranose assignments appear to be more or less tentative and so it is still unclear whether glucofuranose is at all present as a component of biomolecules. Furthermore, galactofuranoses are biosynthesized by UDP-galactopyranose mutase [4], but there are no known enzymes that can produce glucofuranoses for incorporation into biomolecules.

The pyranose forms of D-glucose have been extensively studied by NMR spectroscopy [10–14], but the minor forms are less well characterized. The existing studies on the furanose forms have mainly been performed using ^{13}C NMR, often with ^{13}C-labelled glucose, and have focused on the anomeric signals [1,2,15]. To the best of our knowledge, the only ^{1}H NMR data published on the glucofuranoses are the values reported for the anomeric protons [3], meaning that no full NMR characterization of these glucose forms has been carried out. This is most likely due to their low abundance and the fact that many of their resonances are obscured by pyranose signals, especially in ^{1}H NMR.

In the NMR spectra of dilute glucose solutions, the glucofuranose signals are typically below the limit of detection due to their low intensity compared to the pyranose signals. However, if a sample contains large amounts of glucose and small amounts of other compounds, the concentration of the glucofuranose forms may be comparable to or even higher than that of certain compounds of interest. Such samples may be encountered in e.g. food science, in particular food or beverage quality control, and knowledge about the glucofuranose NMR signals may facilitate the study of low-abundant compounds in these cases. Furthermore, if one is unaware of the glucofuranose NMR chemical shifts and coupling patterns, these signals may erroneously be taken for impurities.

In this work, all ^{1}H and ^{13}C NMR signals of α- and β-D-glucofuranose in D_2O are characterized for the first time. Signal assignment, chemical shifts, and homo- and heteronuclear coupling constants are reported. The results were obtained using a combination of selective and non-selective 1D and 2D NMR experiments, as well as spin simulations and iterative spectral analysis.

* Corresponding authors.
E-mail addresses: elin.alexandersson@slu.se (E. Alexandersson), gustav.nestor@slu.se (G. Nestor).

https://doi.org/10.1016/j.carres.2021.108477
Received 11 October 2021; Received in revised form 5 November 2021; Accepted 8 November 2021
Available online 9 November 2021

2. Results and discussion

2.1. Assignment of the D-glucofuranose NMR signals

When inspecting the ^{1}H NMR spectrum of D-glucose, minor signals can be observed between 4.07 ppm and 4.32 ppm, i.e. in the area between the signals from the pyranose ring protons and anomeric protons, as well as at 5.49 ppm (Fig. 2a). Additional low-intensity signals, which overlap with glucopyranose signals in the ^{1}H but not the ^{13}C dimension, are visible in the corresponding ^{1}H,^{13}C-HSQC spectrum (Fig. 2b). All of the signals have approximately the same intensity, implying that they either belong to the same species or to different species with comparable concentration. Analysis of [1–^{13}C]-glucose revealed that the two most downfield of the minor HSQC signals are the H1/C1 atoms of two glucose tautomers (Figure S1) and based on the intensity of the signals relative to the pyranose signals, they most likely belong to the two anomeric forms of glucofuranose. The chemical shifts of the signals also agree well with previously reported NMR data for the glucofuranose anomeric carbons [1,2,15] and protons [3].

To further investigate the observed signals, selective ^{1}H NMR experiments were performed. Exciting the proton signals at 5.49 ppm, 4.31 ppm, and 4.24 ppm in separate 1D-TOCSY experiments revealed two different spin systems, each containing seven signals (Fig. 3). The signal at 5.49 ppm has previously been assigned to the anomeric proton of α-glucofuranose [3] and has a coupling constant of 3.96 Hz, which is similar to values observed for the anomeric signal of other furanoid 1, 2-*cis* tautomers [16–19]. The other anomeric furanose signal overlaps with the anomeric α-glucopyranose signal in the ^{1}H NMR spectrum, but is clearly observed in the selective 1D-TOSCY spectrum at 5.21 ppm (Fig. 3C). No splitting can be observed for this signal and a small coupling constant is expected for the anomeric signal of a furanoid 1, 2-*trans* tautomer such as β-glucofuranose [3,16–19]. Adding this to the indications described earlier meant that the NMR signals at 5.49 and 5.21 ppm could be assigned to α- and β-glucofuranose, respectively. The remaining furanose signals were then assigned based on series of selective 1D-TOCSY experiments where the mixing time was varied between 20 and 120 ms (Fig. 3 and S2-S4).

The ^{13}C resonances were assigned from HSQC spectra based on the ^{1}H assignments determined above. To observe only the glucofuranoses and to avoid spectral overlap with the pyranose resonances, an f2-band-selective TOCSY-HSQC pulse sequence [20] was employed. In this experiment, a certain part of the ^{1}H spectrum is excited after which a TOCSY spin-lock restores all signals that are *J*-coupled to any of the excited spins. Connecting this pulse sequence to an HSQC sequence thus enables selected spin systems to be visualized in both the ^{1}H and the ^{13}C dimension at the same time. Here, the ^{1}H spectral region 4.10–4.40 ppm was selectively excited to observe the two glucofuranose spin systems without interference from the pyranose forms (Fig. 4).

2.2. Chemical shifts and coupling constants of α- and β-D-glucofuranose

The glucofuranose ^{13}C NMR chemical shifts are listed in Table 1, together with previously reported ^{13}C shifts. The chemical shifts were collected from a 1D-^{13}C spectrum, apart from α-furanose C2 and β-furanose C5 that were read from an HSQC spectrum due to spectral overlap in the 1D spectrum. To facilitate comparison with previous studies [2,15] the chemical shifts have been referenced to both DSS and TMS, since the ^{13}C shifts of these reference compounds differ by about 2 ppm [21]. The ^{13}C chemical shifts agree well with the previously reported data, although the earlier assignments of β-glucofuranose C2 and C4 [15] need to be interchanged. The ^{13}C chemical shifts are also in close agreement with those of the corresponding methyl glucofuranosides [22], apart from the C1 resonances that are affected by the methyl group in the methyl furanosides (Table 1). One-bond ^{13}C–^{1}H coupling constants, also listed in Table 1, were determined from an f2-band-selective TOCSY-HSQC spectrum recorded without carbon decoupling. A previous study on methyl furanosides found that the $J_{C1,H1}$ is 174.0 Hz for methyl α-glucofuranoside and 172.5 Hz for methyl

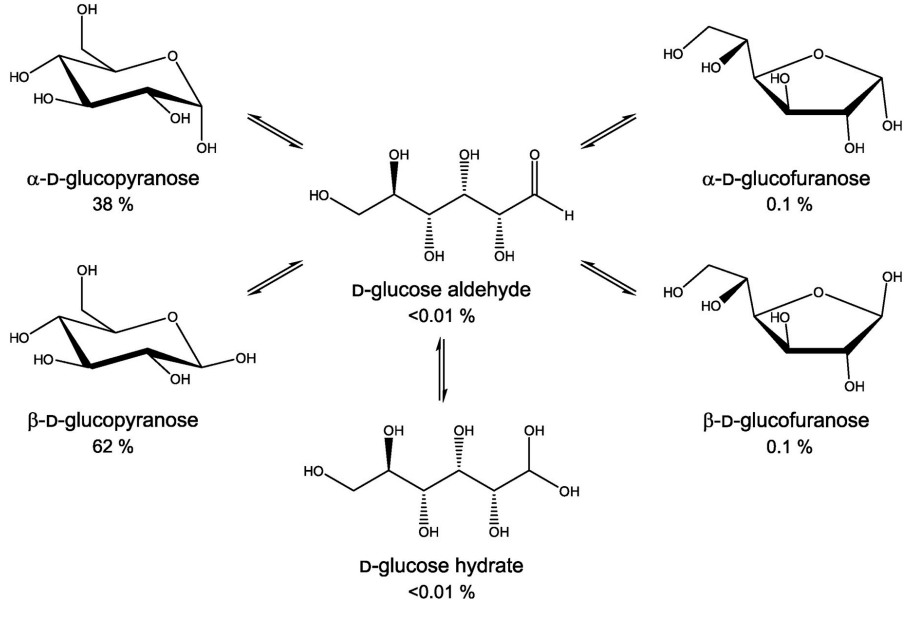


Fig. 1. The anomeric equilibrium of D-glucose in aqueous solution, with relative abundances at 25 °C and pD 7.

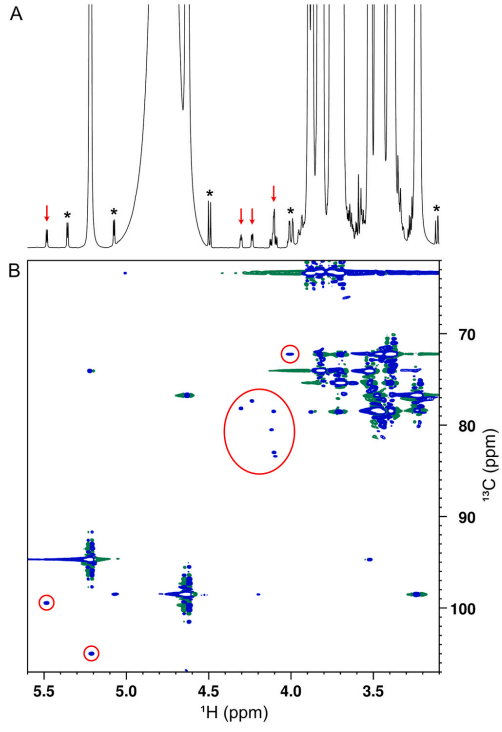


Fig. 2. A) 1D ^{1}H NMR spectrum of D-glucose in D_2O at 25 °C. The arrows denote minor signals not corresponding to ^{13}C satellites of the glucopyranose signals (indicated with asterisks). B) ^{1}H,^{13}C-HSQC spectrum of D-glucose in D_2O (25 °C), with the minor signals encircled.

β-glucofuranoside [23], which is very close to what was found here for the free furanoses.

Spin simulations and iterative spectral analysis were performed to verify the glucofuranose ^{1}H chemical shifts and to determine homonuclear two- and three-bond ^{1}H coupling constants (Table 2 and Fig. 5). Non-selective 1D-^{1}H spectra were used as input for the simulations when possible, however the signals of α- and β-furanose H5, H6a, and H6b, as well as β-furanose H1, are covered by pyranose signals and therefore selective 1D-^{1}H spectra were used to analyse these signals (see Figure S5). Because selective spectra can suffer from distortions in signal shape and intensity, zero-quantum coherence suppression was used to minimize anti-phase components; however the accuracy of the extracted coupling constants may still be compromised. Furthermore, the $J_{H1,H2}$ and $J_{H2,H3}$ of the β anomer are both around 1 Hz in size and difficult to determine with high accuracy because of broad, overlapping signals. Long-range ^{1}H–^{1}H coupling constants have earlier been observed for the pyranose form of α-D-glucose [12] and are likely to be present in the furanose forms as well. However, no long-range couplings were resolved

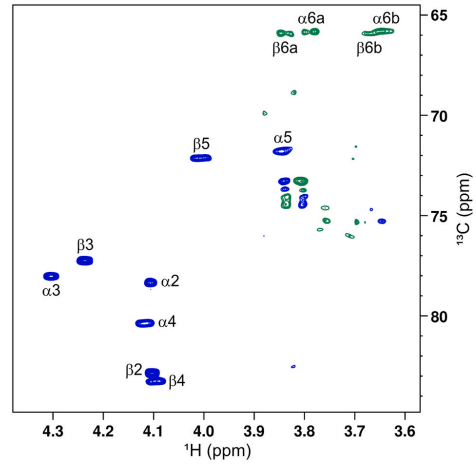


Fig. 4. Multiplicity-edited f2-band-selective ^{1}H,^{13}C-HSQC spectrum of D-glucose in D_2O (25 °C) with assignment of the glucofuranose ring signals. Non-assigned cross-peaks are mainly from residual pyranose signals.

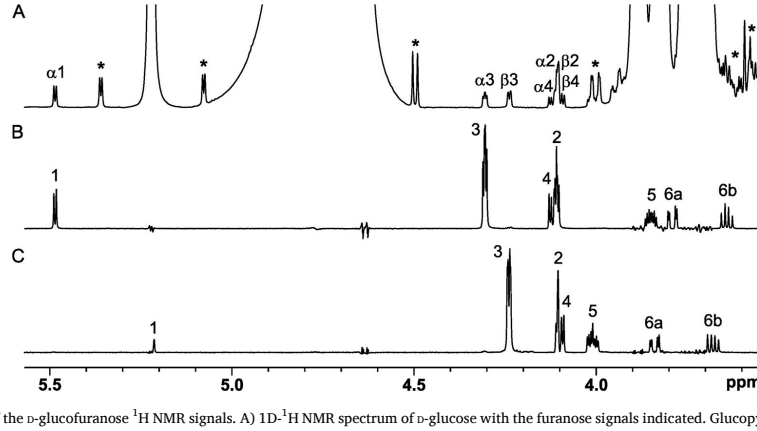


Fig. 3. Assignment of the D-glucofuranose ^{1}H NMR signals. A) 1D-^{1}H NMR spectrum of D-glucose with the furanose signals indicated. Glucopyranose ^{13}C satellites are marked with asterisks. B) Selective 1D-TOCSY spectrum of α-glucofuranose obtained by exciting the H3 signal at 4.30 ppm. C) Selective 1D-TOCSY spectrum of β-glucofuranose obtained by exciting the H3 signal at 4.24 ppm. The mixing time was 100 ms in both B) and C).

Table 1
^{13}C NMR chemical shifts (δ, ppm) and $^{1}J_{C,H}$ coupling constants (Hz) of α- and β-D-glucofuranose in D_2O (25 °C, pD 7.0). Results from previous studies (in parentheses) are included as reference.

		C1	C2	C3	C4	C5	C6
α	δ_{DSS}	99.43	78.49[a]	78.16	80.51	71.92	65.95
	δ_{TMS}	97.64	76.69[a]	76.36	78.71	70.12	64.15
	$^{1}J_{C,H}$[b]	172.6	152.6	153.1	146.3	n.d.[c]	143.3 (6a) 142.7 (6b)
	δ_{TMS}[d] [2]	(97.6)	-	-	-	-	-
	δ_{TMS}[e] [22]	(104.0)	(77.7)	(76.6)	(78.8)	(70.7)	(64.2)
β	δ_{DSS}	104.98	82.98	77.38	83.40	72.27[a]	66.08
	δ_{TMS}	103.18	81.18	75.58	81.60	70.47[a]	64.28
	$^{1}J_{C,H}$[b]	172.7	154.5	153.1	147.8	146.8	143.3 (6a) 142.7 (6b)
	δ_{TMS}[f] [15]	(103.8)	(82.1)	-	(81.8)	-	-
	δ_{TMS}[d] [2]	(103.2)	-	-	-	-	-
	δ_{TMS}[e] [22]	(110.0)	(80.6)	(75.8)	(82.3)	(70.7)	(64.7)

[a] The chemical shift was determined from an HSQC spectrum, with an uncertainty of around 0.01 ppm.
[b] The standard error was estimated to 0.1 Hz, except for the C6/H6 couplings where the error was estimated to 0.5 Hz.
[c] Not determined.
[d] Externally referenced to the chemical shift of C1 in α-D-[1–^{13}C]mannopyranose (95.0 ppm). Recorded at 30 °C.
[e] Methyl glucofuranoside.
[f] Recorded at 41 °C.

Table 2
^{1}H NMR chemical shifts (δ, ppm) and $J_{H,H}$ coupling constants (Hz) of α- and β-D-glucofuranose in D_2O (25 °C, pD 7.0).

		H1	H2	H3	H4	H5	H6a	H6b
α	δ_{DSS}	5.486	4.108	4.306	4.119	3.850	3.790	3.642
	$J_{H,H}$[a]	$^{3}J_{1,2}$ = 3.96	$^{3}J_{2,3}$ = 2.42	$^{3}J_{3,4}$ = 3.85	$^{3}J_{4,5}$ = 8.50	$^{3}J_{5,6a}$ = 2.87 $^{3}J_{5,6b}$ = 6.22	$^{2}J_{6a,6b}$ = 12.00	= (−)
β	δ_{DSS}	5.213	4.104	4.238	4.098	4.009	3.838	3.679
	$J_{H,H}$[a]	$^{3}J_{1,2}$ < 1	$^{3}J_{2,3}$ ≈ 1.2	$^{3}J_{3,4}$ = 4.19	$^{3}J_{4,5}$ = 9.00	$^{3}J_{5,6a}$ = 2.75 $^{3}J_{5,6b}$ = 6.05	$^{2}J_{6a,6b}$ = 11.99	= (−)

[a] Standard errors are 0.05 Hz or less, except for α-furanose $^{3}J_{5,6a}$ that has an uncertainty of 0.1 Hz.

in the spectra, indicating that they are probably small in size. Comparing the coupling constants with previous data shows good agreement; Kaufmann et al. [3] reported that α- and β-glucofuranose $J_{H1,H2}$ are 3.9 Hz and "less than 1 Hz", respectively, whereas the corresponding methyl glucofuranoside coupling constants are 4.2 Hz and 1.0 Hz, respectively [19]. Here, 3.96 Hz and <1 Hz were found for the α- and β-glucofuranose $J_{H1,H2}$, respectively.

The ring-opening rate of glucofuranose is about 1 s^{-1} at 87 °C [3], which is much faster than that of the pyranose forms (0.004–0.008 s^{-1} at 87 °C [3] or 0.001–0.002 s^{-1} at 30 °C [24]). Chemical exchange can affect NMR spectra so that signals are broadened or observed as an average between the two exchanging species, but in the case of glucofuranose at room temperature no such signal broadening was observed. Hence, the ring-opening of the glucofuranose ring is slow on the NMR timescale and the reported chemical shifts are thus not considered to be affected by chemical exchange.

2.3. Relative quantification of the D-glucopyranose and furanose anomers

To determine the relative amounts of the different D-glucose anomers at 25 °C and pD 7.0, a 1D-^{1}H NMR spectrum was recorded with a long relaxation delay (>5 × T_1). The α-pyranose was found to account for 37.5% and the β-pyranose for 62.2% of the total glucose, which is within the previously reported α/β ratio [1,2,12,13,15], whereas α-furanose made up 0.12% and β-furanose 0.13%. A similar result was obtained using quantitative ^{13}C NMR on [1–^{13}C]-glucose: 37.5% α-pyranose, 62.3% β-pyranose, and 0.11–0.12% of each furanose form. It should be noted that the anomeric signals were used for quantification in ^{13}C NMR, but other signals had to be used in ^{1}H NMR since the α-pyranose and β-furanose anomeric signals overlap with each other and the anomeric β-pyranose signal overlaps with the residual water signal (see section 4.2 for details). For quantitative purposes, ^{1}H NMR is the preferred choice in order to minimize the effects of the longer relaxation times of ^{13}C that require long recovery delays. Furthermore, the signal to noise ratio is often better in ^{1}H NMR spectra, making the integration more reliable especially when dealing with low-intensity signals as in this case. The results from previous studies are somewhat disparate. One study found that the α- and β-furanose, at 27 °C and pH 4.7, constitute 0.14% and 0.15%, respectively [1], which is very similar to the results obtained here. Another study, performed at 30 °C, found that the amount of α-furanose is significantly lower than that of β-furanose: 0.108% and 0.28%, respectively [2]. Both these studies quantified the anomers using ^{13}C NMR on [1–^{13}C]-glucose. It is known that the anomeric equilibrium of sugars is affected by factors such as sugar concentration [13,15],

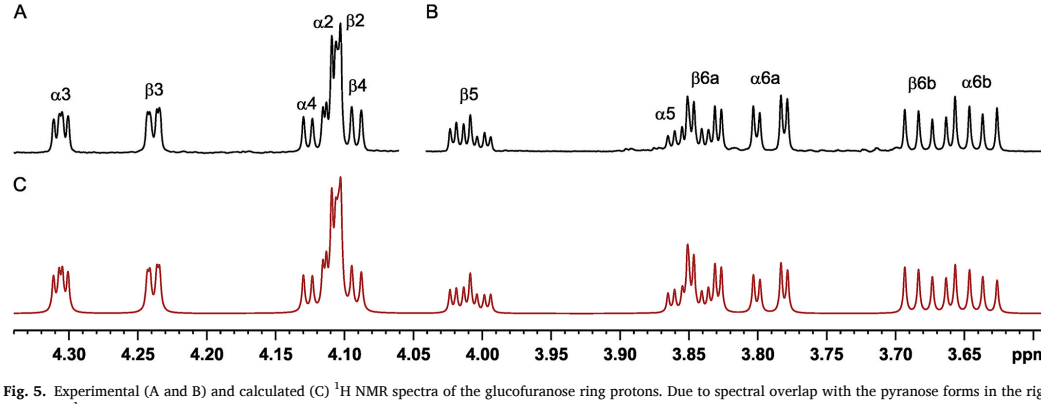


Fig. 5. Experimental (A and B) and calculated (C) ^{1}H NMR spectra of the glucofuranose ring protons. Due to spectral overlap with the pyranose forms in the right part of the ^{1}H NMR spectrum, A) is a standard 1D spectrum whereas B) is a 1D-TOCSY spectrum obtained by band-selective excitation of the spectral region 4.10-4.40 ppm. See also Figure S5-S7.

temperature [1], pH, and buffer concentration [25,26], which may explain some of the discrepancy. Here, the glucose concentration was 1 M and a dilute phosphate buffer was used to keep the pD at 7.0.

3. Conclusions

In this study, the complete ^{1}H and ^{13}C NMR chemical shifts of both D-glucofuranose anomers are reported for the first time. This new knowledge can be utilized e.g. when studying glucofuranose conformation and glucose ring-chain tautomerism. Furthermore, although the furanoses constitute less than 1% of the total glucose, knowing their chemical shifts can be essential for studies of low-abundant compounds in solutions containing high amounts of glucose, where some of the furanose signals may interfere with the signals of interest. This study also highlights the usefulness of selective and band-selective NMR experiments for characterization of low-abundant compounds when other compounds in high concentration cause spectral interference.

4. Experimental

4.1. General methods

Anhydrous D(+)-glucose was purchased from VWR. [1−^{13}C]-glucose was purchased from Cambridge Isotope Laboratories, Inc. Glucose solutions (1 M) were prepared using 20 mM KH_2PO_4 buffer in D_2O, pD 7.0 (apparent pH 6.6), as solvent. DSS-d_6 (sodium 3-(trimethylsilyl)propane-1-sulfonate-d_6) was added as a chemical shift reference in a concentration of approximately 5 mM. To some of the samples, TMS (tetramethylsilane) was added as well. The solutions were equilibrated in room temperature for at least 24 h prior to NMR analysis. Before NMR acquisition, the pD was checked again and adjusted to 7.0 using NaOH dissolved in D_2O if necessary.

4.2. NMR spectroscopy

NMR analysis was performed on a Bruker Avance III 600 MHz spectrometer with a 5 mm ^{1}H/^{13}C/^{15}N/^{31}P inverse detection cryoprobe or a 5 mm broadband observe detection SmartProbe, both equipped with a z gradient. Spectra were recorded at 25 °C and were processed with TopSpin 4.0.6. The spectrometer temperature was calibrated in connection to the experiments using 4% methanol in methanol-d_4. The carrier frequency was placed on the HDO signal (4.70 ppm). SUN (Suppression of UNwanted signals) pulse sequences [20] (1D-^{1}H, ^{1}H,^{1}H-TOCSY, and ^{1}H,^{13}C-HSQC) were used for selective excitation experiments. As the 180° selective excitation pulse, either an IBURP-2 (25.8 ms) or a Gaussian pulse (80 ms) was used depending on the width of the selected area; when only one spin was excited the latter pulse was employed due to its narrower excitation profile. The TOCSY spin-lock was 100 ms if not stated otherwise. For quantitative ^{1}H NMR, a relaxation delay of 20 s (including the acquisition time) was used and the following signals were integrated: α-pyranose H1 or H2, β-pyranose H2 or H6a, α-furanose H1 or H3, and β-furanose H3. For quantitative 1D ^{13}C NMR, the inverse-gated decoupled Bruker pulse sequence *zgig30* was used, with a relaxation delay of 40 s. Before quantitation, spectral baseline points were defined after which the baseline was corrected using cubic spline. The number of data points collected in HSQC was 1024 or 2048, apart from the HSQC without decoupling where 8192 data points were collected.

4.3. Spin simulations

Spin simulations and iterative spectral analysis were performed using the Daisy module in TopSpin. Proton chemical shifts and coupling constants that could be extracted directly from the spectra (both selective and non-selective) were used as starting values for the simulations. Thereafter, the *J*-value, chemical shift, and line width for each spin were iterated until the best possible fit was achieved, judged by visual evaluation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carres.2021.108477.

References

[1] S.R. Maple, A. Allerhand, Detailed tautomeric equilibrium of aqueous d-glucose. Observation of six tautomers by ultrahigh resolution carbon-13 NMR, J. Am. Chem. Soc. 109 (1987) 3168–3169, https://doi.org/10.1021/ja00244a063.

[2] Y. Zhu, J. Zajicek, A.S. Serianni, Acyclic forms of [1-^{13}C]aldohexoses in aqueous solution: quantitation by ^{13}C NMR and deuterium isotope effects on tautomeric equilibria, J. Org. Chem. 66 (2001) 6244–6251, https://doi.org/10.1021/jo010541m.

[3] M. Kaufmann, C. Mügge, L.W. Kroh, NMR analyses of complex d-glucose anomerization, Food Chem. 265 (2018) 222–226, https://doi.org/10.1016/j.foodchem.2018.05.100.

[4] M.R. Richards, T.L. Lowary, Chemistry and biology of galactofuranose-containing polysaccharides, Chembiochem 10 (2009) 1920–1938, https://doi.org/10.1002/cbic.200900208.

[5] B. Tefsen, A.F.J. Ram, I. van Die, F.H. Routier, Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact, Glycobiology 22 (2012) 456–469, https://doi.org/10.1093/glycob/cwr144.

[6] T.C. Ray, A.R.W. Smith, R. Wait, R.C. Hignett, Structure of the sidechain of lipopolysaccharide from *Erwinia amylovora* T, Eur. J. Biochem. 170 (1987) 357–361, https://doi.org/10.1111/j.1432-1033.1987.tb13707.x.

[7] Y.A. Knirel, Structure of O-antigens, in: Y.A. Knirel, M.A. Valvano (Eds.), Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells, Springer, Vienna, 2011, pp. 41–115.

[8] X. Wang, R. Sun, J. Zhang, Y. Chen, N. Liu, Structure and antioxidant activity of polysaccharide POJ-U1a extracted by ultrasound from *Ophiopogon japonicus*, Fitoterapia 83 (2012) 1576–1584, https://doi.org/10.1016/j.fitote.2012.09.005.

[9] M.K. Park, J.H. Park, Y.G. Shin, W.Y. Kim, J.H. Lee, K.H. Kim, A. Neoaloesin, A new *C*-glucofuranosyl chromone from *Aloe barbadensis*, Planta Med. 62 (1996) 363–365, https://doi.org/10.1055/s-2006-957907.

[10] T.E. Walker, R.E. London, T.W. Whaley, R. Barker, N.A. Matwiyoff, Carbon-13 nuclear magnetic resonance spectroscopy of [1-^{13}C] enriched monosaccharides. Signal assignments and orientational dependence of geminal and vicinal carbon-carbon and carbon-hydrogen spin-spin coupling constants, J. Am. Chem. Soc. 98 (1976) 5807–5813, https://doi.org/10.1021/ja00435a011.

[11] S.J. Perkins, L.N. Johnson, D.C. Phillips, R.A. Dwek, High resolution ^{1}H- and ^{13}C-N.M.R spectra of d-glucopyranose, 2-acetamido-2-deoxy-d-glucopyranose, and related compounds in aqueous media, Carbohydr. Res. 59 (1977) 19–34, https://doi.org/10.1016/S0008-6215(00)83289-9.

[12] M.U. Roslund, P. Tähtinen, M. Niemitz, R. Sjöholm, Complete assignments of the ^{1}H and ^{13}C chemical shifts and $J_{H,H}$ coupling constants in NMR spectra of d-glucopyranose and all d-glucopyranosyl-d-glucopyranosides, Carbohydr. Res. 343 (2008) 101–112, https://doi.org/10.1016/j.carres.2007.10.008.

[13] M. Maebayashi, M. Ohba, T. Takeuchi, Anomeric proportions of d-glucopyranose at the equilibrium determined from ^{1}H-NMR spectra I. Investigation of experimental conditions and concentration dependence at 25.0 °C, J. Mol. Liq. 232 (2017) 408–415, https://doi.org/10.1016/j.molliq.2017.02.062.

[14] A.S. Perlin, B. Casu, H.J. Koch, Configurational and conformational influences on the carbon-13 chemical shifts of some carbohydrates, Can. J. Chem. 48 (1970) 2596–2606, https://doi.org/10.1139/v70-435.

[15] C. Williams, A. Allerhand, Detection of β-d-glucofuranose in aqueous solutions of d-glucose. Application of carbon-13 fourier-transform n.m.r. spectroscopy, Carbohydr. Res. 56 (1977) 173–179, https://doi.org/10.1016/S0008-6215(00)84250-0.

[16] S.J. Angyal, V.A. Pickles, Equilibria between pyranoses and furanoses. II. Aldoses, Aust. J. Chem. 25 (1972) 1695–1710, https://doi.org/10.1071/CH9721695.

[17] D.E. Kiely, L. Benzing-Nguyen, Oxidation of carbohydrates with chromic acid. Synthesis of 6-acetamido-6-deoxy-d-xylo-hexos-5-ulose, J. Org. Chem. 40 (1975) 2630–2634, https://doi.org/10.1021/jo00906a012.

[18] L.D. Hayward, S.J. Angyal, A symmetry rule for the circular dichroism of reducing sugars, and the proportion of carbonyl forms in aqueous solutions thereof, Carbohydr. Res. 53 (1977) 13–20, https://doi.org/10.1016/S0008-6215(00)85450-6.

[19] A. Lubineau, J.-C. Fischer, High-yielding one-step conversion of d-Glucose and d-galactose to the corresponding α and β methyl-d-glucofuranosides and galactofuranosides, Synth. Commun. 21 (1991) 815–818, https://doi.org/10.1080/00397919108019762.

[20] E. Alexandersson, C. Sandström, L.C.E. Lundqvist, G. Nestor, Band-selective NMR experiments for suppression of unwanted signals in complex mixtures, RSC Adv. 10 (2020) 32511–32515, https://doi.org/10.1039/D0RA06828D.
[21] R.K. Harris, E.D. Becker, S.M. Cabral de Menezes, R. Goodfellow, P. Granger, NMR nomenclature: nuclear spin properties and conventions for chemical shifts (IUPAC recommendations 2001), Pure Appl. Chem. 73 (2001) 1795–1818, https://doi.org/10.1351/pac200173111795.
[22] R.G.S. Ritchie, N. Cyr, B. Korsch, H.J. Koch, A.S. Perlin, Carbon-13 chemical shifts of furanosides and cyclopentanols. Configurational and conformational influences, Can. J. Chem. 53 (1975) 1424–1433, https://doi.org/10.1139/v75-197.
[23] N. Cyr, A.S. Perlin, The conformations of furanosides. A ^{13}C nuclear magnetic resonance study, Can. J. Chem. 57 (1979) 2504–2511, https://doi.org/10.1139/v79-399.
[24] B.E. Lewis, N. Choytun, V.L. Schramm, A.J. Bennet, Transition states for glucopyranose interconversion, J. Am. Chem. Soc. 128 (2006) 5049–5058, https://doi.org/10.1021/ja0573054.
[25] J.M. Los, L.B. Simpson, K. Wiesner, The kinetics of mutarotation of d-glucose with consideration of an intermediate free-aldehyde form, J. Am. Chem. Soc. 78 (1956) 1564–1568, https://doi.org/10.1021/ja01589a017.
[26] S.J. Angyal, The composition of reducing sugars in solution, in: R.S. Tipson, D. Horton (Eds.), Advances in Carbohydrate Chemistry and Biochemistry, Academic Press, 1984, pp. 15–68.

Electronic supplementary information

Complete ^{1}H and ^{13}C NMR spectral assignment of D-glucofuranose

Elin Alexandersson and Gustav Nestor

Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

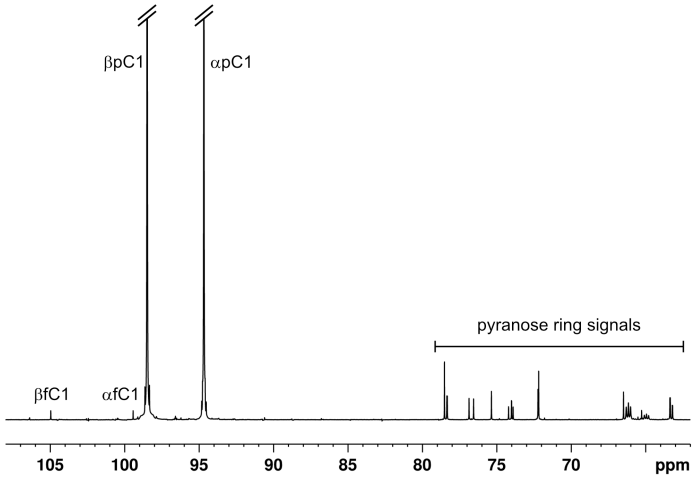


Figure S1. 1D-^{13}C NMR spectrum of [1-^{13}C]-glucose.

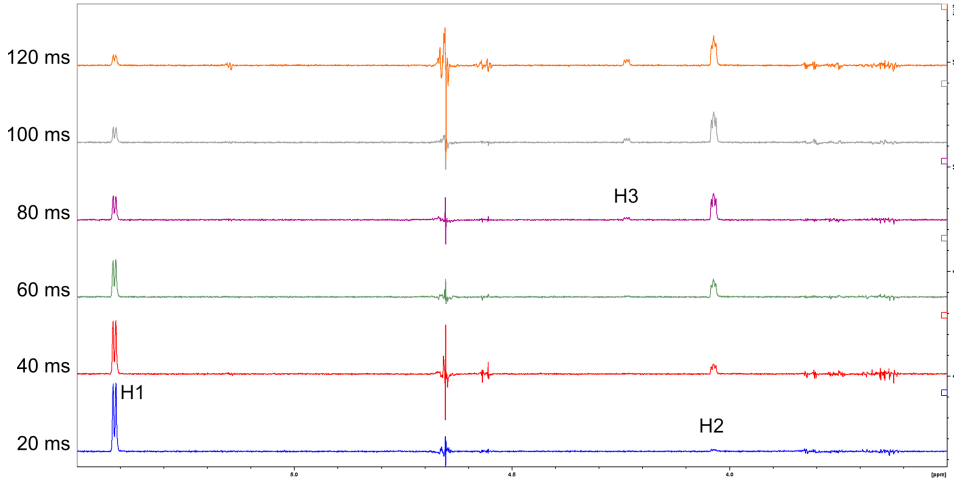


Figure S2. 1D-TOCSY with selective excitation of the anomeric signal of α-glucofuranose. The mixing time was varied between 20 and 120 ms to enable step-wise assignment of the spin system.

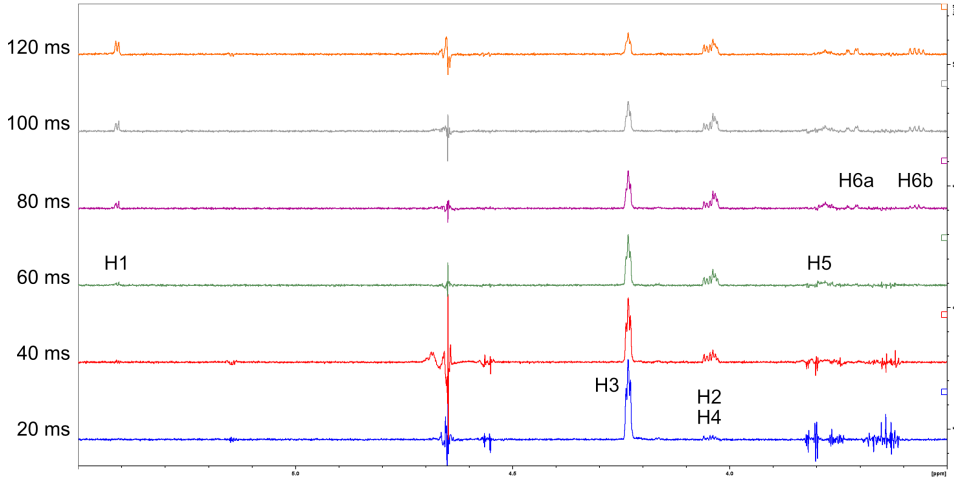


Figure S3. 1D-TOCSY with selective excitation of the α-glucofuranose H3 signal. The mixing time was varied between 20 and 120 ms to enable step-wise assignment of the spin system.

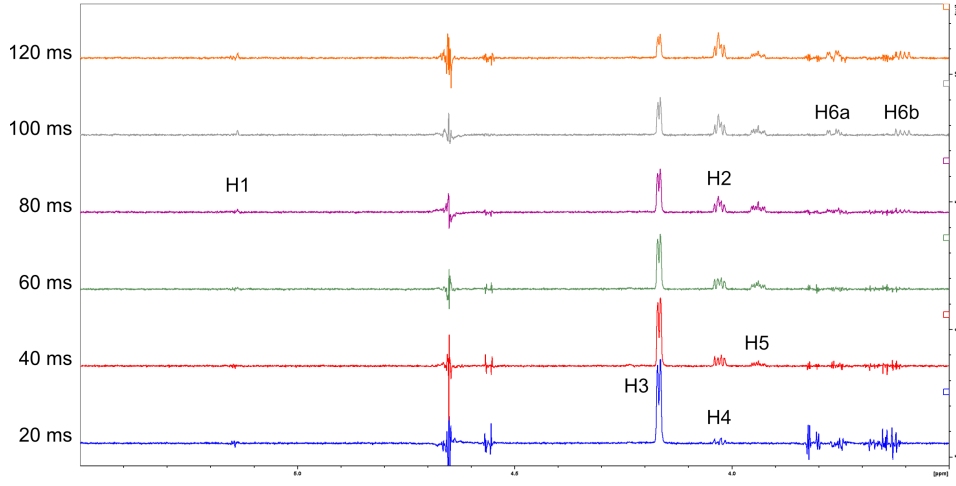


Figure S4. 1D-TOCSY with selective excitation of the β-glucofuranose H3 signal. The mixing time was varied between 20 and 120 ms to enable step-wise assignment of the spin system. H1 and H2 have a small coupling constant to H3 and therefore require a relatively long mixing time to appear in the spectrum.

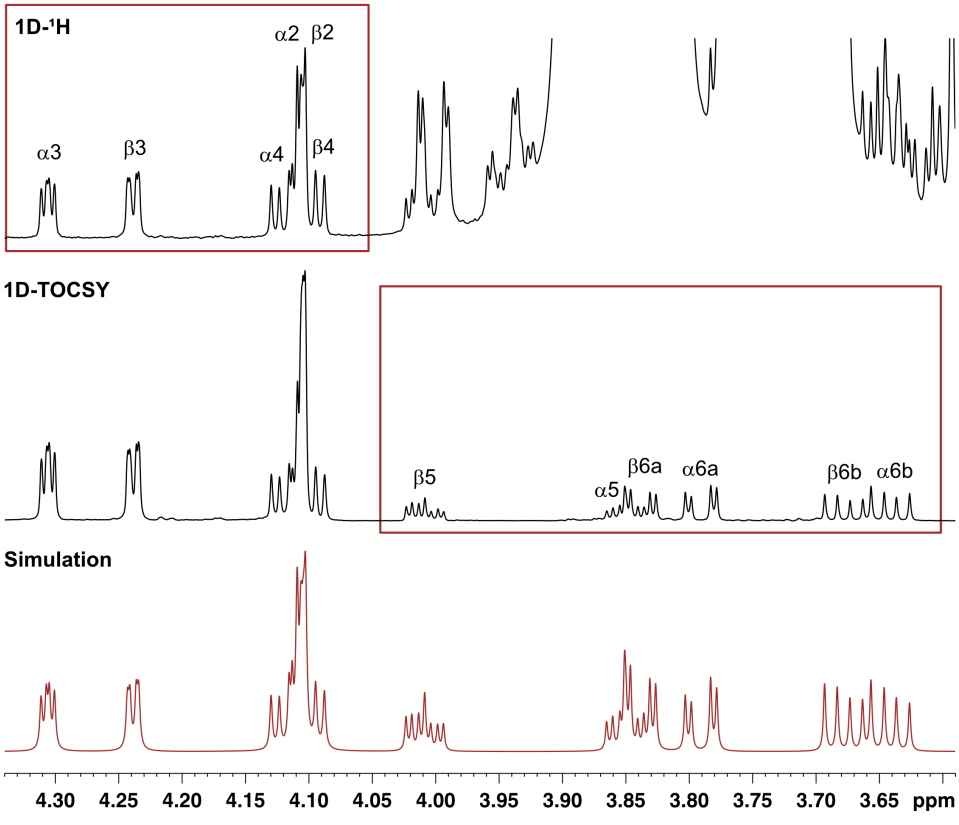


Figure S5. Experimental 1D-^{1}H spectrum (top), 1D-TOCSY spectrum (middle), and calculated spectrum (bottom) of the glucofuranose ring protons. The 1D-TOCSY spectrum was obtained by band-selective excitation of the spectral region 4.10-4.40 ppm. The region of each experimental spectrum that was used as input for the simulation is marked in red.

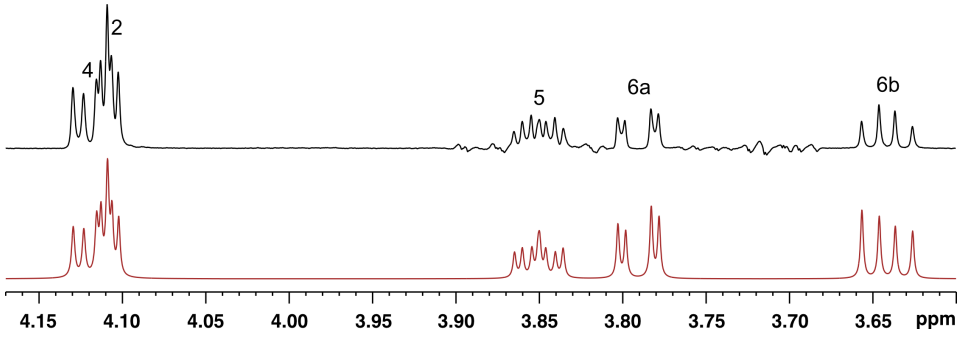


Figure S6. Experimental 1D-TOCSY spectrum (top) and spin simulation (bottom) of α-glucofuranose H2 and H4-H6. The experimental spectrum was obtained by exciting the αH3 signal at 4.30 ppm.

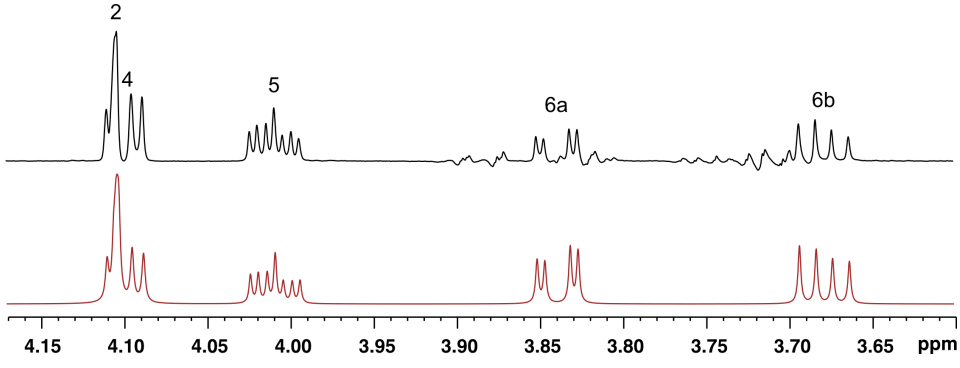


Figure S7. Experimental 1D-TOCSY spectrum (top) and spin simulation (bottom) of β-glucofuranose H2 and H4-H6. The experimental spectrum was obtained by exciting the βH3 signal at 4.24 ppm.

III

Metabolomics (2024) 20:11
https://doi.org/10.1007/s11306-023-02073-z

ORIGINAL ARTICLE

Extended automated quantification algorithm (AQuA) for targeted ^{1}H NMR metabolomics of highly complex samples: application to plant root exudates

Elin Alexandersson[1] · Corine Sandström[1] · Johan Meijer[2] · Gustav Nestor[1] · Anders Broberg[1] · Hanna E. Röhnisch[1]

Received: 22 June 2023 / Accepted: 28 November 2023

Abstract

Introduction The Automated Quantification Algorithm (AQuA) is a rapid and efficient method for targeted NMR-based metabolomics, currently optimised for blood plasma. AQuA quantifies metabolites from 1D-^{1}H NMR spectra based on the height of only one signal per metabolite, which minimises the computational time and workload of the method without compromising the quantification accuracy.

Objectives To develop a fast and computationally efficient extension of AQuA for quantification of selected metabolites in highly complex samples, with minimal prior sample preparation. In particular, the method should be capable of handling interferences caused by broad background signals.

Methods An automatic baseline correction function was combined with AQuA into an automated workflow, the extended AQuA, for quantification of metabolites in plant root exudate NMR spectra that contained broad background signals and baseline distortions. The approach was evaluated using simulations as well as a spike-in experiment in which known metabolite amounts were added to a complex sample matrix.

Results The extended AQuA enables accurate quantification of metabolites in 1D-^{1}H NMR spectra with varying complexity. The method is very fast (< 1 s per spectrum) and can be fully automated.

Conclusions The extended AQuA is an automated quantification method intended for 1D-^{1}H NMR spectra containing broad background signals and baseline distortions. Although the method was developed for plant root exudates, it should be readily applicable to any NMR spectra displaying similar issues as it is purely computational and applied to NMR spectra post-acquisition.

Keywords Targeted metabolomics · Automated quantification · Baseline correction · AQuA · Root exudate · NMR

1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is commonly used in metabolomics for identification and quantification of metabolites in different biological samples (Crook & Powers, 2020). NMR has many advantages; it is inherently quantitative, highly reproducible, non-destructive, and enables analysis of compounds with different chemical properties in one single experiment. However, the complex mixtures of natural products that are studied in metabolomics typically yield complicated 1D-^{1}H NMR spectra with extensive spectral overlap, which can make both identification and quantification of individual metabolites challenging. Spectral overlap occurs because one metabolite can generate several NMR signals, and signals from different compounds often appear at similar chemical shifts. The resulting signal interferences are especially problematic for quantitative studies because concentrations of individual metabolites will be overestimated unless the interferences are properly accounted for. Two-dimensional NMR experiments can be used to increase signal dispersion, but 2D spectra typically take longer time both to acquire and to analyse than 1D

✉ Elin Alexandersson
elin.alexandersson@slu.se

[1] Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

[2] Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden

Published online: 23 December 2023

spectra. Furthermore, quantification based on 2D spectra is not straightforward since the intensity of individual peaks is influenced by their coupling constants and transverse relaxation times. Accordingly, calibration with pure reference compounds, either externally or internally, is required for accurate quantification (Crook & Powers, 2020; Martineau et al., 2020). Therefore, 1D-^{1}H NMR experiments are still the most common in high-throughput studies and there continues to be a high demand for methods that can accurately quantify metabolites based on 1D-^{1}H spectra. Various approaches have been developed, both manual (Weljie et al., 2006) and automated (Zheng et al., 2011; Hao et al., 2012; Ravanbakhsh et al., 2015; Tardivel et al., 2017; Lefort et al., 2019; Häckl et al., 2021; Rout et al., 2023).

An Automated Quantification Algorithm (AQuA) for targeted metabolomics has previously been developed in our group (Röhnisch et al., 2018, 2021). This method quantifies metabolites from 1D-^{1}H NMR spectra using only one signal per metabolite, which reduces the computational time and workload substantially compared to e.g. curve-fitting quantification algorithms (Zheng et al., 2011; Hao et al., 2012; Ravanbakhsh et al., 2015; Tardivel et al., 2017). At the same time, AQuA corrects for signal interferences between different metabolites as well as inter-spectral variation in signal position. Currently, AQuA is optimised for ultra-filtered human plasma samples but it would be desirable to extend its use to other, more heterogeneous, sample types as well, preferably without any time-consuming sample preparation.

Whereas human blood plasma and serum are well studied by NMR and the majority of signals have been assigned (Psychogios et al., 2011; Nagana Gowda et al., 2015), many other biological samples are less well characterised. Plant samples, for example, are very complex with numerous different metabolites of widely different concentrations, which complicates NMR analysis (Deborde et al., 2017). In the present study, aqueous oilseed rape (*Brassica napus*) root exudate samples were used as a model system to develop the proposed workflow (Fig. 1). Root exudates consist of all substances that are excreted by plant roots during growth, including sugars, organic acids, and amino acids (Vives-Peris et al., 2020). In addition, the samples used in this study all contained various unknown compounds, likely lipids, that gave rise to broad signals in the spectra (Fig. 1b and c). Before accurate quantification can be performed, these signals need to be accounted for in some way.

In blood plasma and serum, macromolecules giving rise to broad signals are routinely removed by ultrafiltration or precipitation with organic solvents before NMR analysis (Daykin et al., 2002; Nagana Gowda & Raftery, 2014). Other options are to use certain NMR experiments that target broad signals, such as the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (Carr & Purcell, 1954; Meiboom & Gill, 1958) or diffusion-edited experiments (Liu et al., 1996; de Graaf & Behar, 2003; Bliziotis et al., 2020). There are also methods solely based on computations, such as the Small Molecule Enhancement Spectroscopy (SMolESY) method (Takis et al., 2020, 2021) that utilises the first derivative of the imaginary part of the NMR data to generate a spectrum devoid of broad signals. SMolESY is capable of performing automated relative quantification in blood samples, but for more complex spectra remaining metabolite signal interferences may appear. Because the NMR signals in a SMolESY spectrum are not Lorentzian shaped, standard spectral libraries cannot be used to model these interferences to obtain absolute concentrations. Another strategy is to include broad signals in the quantification methods, either by modelling them as signals using e.g. wavelets (Hao et al., 2012) or Lorentzians (de Graaf et al., 2015), or by treating them as baseline distortions and removing their interference by approximating a baseline correction function through the broad signal (Zheng et al., 2011; Jacob et al., 2017). Most of these methods are developed for plasma, but could potentially

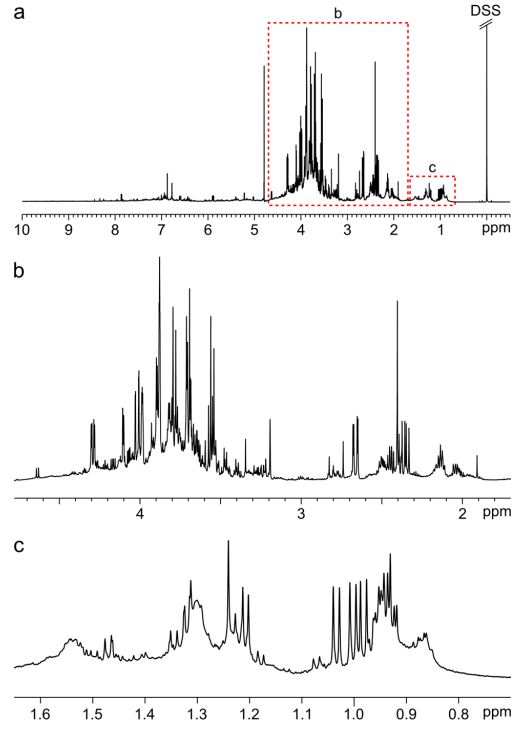


Fig. 1 **a** Typical ^{1}H NMR spectrum of oilseed rape root exudate dissolved in D_2O, **b** Magnification of spectral region 1.7–4.8 ppm, **c** Magnification of spectral region 0.70–1.65 ppm

also be applied to the plant root exudate samples used as test system in the current study.

The aim of the current study was to develop a rapid, straightforward, and computationally efficient extension of AQuA for absolute quantification of selected metabolites in highly complex spectra containing broad background signals. The method should require minimal sample preparation without compromising the quantitative accuracy. Because of speed and computational cost, we decided to remove the interferences caused by the broad signals before the AQuA computation. This was done using an automatic baseline correction function; here we employed the widely used adaptive iteratively reweighted penalised least squares (airPLS) algorithm (Zhang et al., 2010). The combined method, called extended AQuA, was evaluated using simulations as well as a spike-in experiment performed in a complex sample matrix. This showed that the approach is both accurate, linear, and robust. Furthermore, the proposed workflow is fast and flexible and can easily be fine-tuned for individual samples.

2 Materials and methods

2.1 Root exudate collection

Seeds of various spring varieties of oilseed rape (*Brassica napus*) were kindly provided by Scandinavian Seed AB and Lantmännen Seed AB. All glassware was rinsed extensively with MilliQ water and autoclaved before use to minimise traces of detergents. Seeds were surface sterilised (10% chlorine bleach for 5 min with mild shaking) and then rinsed with autoclaved MilliQ water four times. The seeds were germinated on petri dishes containing 0.5× Murashige-Skoog medium, including vitamins (MS0222, Duchefa Biochemie B.V., Haarlem, Netherlands) and 0.6% bacto agar, in a growth chamber at 22/20 °C (day/night), 16/8 h photoperiod with 110 µE. After three to five days of germination, when cotyledons and rootlets were expanded, plantlets ($n = 8$) were transferred to sterile plastic nets attached to 50 ml plastic tubes filled with autoclaved MilliQ water, so that the seedling roots were immersed into the water. This procedure was done in a sterile laminar flow hood. The samples were placed in a sterilised transparent plastic box and kept for four days with slow agitation in a growth chamber at 22/20 °C (day/night), 16/8 h photoperiod with 110 µE. Exudates were collected into glass bottles in a sterile laminar flow hood, shell frozen and lyophilised in darkness. Aliquots of the exudates were spread on plates containing LB agar or 0.5× Murashige-Skoog agar and stored for 48 h to assess any microbial contamination. Blank samples did not contain any seedlings but were otherwise treated as described above.

Lyophilised root exudate and blank samples were dissolved in a few millilitres of MilliQ water, transferred to 15 ml plastic tubes, and dried in a vacuum centrifuge. Dried samples were stored in a desiccator until use.

2.2 Sample preparation

NMR samples were prepared in a similar fashion to a previously published protocol (Kim et al., 2010). All experimental work was performed at room temperature. 750 µl KH_2PO_4 buffer in D_2O (45 mM, pD 7.0 (apparent pH 6.6) containing approximately 0.29 mM DSS-d_6 (sodium 3-(trimethylsilyl) propane-1-sulfonate-d_6) was added to each sample. The samples were vortexed 30 s followed by 10 min ultrasonication. This procedure was repeated once. The samples were then transferred to 1.5 ml plastic tubes and centrifuged for 10 min at 17 000×g. For each sample, 600 µl of the supernatant was added to a 5 mm NMR tube.

2.3 NMR spectroscopy and spectral processing

NMR spectra were acquired on a Bruker Avance III 600 MHz spectrometer with a 5 mm ^{1}H/^{13}C/^{15}N/^{31}P inverse detection cryoprobe equipped with a z gradient. 1D-^{1}H NMR spectra (256 transients) were recorded at 25 °C using a NOESY presaturation pulse sequence (Bruker's *noesypr1d*) with 1 s relaxation delay, 100 ms mixing time, 4.5 s acquisition time, and 12 ppm spectral width, to enable absolute quantification based on the Chenomx library. 65 536 data points were collected and the carrier frequency was placed on the HDO signal (4.70 ppm). After acquisition, an exponential line broadening of 0.3 Hz was applied and the spectral quality was evaluated by assessing the full width half maximum (FWHM) of the DSS signal. If $FWHM_{DSS}$ was greater than 1.20 Hz, a new spectrum was recorded. Spectra were processed (zero-filling, line broadening, phase correction, crude baseline correction) using Chenomx NMR Suite Professional Software package (version 8.6, Chenomx Inc., Edmonton, Canada). The line-broadening factor was adjusted for each spectrum to obtain $FWHM_{DSS} = 1.20$ Hz. If necessary, a crude baseline correction was applied to obtain a flat baseline around the internal standard signal before determining $FWHM_{DSS}$. The processed spectra were subjected to spectral binning (−0.50 to 4.68 ppm and 4.98 to 10.00 ppm, 0.0002 ppm/bin, 51 000 bins in total) and imported to MATLAB (version R2020a, MathWorks Inc., Natick (MA), USA).

To verify metabolite identification, ^{1}H,^{1}H-TOCSY (Bruker pulse sequence *dipsi2gpphpr*) and ^{1}H,^{13}C-HSQC (Bruker pulse sequence *hsqcedetgpsisp.2*) spectra were recorded for some of the samples. These spectra were processed with TopSpin 4.0.6 (Bruker BioSpin).

2.4 Metabolite identification and quantification

AQuA does not attempt at automated metabolite identification, hence metabolite signals have to be selected prior to AQuA computation. Here, identification of metabolites was based on previous literature (Vives-Peris et al., 2020) and reference NMR spectra included in the Chenomx library. The identity of the metabolites was verified with ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC NMR spectra recorded for some of the samples. ^{13}C NMR chemical shifts were compared with those available in the Biological Magnetic Resonance Data Bank (Ulrich et al., 2008). Only metabolites verified by 2D NMR experiments, or displaying an excellent fit for several signals with the Chenomx library, were included in the quantification model. Furthermore, only primary metabolites were included since one of the aims was to develop a method capable of quantifying only a subset of all metabolites in an NMR spectrum.

Binned processed NMR spectra (see Sect. 2.3) were imported to MATLAB and subjected to the airPLS algorithm (Zhang et al., 2010) to fine-tune the baseline where affected by irregularities or the presence of broad signals. As default, the airPLS smoothing factor λ was set to 1×10^7, but a local value was determined for spectral regions where the default λ failed to yield a satisfactory baseline correction. The values for the other parameters in the airPLS algorithm were used as default (order = 2, weight exception proportion = 0.1, asymmetry parameter = 0.05, and maximum iteration time = 20). The airPLS algorithm, using the optimised λ values, was incorporated in an automated joint workflow with AQuA in MATLAB. This workflow is referred to as the extended AQuA. Metabolite quantification using AQuA was performed on the corrected spectra according to the strategy previously described (Röhnisch et al., 2018), using the Chenomx library as a basis to model metabolite signals. In total, 24 metabolites were targeted for quantification, including various amino acids, organic acids, and sugars (Table S1). One reporter signal to be used for quantification was selected for each metabolite (Table S1). Additionally, a few unknown signals were included in the model as Lorentzians generated in Chenomx (Fig. S1).

2.5 Simulations

A simple smoothing algorithm developed in-house was applied to one root exudate spectrum to model the spectral background. The algorithm was built in MATLAB based on the 'smooth' function. In short, the following steps were employed: (1) localisation of narrow high-intensity signals (spikes), (2) determination of spike borders, (3) spike depletion by linear regression inside spike borders, and (4) average-based smoothing of the spike-depleted spectrum (for more information and a visual description of the process, see Supplementary Information Sect. 3, especially Fig. S5). In the final step, three levels of smoothing (low, medium, and high) were used to obtain three distinct spectral background models (referred to as A, B, and C, respectively, see Fig. S6). Normalised reference spectra of 24 metabolites (Table S1) were summed together and added to each spectral background in seven different scaling levels, thus yielding 21 simulated spectra. The spectra were corrected with the airPLS algorithm using three different λ values (1×10^6, 1×10^7, and 1×10^8) applied to the whole spectra. Peak picking of one signal per metabolite was performed as previously described (Röhnisch et al., 2018) to obtain signal intensities in the corrected spectra.

2.6 Spike-in experiment

Six of the analysed root exudate samples were pooled together and then divided into five portions. Five metabolites (γ-aminobutyric acid (GABA), DL-asparagine, L(+)-tartaric acid, L-threonine, and D-xylose) not present in the pooled sample were added to different concentrations. As control, five identical blank samples were spiked the same way. The chosen metabolites have signals in different spectral regions with different multiplicities and differ in how much they are affected by broad signals or baseline distortions. The large variation in concentration (10 µM-3200 µM) between the spiked metabolites reflects the large dynamic range observed in the experimental data set, both between different metabolites in the same sample and between the same metabolite in different samples. See Supplementary Information, Sect. 4.1, for more details about the design of the spike-in experiment.

The spiked root exudate samples were analysed as described above, i.e. NMR analysis, spectral processing, and metabolite quantification using an airPLS-extended AQuA, which had been adjusted to include all spiked metabolites (Table S1).

The spectra of the spiked blank spectra were carefully baseline corrected in Chenomx. For the analysis of these spectra, the airPLS step was omitted and an AQuA that only targeted the five spiked metabolites plus lactic acid was used to calculate metabolite concentrations.

3 Results and discussion

3.1 Extended AQuA: workflow, parameter optimisation, and general considerations

The 1D-^{1}H NMR spectra of oilseed rape root exudates displayed baseline irregularities, including broad background signals, that would impair metabolite concentration estimates if not properly accounted for (Fig. 1). The broad signals in the low-frequency part of the spectra were the most

problematic distortions, due to their interference with several amino acid signals. Different methods for elimination of the baseline distortions were evaluated, utilising sample preparation, spectral editing, and computations, respectively (see Supplementary Information, Sect. 2). It was found that an automatic baseline correction function such as the airPLS algorithm (Zhang et al., 2010) could be employed to yield root exudate spectra suitable for targeted metabolomics, i.e. with well-preserved metabolite signal line shapes, a flat baseline, no pronounced residual broad signals, and no severe intensity modulation (see Fig. S2). Manual baseline correction was not considered feasible due to the complexity of the spectra.

The airPLS algorithm was combined with AQuA into a joint automated workflow, i.e. the extended AQuA, for quantification of metabolites in experimental ^{1}H NMR spectra of root exudates, acquired with minimal prior sample preparation (Fig. 2). The identity of the metabolites was confirmed with ^{1}H,^{1}H-TOCSY and ^{1}H,^{13}C-HSQC experiments. Because the AQuA quantification is based on just one signal per metabolite, the airPLS algorithm was used to obtain a good baseline around these signals only, rather than aiming for a perfect baseline in the entire spectrum. The spectral library used here was created from Chenomx but other sources, e.g. in-house libraries, can be used instead if desired.

For the airPLS algorithm to work properly, the smoothing factor λ needs to be optimised. This parameter, which can be set to any value between 1 and 1×10^{9} (Zhang et al., 2010), strongly affects the result of the baseline correction. If λ is set too high, the fitted baseline does not include enough of the background, whereas if it is set too low, the algorithm starts to remove parts of the metabolite signals (Fig. 3). Here, due to the non-uniform distribution of broad signals and other baseline distortions, a single λ value was not used for an entire spectrum; instead, different λ values were used for different spectral regions (see Sect. 3.2). Despite the virtually unlimited number of options, it was neither difficult nor time-consuming to find suitable λ values. Importantly, the optimised λ values could be kept fairly constant throughout each data set and could thereby be included in the automated workflow. Before applying the extended AQuA to a data set, the result of the baseline correction should be assessed carefully on a representative subset of the spectra, although one has to keep in mind that the procedure is inevitably an estimation and may not exactly match the actual baseline of the spectrum. However, this is true for all baseline correction methods, regardless of if they are manual or automated.

In addition to baseline distortions, interference can also be caused by spectral overlap with narrow unknown signals. In the current study, the aim was to quantify a preselected subset of metabolites while leaving remaining signals in the spectra untargeted. However, other signals that interfere with the metabolite signals used in AQuA need to be included in the quantification model to avoid overestimating the metabolite concentrations. Here, four unknown signals between 0.93 and 0.97 ppm were added to the quantification model as single Lorentzians to obtain a more accurate concentration estimate of leucine based on the signal at 0.96 ppm (Table S1 and Fig. S1).

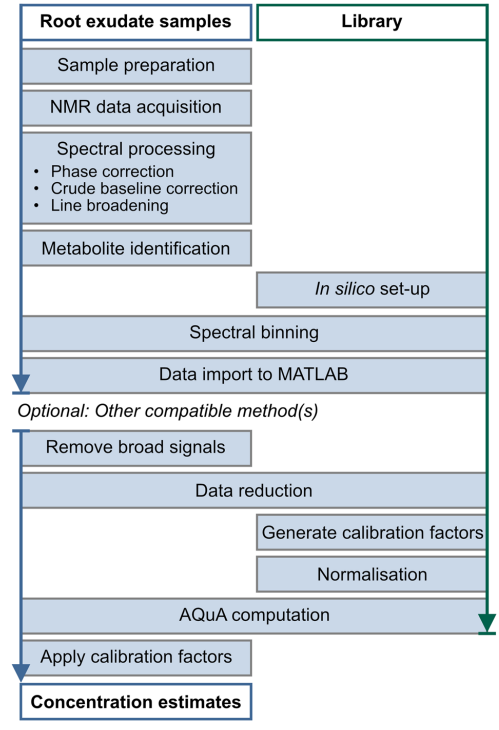


Fig. 2 Proposed workflow for NMR-based quantification of primary metabolites in plant root exudates after minimal sample preparation (see Materials and methods). The workflow includes the airPLS algorithm (Zhang et al., 2010) for removal of broad signals, followed by quantification using AQuA (Röhnisch et al., 2018). The *in silico* library, here created from Chenomx, contains 1D-^{1}H NMR spectra of all targeted metabolites. Other compatible methods, e.g. for signal alignment, can also be included in the workflow if desired

3.2 Evaluation of the extended AQuA

3.2.1 Simulations

The extended AQuA was first evaluated using simulated spectra of root exudates where the contributions of the broad signal background and the narrow metabolite signals were

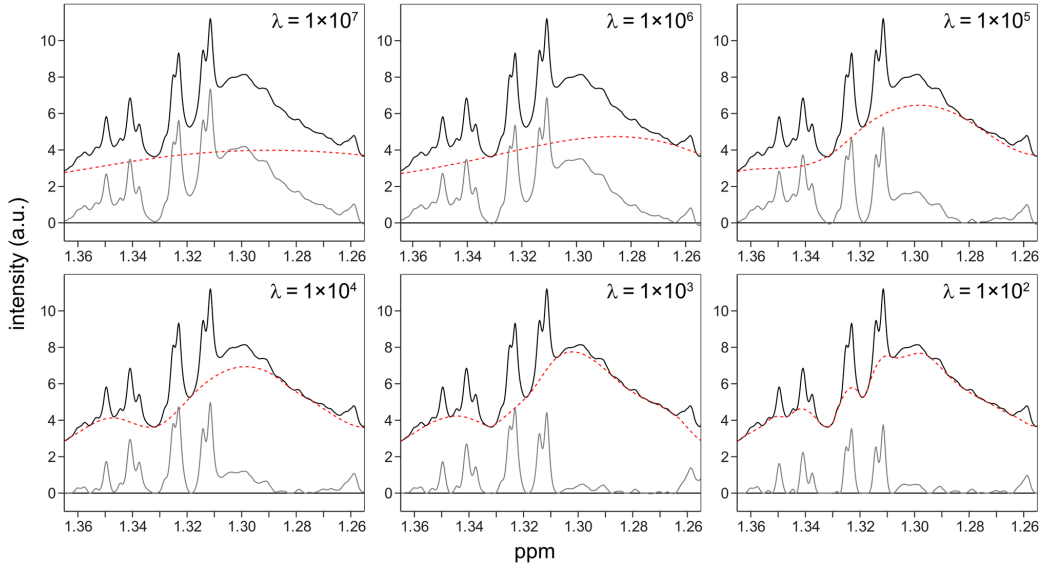


Fig. 3 The effect of different λ values ($1\times10^7 - 1\times10^2$) on the baseline correction of the lactic acid/threonine region of a root exudate spectrum. Black: experimental spectrum before baseline correction, dashed: fitted baseline, grey: experimental spectrum after baseline correction. Note that the experimental and corrected spectra are superimposed, not stacked on top of each other

exactly known (Fig. 4). To test how well the method can handle different types of spectra, three different spectral background models (A, B, and C) with varying smoothness were created (Figs. S5 and S6) and a simulated narrow signal spectrum was added to the backgrounds in seven different intensity levels. In total, 21 simulated spectra were thus obtained with differences in their spectral backgrounds as well as in their ratio between narrow and broad signals (Figs. S7–S10 and Table S2). For reference, Fig. 4a depicts the simulated spectrum created with the medium-smooth background B (Fig. 4b) and an intermediate intensity of the narrow signal spectrum (Fig. 4c). The airPLS algorithm was applied three times to all spectra, with three different λ values, to evaluate the robustness of the method. The signal heights in the airPLS corrected spectra (Fig. 4d) were compared to those in the corresponding narrow signal spectra (Fig. 4c) using linear regression. Thereby, it was possible to precisely assess how well the airPLS algorithm could remove interferences caused by broad signals and baseline irregularities, and to what extent the narrow signal part of the spectrum was affected by the procedure.

In general, the agreement between the intensities in the baseline corrected spectra and the original narrow signal spectra was good for the signals used in AQuA, as indicated by slopes and R^2 coefficients close to one and intercepts close to zero (Fig. 5 and Table S3). This suggests that the airPLS feature specifically corrected the baseline and removed broad background signals without notably affecting the selected metabolite signals. Percentage differences (Table S3) were calculated to condense the accuracy estimate into a single variable. For most metabolites, the difference was less than 10% with at least one of the λ values. The smoother backgrounds B and C were easier to fit than the rougher background A, hence the smaller intercepts (Fig. 5b and c). Overall, when the airPLS algorithm was applied to spectra created using background A the λ value needed to be smaller than for spectra based on background B or C. For metabolite signals situated in spectral regions without background interference, e.g. formic acid and fumaric acid, the accuracy was good for all spectra regardless of which λ value was being used (Fig. 5 and Table S3). In contrast, some metabolites had signal intensities in the corrected spectra that deviated substantially from their true values. Often, this coincided with a pronounced interference from the spectral background (see Table S2 and Fig. S10). For example, the signals of fructose, glyceric acid, lactic acid, and threonine were all highly influenced by the spectral background and so the quantification accuracy of these metabolites was strongly dependent on the performance of the baseline correction. Because of the large variation in signal intensity in the simulated spectra, the results could have been more accurate if the λ value had been optimised for each individual spectrum

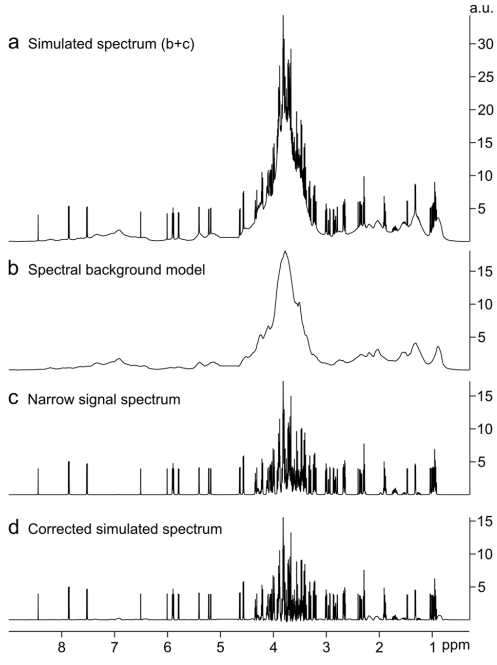


Fig. 4 Examples of simulated spectra used to evaluate the extended AQuA. **a** Simulated root exudate spectrum, constructed from a simulated spectral background **b** and a simulated narrow signal spectrum **c**. **d** The simulated root exudate spectrum after correction with the airPLS algorithm ($\lambda = 1 \times 10^7$ for the entire spectrum). Ideally, the spectra in **c** and **d** should be identical. An intensity scale has been added to all spectra to facilitate comparison between the spectra

(see next section). However, in an experimental data set, the inter-spectral variation is usually not as big. Furthermore, the quantification accuracy for a given metabolite generally increased with increasing signal intensity relative to the spectral background. Thus, the lower the intensity of the narrow signals and the higher the intensity of the background, the more critical it becomes to optimise the method parameters to avoid quantification errors. Ideally, the signals used in the AQuA computation should all have a low degree of interference and high signal to noise ratio (Röhnisch et al., 2018); however, this is not possible for all metabolites. Still, the proposed method appears to be both linear and accurate for most metabolites.

3.2.2 Spike-in experiment

A spike-in experiment was conducted to further evaluate the extended AQuA (Tables S4, S5 and Fig. S11). Five metabolites (asparagine, GABA, tartaric acid, threonine, and xylose) were added both to blank samples and to aliquots of a pooled root exudate sample in concentrations above the limit of quantification (10 × S/N) for each metabolite. The blank spectra displayed minimal signal interference and lacked the broad background signals and baseline distortions that were present in the root exudate spectra. Therefore, these spectra were only subjected to manual baseline correction before the AQuA computation. The spiked root exudate spectra, on the other hand, were baseline corrected with the airPLS algorithm to remove broad background signals. Here, the default λ value gave a satisfactory correction for all metabolites except threonine and GABA, as evaluated by manual inspection. Threonine was the most challenging metabolite to quantify in the spiked root exudate spectra because its selected signal overlapped both with the signal of the methyl group of lactic acid and with a broad signal that was not assigned unambiguously but can be tentatively attributed to a lipid methylene signal (Fig. S11). The latter could not be correctly suppressed unless a lower λ value was used (see Fig. 3). The GABA signal is a broad quintet whose intensity was slightly reduced with the default λ value because the fitted baseline removed a small portion of the signal (Fig. S12). Therefore, the size of λ was increased for this spectral region.

After baseline correction, AQuA computation was performed on both the spectra from the spiked root exudates and the spiked blank samples, and the results were compared with each other using linear regression as well as percent differences (Table 1). The calculated concentrations are listed in Table S6. Because the same amount of metabolites were added to both sample sets, all slopes should theoretically be equal to one, and all intercepts should be equal to zero as none of the spiked metabolites were present in the samples initially. However, since the sample matrices differed somewhat and all metabolite additions were done manually, some deviations could be expected. Still, as shown in Table 1, the R^2 values were > 0.999, all intercepts were close to the origin, and the percent differences were generally small. This was in agreement with the results from the simulations. To enable comparison of the intercepts amongst the different metabolites despite the big differences in concentration, the intercepts are reported both as the actual value and as percent of the highest concentration for each metabolite. The tartaric acid signal was consistently more intense in the spectra of the spiked root exudate samples than in the spectra of the corresponding blank samples (Fig. S13), hence the large slope and percent differences. An experimental error probably occurred when tartaric acid was added to the root exudate samples since the calculated concentration of tartaric acid in the blank samples, but not the root exudate samples, agreed well with the actual concentrations (Tables S6–S8). The values of the intercepts and slopes for the other metabolites indicated that there was no clear, systematic over- or underestimation of the concentrations obtained using the

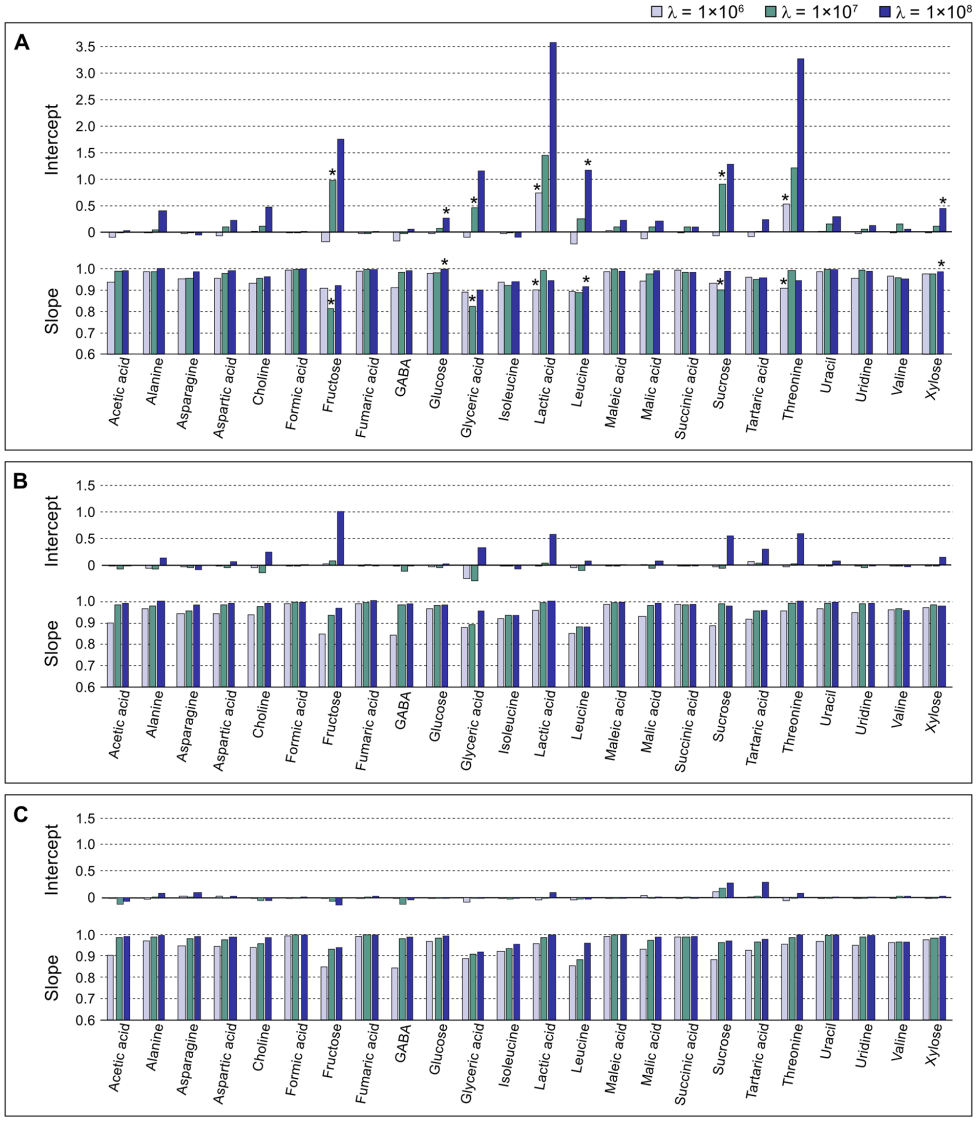


Fig. 5 Evaluation of the extended AQuA applied to seven simulated spectra constructed using spectral background model A, B, or C (see Figs. S5–S9). Linear regression was performed using signal heights in the simulated narrow signal spectra as predictor (x-axis) and signal heights in the corresponding simulated spectra with both broad and narrow signals, after correction with the airPLS algorithm, as response (y-axis). Only the signals used in AQuA were evaluated. Bar heights display values of the intercepts and slopes for each metabolite in the simulated spectra. Bar colour indicates the λ value used in the baseline correction. Asterisks denote linear regressions with $R^2 < 0.9900$

proposed method compared to when the same metabolites, in the absence of baseline distortions, were quantified with the non-extended AQuA.

3.3 Application to plant root exudates

The extended AQuA was applied to a data set consisting of 50 NMR spectra from oilseed rape root exudates and 7

Table 1 Comparison of the concentrations obtained for the spiked blank samples and the concentrations obtained for the spiked root exudate samples[a]

Metabolite	Max conc. (μM)[b]	R^2	Slope	Intercept (μM)	Rel. intercept[c] %	Mean % difference blank-sample[d]
Asparagine	1604	1.0000	0.976	−0.474	−0.0295	2.7
GABA	403	1.0000	1.00	−3.23	−0.801	4.0
Tartaric acid	800	0.9994	1.16	6.70	0.837	20.0
Threonine	164	0.9996	1.01	2.22	1.36	6.2
Xylose	3224	1.0000	0.999	4.49	0.139	1.2

[a]Results from linear regression. Predictor (x-axis): Concentrations for spiked blank samples calculated using an AQuA including only the five spiked metabolites and lactic acid. Response (y-axis): Concentrations for spiked root exudate samples calculated using an airPLS-extended AQuA including the metabolites listed in Table S1. Used airPLS parameters: $\lambda_{default} = 1 \times 10^7$, $\lambda_{Thr} = 1 \times 10^5 - 1 \times 10^6$ (depending on the intensity of the threonine signal), $\lambda_{GABA} = 1 \times 10^8$

[b]Actual value for the spiked sample with the highest concentration

[c]Intercept as percent of the maximum concentration for the metabolite

[d]Average difference (%) for each spiked metabolite, comparing the calculated concentrations found for the spiked root exudate samples with the calculated concentrations for the corresponding blank samples (calculated as $100 \times |C_{blank} - C_{sample}|/C_{blank}$). See Table S6 for a complete list of difference values

blank spectra. Concentration estimates were computed for 24 metabolites (Table S1). Additionally, four unknown signals were included to model signal interferences (Fig. S1) but they were not quantitatively interpreted.

The extended AQuA process (i.e. baseline correction followed by the quantification of 24 target metabolites) applied to all 57 spectra was typically completed in less than 30 s on a standard personal computer. The same method parameters were used for all spectra. In addition to the default λ value, two local values were used in the airPLS baseline correction ($\lambda = 1 \times 10^6$ for the spectral region 0.899–0.967 ppm and $\lambda = 1 \times 10^5$ for the region 1.225–1.334 ppm). If only one λ value was used, the total computation time decreased to around 10 s. It has been shown that AQuA requires less than one second to quantify 67 metabolites in 1342 spectra (Röhnisch et al., 2018). Introducing the airPLS step thus increases the computation time but the combined method is still very rapid. Because the airPLS algorithm is the rate-limiting step, the computation time increases notably with the number of spectra and λ values whereas it is negligibly affected by the number of metabolites targeted for quantification.

3.4 Advantages and limitations

The method described here allows for quantification of metabolites in complex spectra that contain broad signals and baseline distortions. Only minimal sample preparation is required and because the method is purely computational, knowledge about the compounds causing the broad signals is not needed. However, in case of binding interactions between metabolites and other compounds such as proteins, application of the method would be more challenging. The occurrence of such interactions can be estimated by assessing the line width and shape of the internal standard signal, since both DSS and TSP are known to interact with macromolecules (Bell et al., 1989; Kriat et al., 1992; Shimizu et al., 1994; Nowick et al., 2003). Here, both metabolite signals and the internal standard signal were narrow and symmetric, which indicated that no significant macromolecular interaction was taking place.

As shown here, the baseline correction method airPLS and the quantification method AQuA can be combined into a fully automated workflow, provided that prior metabolite identification and parameter optimisation have been conducted. Optimising the airPLS algorithm is straightforward and depends only on the parameter λ. Here, we did not strive for an optimal baseline in the whole spectrum but only in regions containing signals used in AQuA, which reduced the optimisation time and effort. The combined method is extremely fast and typically requires less than one second per spectrum. This is due both to the sparse matrix characteristic of the airPLS algorithm, but more importantly the AQuA data reduction strategy. AQuA only considers a set of pre-selected signals in the quantitative process, one for each metabolite, which facilitates very rapid computations whilst still accounting for interferences between metabolites.

Here, we chose to use the airPLS algorithm for baseline correction but it is possible to use other methods instead, as long as they are compatible with AQuA. The metabolite library can also be exchanged if e.g. in-house spectral libraries are preferred.

There are also some limitations of the method. Relying on the height of one single metabolite signal for deriving concentrations may make the method more sensitive to systematic errors caused by database discrepancies compared to when several signals are used (see Supplementary

Information, Sect. 4.2). However, this has not been fully evaluated, neither have we investigated whether other quantification methods are less susceptible to this kind of errors. Since AQuA is not an identification method, there is also a risk of erroneous metabolite quantification if the chemical shift windows have not been properly selected or if there are unknown signals present in some spectra that have not been accounted for. If a signal from another compound, metabolite or impurity, with higher intensity than the intended metabolite signal resides in the chemical shift window, the algorithm will pick this signal for quantification instead. For reliable results, metabolite identification should ideally be assessed manually. However, the problem with possible false identification of metabolites is not unique to AQuA, especially when the targeted metabolite signals are singlets.

4 Conclusions

We have here presented a fast and accurate approach for automated quantification of selected metabolites in complex NMR spectra. The spectra of minimally handled plant root exudate samples were successfully analysed with the proposed method, despite the presence of unknown broad signals, baseline distortions, and extensive spectral overlap. Although not evaluated here, the method is theoretically applicable to any spectrum with similar characteristics, as long as the metabolite signals are unaffected by macromolecular interactions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11306-023-02073-z.

Acknowledgements The authors thank Dr Jan Eriksson for giving valuable comments about the manuscript.

Author contributions Conceptualization: CS, EA, HER; Methodology: AB, CS, EA, HER, GN, JM; Formal analysis: EA, HER; Investigation: EA, JM; Writing—original draft preparation: EA; Writing—review and editing: AB, CS, EA, HER, GN, JM; Funding acquisition: CS; Resources: CS, JM; Supervision: AB, CS, HER, GN, JM. All authors read and approved the final manuscript.

Funding Open access funding provided by Swedish University of Agricultural Sciences. Strategic funding for metabolomics-based research was provided by the Swedish University of Agricultural Sciences.

Data availability The NMR data has been uploaded to the Swedish National Data Service and is available at https://doi.org/10.5878/8t82-d090.

Declarations

Conflict of interest There are no conflict of interest to declare.

Ethical approval This article does not contain any studies with human or animal participants performed by any of the authors.

References

Bell, J. D., Brown, J. C. C., & Sadler, P. J. (1989). NMR studies of body fluids. *NMR in Biomedicine,2*(5–6), 246–256. https://doi.org/10.1002/nbm.1940020513

Bliziotis, N. G., Engelke, U. F. H., Aspers, R. L. E. G., Engel, J., Deinum, J., Timmers, H. J. L. M., Wevers, R. A., & Kluijtmans, L. A. J. (2020). A comparison of high-throughput plasma NMR protocols for comparative untargeted metabolomics. *Metabolomics,16*(5), 64. https://doi.org/10.1007/s11306-020-01686-y

Carr, H. Y., & Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review,94*(3), 630–638. https://doi.org/10.1103/PhysRev.94.630

Crook, A. A., & Powers, R. (2020). Quantitative NMR-based biomedical metabolomics: Current status and applications. *Molecules,25*(21), 5128. https://doi.org/10.3390/molecules25215128

Daykin, C. A., Foxall, P. J. D., Connor, S. C., Lindon, J. C., & Nicholson, J. K. (2002). The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by ^{1}H nuclear magnetic resonance spectroscopy. *Analytical Biochemistry,304*(2), 220–230. https://doi.org/10.1006/abio.2002.5637

de Graaf, R. A., & Behar, K. L. (2003). Quantitative ^{1}H NMR spectroscopy of blood plasma metabolites. *Analytical Chemistry,75*(9), 2100–2104. https://doi.org/10.1021/ac020782+

de Graaf, R. A., Prinsen, H., Giannini, C., Caprio, S., & Herzog, R. I. (2015). Quantification of ^{1}H NMR spectra from human plasma. *Metabolomics,11*(6), 1702–1707. https://doi.org/10.1007/s11306-015-0828-1

Deborde, C., Moing, A., Roch, L., Jacob, D., Rolin, D., & Giraudeau, P. (2017). Plant metabolism as studied by NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy,102–103*, 61–97. https://doi.org/10.1016/j.pnmrs.2017.05.001

Häckl, M., Tauber, P., Schweda, F., Zacharias, H. U., Altenbuchinger, M., Oefner, P. J., & Gronwald, W. (2021). An R-package for the deconvolution and integration of 1D NMR data: MetaboDecon1D. *Metabolites,11*(7), 452. https://doi.org/10.3390/metabo11070452

Hao, J., Astle, W., De Iorio, M., & Ebbels, T. M. D. (2012). BATMAN—An R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a bayesian model. *Bioinformatics,28*(15), 2088–2090. https://doi.org/10.1093/bioinformatics/bts308

Jacob, D., Deborde, C., Lefebvre, M., Maucourt, M., & Moing, A. (2017). NMRProcFlow: A graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. *Metabolomics,13*(4), 36. https://doi.org/10.1007/s11306-017-1178-y

Kim, H. K., Choi, Y. H., & Verpoorte, R. (2010). NMR-based metabolomic analysis of plants. *Nature Protocols,5*(3), 536–549. https://doi.org/10.1038/nprot.2009.237

Kriat, M., Confort-Gouny, S., Vion-Dury, J., Sciaky, M., Viout, P., & Cozzone, P. J. (1992). Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards. *NMR Biomedicine,5*(4), 179–184. https://doi.org/10.1002/nbm.1940050404

Lefort, G., Liaubet, L., Canlet, C., Tardivel, P., Père, M. C., Quesnel, H., Paris, A., Iannuccelli, N., Vialaneix, N., & Servien, R. (2019). ASICS: An R package for a whole analysis workflow of 1D ^{1}H NMR spectra. *Bioinformatics,35*(21), 4356–4363. https://doi.org/10.1093/bioinformatics/btz248

Liu, M., Nicholson, J. K., & Lindon, J. C. (1996). High-resolution diffusion and relaxation edited one- and two-dimensional ^{1}H NMR spectroscopy of biological fluids. *Analytical Chemistry,68*(19), 3370–3376. https://doi.org/10.1021/ac960426p

Martineau, E., Dumez, J. N., & Giraudeau, P. (2020). Fast quantitative 2D NMR for metabolomics and lipidomics: A tutorial. *Magnetic Resonance in Chemistry,58*(5), 390–403. https://doi.org/10.1002/mrc.4899

Meiboom, S., & Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments,29*(8), 688–691. https://doi.org/10.1063/1.1716296

Nagana Gowda, G. A., Gowda, Y. N., & Raftery, D. (2015). Expanding the limits of human blood metabolite quantitation using NMR spectroscopy. *Analytical Chemistry,87*(1), 706–715. https://doi.org/10.1021/ac503651e

Nagana Gowda, G. A., & Raftery, D. (2014). Quantitating metabolites in protein precipitated serum using NMR spectroscopy. *Analytical Chemistry,86*(11), 5433–5440. https://doi.org/10.1021/ac5005103

Nowick, J. S., Khakshoor, O., Hashemzadeh, M., & Brower, J. O. (2003). DSA: A new internal standard for NMR studies in aqueous solution. *Organic Letters,5*(19), 3511–3513. https://doi.org/10.1021/ol035347w

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., … Wishart, D. S. (2011). The human serum metabolome. *PLoS ONE,6*(2), e16957. https://doi.org/10.1371/journal.pone.0016957

Ravanbakhsh, S., Liu, P., Bjordahl, T. C., Mandal, R., Grant, J. R., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R., & Wishart, D. S. (2015). Accurate, fully-automated NMR spectral profiling for metabolomics. *PLoS ONE,10*(5), e0124219. https://doi.org/10.1371/journal.pone.0124219

Röhnisch, H. E., Eriksson, J., Müllner, E., Agback, P., Sandström, C., & Moazzami, A. A. (2018). AQuA: An automated quantification algorithm for high-throughput NMR-based metabolomics and its application in human plasma. *Analytical Chemistry,90*(3), 2095–2102. https://doi.org/10.1021/acs.analchem.7b04324

Röhnisch, H. E., Eriksson, J., Tran, L. V., Müllner, E., Sandström, C., & Moazzami, A. A. (2021). Improved automated quantification algorithm (AQuA) and its application to NMR-based metabolomics of EDTA-containing plasma. *Analytical Chemistry,93*(25), 8729–8738. https://doi.org/10.1021/acs.analchem.0c04233

Rout, M., Lipfert, M., Lee, B. L., Berjanskii, M., Assempour, N., Vazquez Fresno, R., Serra Cayuela, A., Dong, Y., Johnson, M., Shahin, H., Gautam, V., Sajed, T., Oler, E., Peters, H., Mandal, R., & Wishart, D. S. (2023). MagMet: A fully automated web server for targeted nuclear magnetic resonance metabolomics of plasma and serum. *Magnetic Resonance in Chemistry*. https://doi.org/10.1002/mrc.5371

Shimizu, A., Ikeguchi, M., & Sugai, S. (1994). Appropriateness of DSS and TSP as internal references for ^{1}H NMR studies of molten globule proteins in aqueous media. *Journal of Biomolecular NMR,4*(6), 859–862. https://doi.org/10.1007/BF00398414

Takis, P. G., Jiménez, B., Al-Saffar, N. M. S., Harvey, N., Chekmeneva, E., Misra, S., & Lewis, M. R. (2021). A computationally lightweight algorithm for deriving reliable metabolite panel measurements from 1D ^{1}H NMR. *Analytical Chemistry,93*(12), 4995–5000. https://doi.org/10.1021/acs.analchem.1c00113

Takis, P. G., Jiménez, B., Sands, C. J., Chekmeneva, E., & Lewis, M. R. (2020). SMolESY: An efficient and quantitative alternative to on-instrument macromolecular ^{1}H-NMR signal suppression. *Chemical Science,11*(23), 6000–6011. https://doi.org/10.1039/D0SC01421D

Tardivel, P. J. C., Canlet, C., Lefort, G., Tremblay-Franco, M., Debrauwer, L., Concordet, D., & Servien, R. (2017). ASICS: An automatic method for identification and quantification of metabolites in complex 1D ^{1}H NMR spectra. *Metabolomics,13*(10), 109. https://doi.org/10.1007/s11306-017-1244-5

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Wenger, R. K., Yao, H., & Markley, J. L. (2008). BioMagResBank. *Nucleic Acids Research,36*(suppl_1), D402–D408. https://doi.org/10.1093/nar/gkm957

Vives-Peris, V., de Ollas, C., Gómez-Cadenas, A., & Pérez-Clemente, R. M. (2020). Root exudates: From plant to rhizosphere and beyond. *Plant Cell Reports,39*(1), 3–17. https://doi.org/10.1007/s00299-019-02447-5

Weljie, A. M., Newton, J., Mercier, P., Carlson, E., & Slupsky, C. M. (2006). Targeted profiling: Quantitative analysis of 1H NMR metabolomics data. *Analytical Chemistry,78*(13), 4430–4442. https://doi.org/10.1021/ac060209g

Zhang, Z. M., Chen, S., & Liang, Y. Z. (2010). Baseline correction using adaptive iteratively reweighted penalized least squares. *The Analyst,135*(5), 1138–1146. https://doi.org/10.1039/B922045C

Zheng, C., Zhang, S., Ragg, S., Raftery, D., & Vitek, O. (2011). Identification and quantification of metabolites in ^{1}H NMR spectra by bayesian model selection. *Bioinformatics,27*(12), 1637–1644. https://doi.org/10.1093/bioinformatics/btr118

Supplementary information

Extended Automated Quantification Algorithm (AQuA) for targeted ^{1}H NMR metabolomics of highly complex samples: Application to plant root exudates

Elin Alexandersson[1*], Corine Sandström[1], Johan Meijer[2], Gustav Nestor[1], Anders Broberg[1] & Hanna E. Röhnisch[1]

[1] Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden
[2] Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden

* Corresponding author. E-mail: elin.alexandersson@slu.se

Table of content

1. Metabolite library

Table S1 The metabolites included in AQuA (indicated by "X") applied to the different data sets

	Signal used in AQuA				Data set			
Metabolite	**δ (ppm)**	**mult.**	**no. H**	**structure**	**Root exudates**	**Simulations**	**Spike-in, samples**	**Spike-in, blanks**
Acetic acid	1.91	s	3	$–CH_3$	X	X	X	
Alanine	1.48	d	3	$–CH_3$	X	X	X	
Asparagine	2.86	dd	1	$–CH_2–$	X	X	X	X
Aspartic acid	2.68	dd	1	$–CH_2–$	X	X	X	
Choline	3.19	s	9	$–CH_3$	X	X	X	
Formic acid	8.45	s	1	=CH–	X	X	X	
Fructose	4.10	m	2	>CH–	X	X	X	
Fumaric acid	6.51	s	2	=CH–	X	X	X	
GABA	1.89	m	2	$–CH_2–$	X	X	X	X
Glucose	5.22	d	1	>CH–	X	X	X	
Glutamine	2.46	m	2	$–CH_2–$	X			
Glyceric acid	4.07	dd	1	>CH–	X	X	X	
Isoleucine	1.01	d	3	$–CH_3$	X	X	X	
Lactic acid	1.31	d	3	$–CH_3$	X	X	X	X
Leucine	0.96	m	6	$–CH_3$	X	X	X	
Maleic acid	6.01	s	2	=CH–	X	X	X	
Malic acid	2.68	dd	1	$–CH_2–$	X	X	X	
Pyroglutamic acid	4.17	dd	1	>CH–	X			
Succinic acid	2.40	s	4	$–CH_2–$	X	X	X	
Sucrose	4.22	d	1	>CH–	X	X	X	
Tartaric acid	4.31	s	2	>CH–		X	X	X
Threonine	1.31	d	3	$–CH_3$	X	X	X	X
Unknown-1[a]	0.94	s	-	-	X			
Unknown-2[a]	0.95	s	-	-	X			
Unknown-3[a]	0.96	s	-	-	X			
Unknown-4[a]	0.97	s	-	-	X			
Uracil	5.80	d	1	=CH–	X	X	X	
Uridine	5.88	d	1	=CH–	X	X	X	
Valine	1.03	d	3	$–CH_3$	X	X	X	
Xylose	5.18	d	1	>CH–		X	X	X

[a] The unknown signals were included as Lorentzians to model signal interference in the root exudate spectra (see Fig. S1) but they were not quantitatively interpreted.

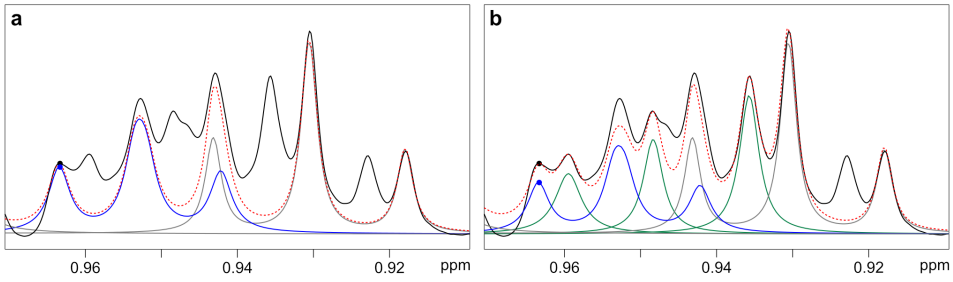


Fig. S1 Overlap with unknown signals. The leucine signal that was selected for AQuA (blue) overlaps with unknown signals that need to be included in the model for accurate concentration estimation. The red dashed line represents the model sum. When only known metabolites (blue and grey) are included in the model, the calculated height of the leucine signal (blue dot) closely matches that of the experimental target signal (black dot) (a). When unknown signals (green) are added to the model, the leucine signal is estimated to be less intense, because interference contributions from the unknown signals are now taken into account (b). The library signals of the unknowns were generated in Chenomx as Lorentzians whose positions and signal line widths were matched manually, based on visual inspection, to one representative root exudate NMR spectrum.

2. Evaluation of different methods for suppression of broad signals in 1D-^{1}H NMR spectra

2.1 Evaluation results

Ultrafiltration is the routine procedure to remove macromolecules from blood samples. However, the broad signals in the root exudate sample were unaffected by this procedure, meaning that they are likely not caused by macromolecules larger than 3 kDa (Fig. S2). The sample was also passed through a C18 solid phase extraction (SPE) column. In the aqueous filtrate, some of the broad signals have been completely removed whereas others have been diminished. The baseline was also generally improved. However, this procedure inevitably disturbs the integrity of the sample and some compounds, including the internal standard DSS, were not fully recovered in the first elution step but appear in the subsequent methanolic eluate as well (Fig. S3). Thus, the method cannot be used for absolute metabolite quantification. Furthermore, since some broad signals are still present in the aqueous filtrate, it can be concluded that the underlying compounds are not particularly hydrophobic.

The CPMG pulse sequence (Carr & Purcell 1954; Meiboom & Gill 1958) is a common approach to selectively suppress broad NMR signals based on their short transverse relaxation times (T_2) compared to narrower signals. Here, with the CPMG experiment it was possible to reduce the intensity of the broad signals, but not without severely affecting certain metabolite signals (Fig. S2). Diffusion experiments, in which resonances from fast-diffusing compounds are defocused, also allowed the intensity of the broad signals in the difference spectrum to be reduced, but other signals were suppressed as well (Fig. S2 and S4). Thus, it appears that there is no distinct difference in T_2 relaxation times or diffusion coefficients between the compounds causing the broad signals and the metabolites of interest, and it was therefore not possible to find NMR parameters that selectively targeted all broad signals, but no metabolite signals, using either CPMG or diffusion experiments. The diffusion experiments also affected individual broad signals differently, implying that the underlying compounds have different diffusion coefficients. This particular problem has been addressed in a previous study where the intensities of individual lipoprotein signals were allowed to vary ±10 % to achieve a good fit with the original 1D-^{1}H spectrum (de Graaf et al. 2015). However, because the spread in diffusion coefficients is much larger in our study, a higher intensity variation would have to be allowed to accurately fit the spectra.

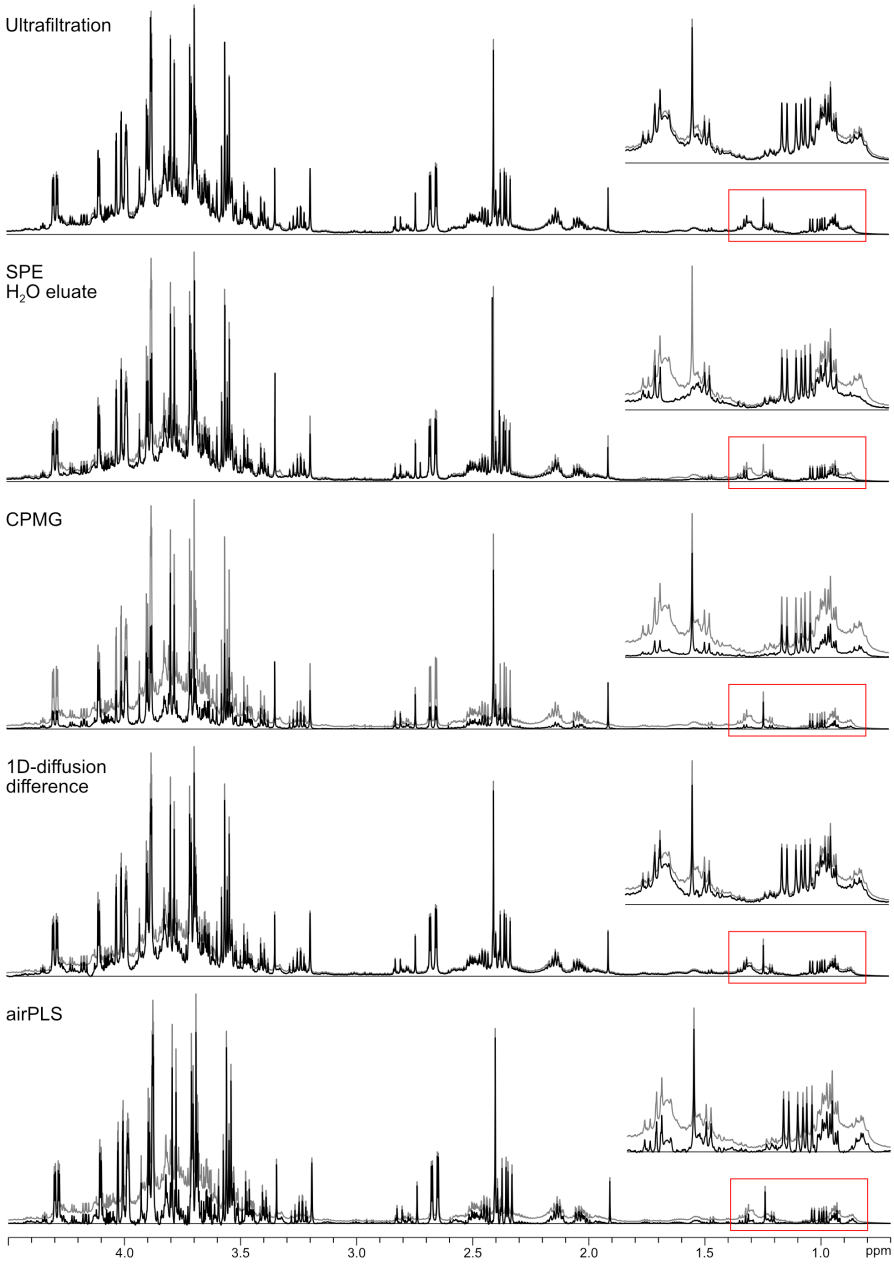


Fig. S2 Ultrafiltration, SPE, CPMG, 1D-diffusion, and airPLS applied to the same root exudate sample (black spectra). All NMR spectra are overlaid with a 1D-NOESY presaturation spectrum recorded on the intact sample (grey). A baseline at zero intensity has been added to all spectra. The insets show magnifications of the spectral region 0.8-1.4 ppm. For the airPLS computation, the default λ value was 1×10^7. In addition, two local λ values were used: 1×10^5 for the spectral region 1.23-1.33 ppm and 1×10^6 for the region 0.90-0.97 ppm. For the other methods, see section 2.2 for parameter details.

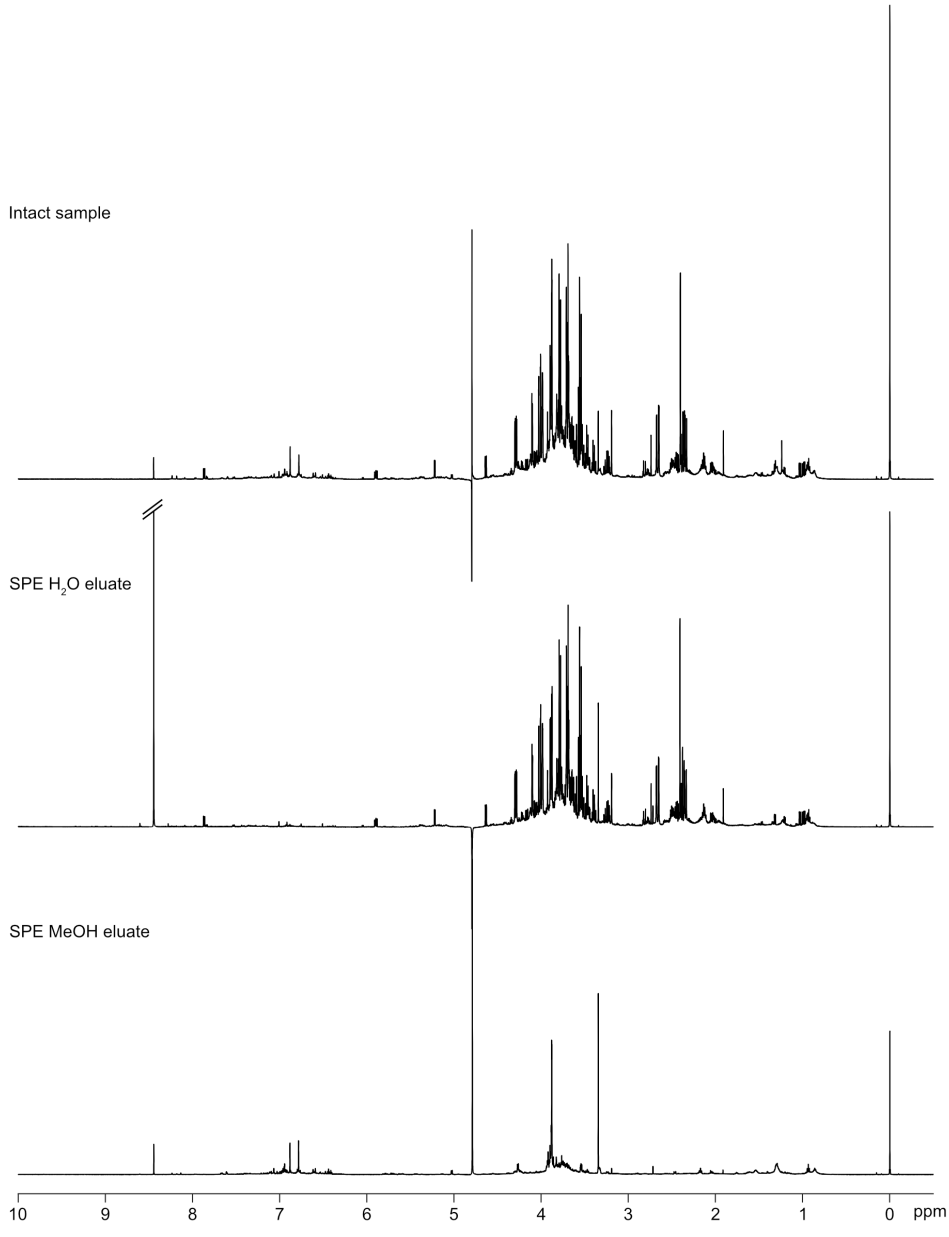


Fig. S3 1D-^{1}H NMR spectra of a pooled root exudate sample either analysed directly (top) or after it was passed through an SPE column (middle and bottom)

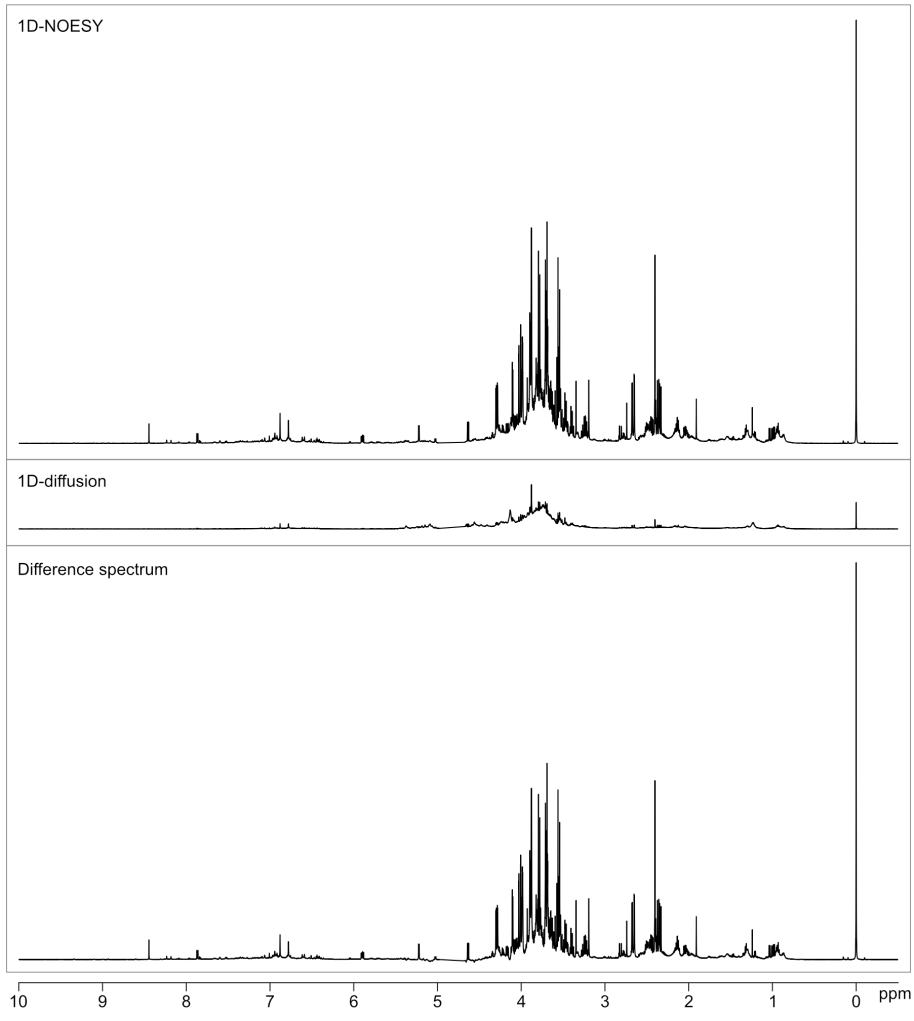


Fig. S4 1D-NOESY-presat, 1D-diffusion, and the difference spectrum (1D-diffusion subtracted from 1D-NOESY), respectively, of a pooled root exudate sample

2.2 Sample preparation and acquisition parameters

Ten of the analysed root exudate samples were pooled together to obtain a large enough sample volume for comparison of different methods. 600 µl of the pooled sample was transferred directly to an NMR tube to be used for 1D-NOESY-presat, CPMG, and diffusion-edited experiments. The remaining sample volume was used to evaluate ultrafiltration and solid phase extraction (SPE). Except where noted, NMR spectra were recorded and processed as described in the Materials and methods section of the paper.

Ultrafiltration was performed in two replicates using Nanosep filters with 3 kDa cut-off (Pall Life Science, Port Washington (NY), USA). The filters were washed nine times with MilliQ water (500 µl, 36 °C, 2000 *g*, 15 min) and then once with D_2O before the sample was added. To each filter, 500 µl of the pooled sample was added and the samples were spun for 10 min at 13 000 *g* and 4 °C, after which an additional volume of 200 µl was added and the centrifugation was repeated until most of the sample had passed through the filter. 600 µl of the filtered samples were then transferred to NMR tubes.

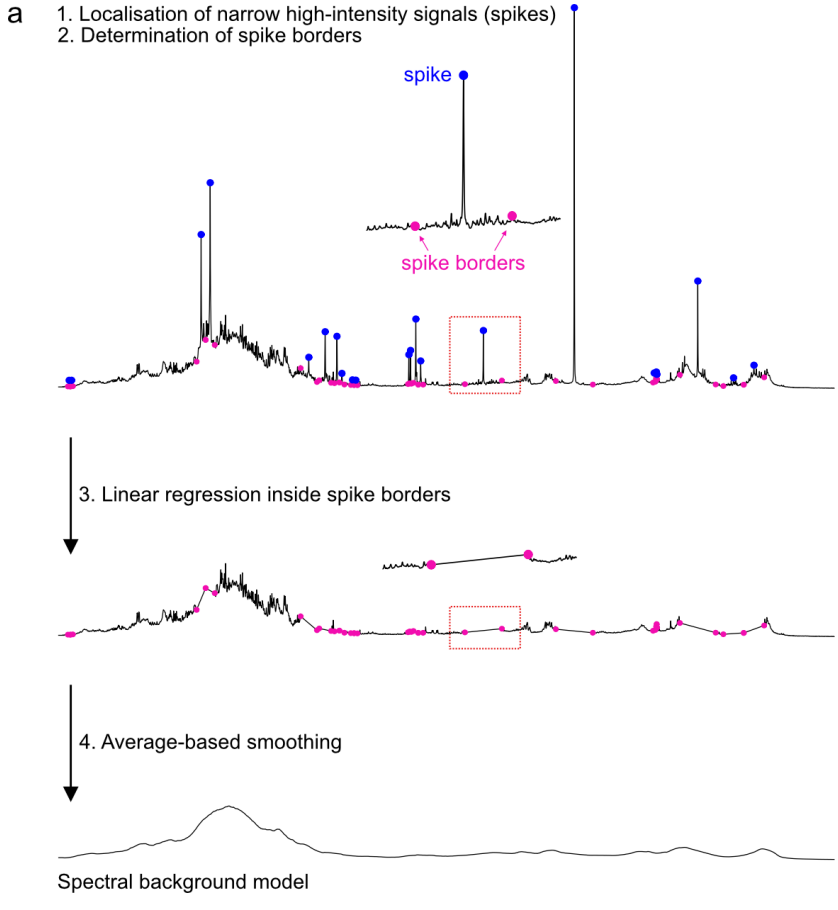
SPE was performed in two replicates using Isolute C18(EC) cartridges (50 mg, 1 ml) (Biotage, Uppsala, Sweden). The cartridges were activated with 1 ml methanol and then washed with 1 ml MilliQ water. Thereafter, 600 µl of the pooled sample was passed through each column followed by 1 ml MilliQ water to wash out all polar metabolites. To elute non-polar metabolites from the SPE columns, 1 ml methanol was added and the eluate was collected separately. Both the aqueous (sample and subsequent MilliQ washing) and the methanolic eluates were dried in a vacuum centrifuge. The dried samples were dissolved in 600 µl D_2O and transferred to NMR tubes.

CPMG spectra were recorded with water presaturation using the Bruker pulse sequence *cpmgpr1d* (relaxation delay-90°-$(\tau$-180°-$\tau)_n$-acquire). The best suppression of broad signals was achieved with a 2 s relaxation delay, n = 450 loops, and a spin echo delay τ of 1 ms, yielding a total echo time of 900 ms.

Diffusion-edited spectra were recorded using a longitudinal-eddy current delay experiment with bipolar gradients (the Bruker pulse sequence *ledbpgp2s1d*). The best suppression of broad signals in the difference spectrum was achieved with a diffusion time of 100 ms and an effective gradient pulse duration (δ) of 2 ms. The gradient strength was set to 95 % of its maximum value (the maximum z-gradient was 48.15 G/cm). 1024 transients were recorded in the diffusion experiment to improve the signal to noise ratio.

3. Simulated spectra

The spectral background of one root exudate spectrum was modelled using average-based smoothing. The procedure is outlined in Fig. S5a. Before applying the smoothing function, narrow high-intensity signals ("spikes") were removed from the spectrum. This was done by defining spike borders - two spectral data points on either side of the spike - and performing linear regression between these two points. Smoothing applied directly to the root exudate spectrum resulted in undesirable artefacts in spike regions (Fig. S5b), which is why spikes were removed before the smoothing step.



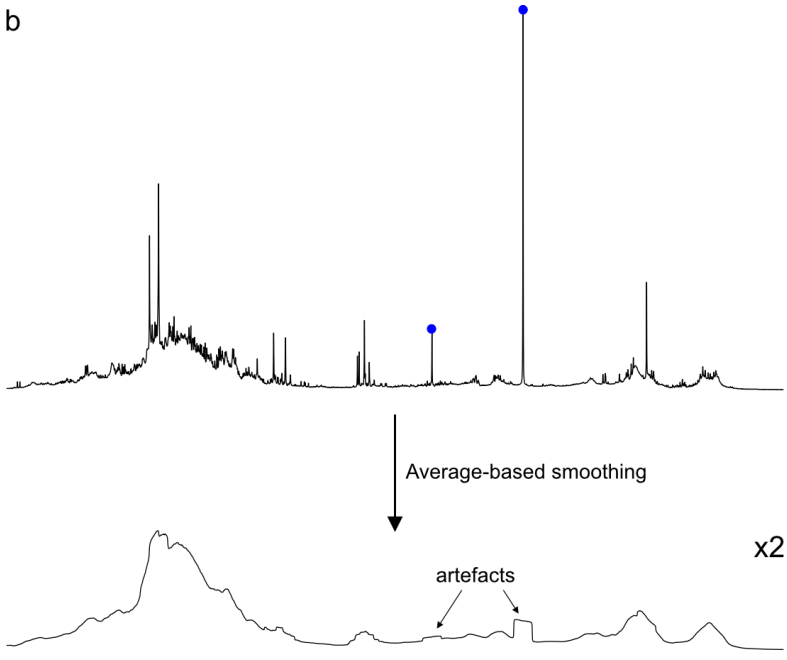


Fig. S5 Modelling the spectral background of a root exudate spectrum. a) The procedure used here to generate the spectral background models (background B is shown), b) Average-based smoothing applied to the root-exudate spectrum with spikes creates artefacts in spike regions (indicated here for two of the spikes).

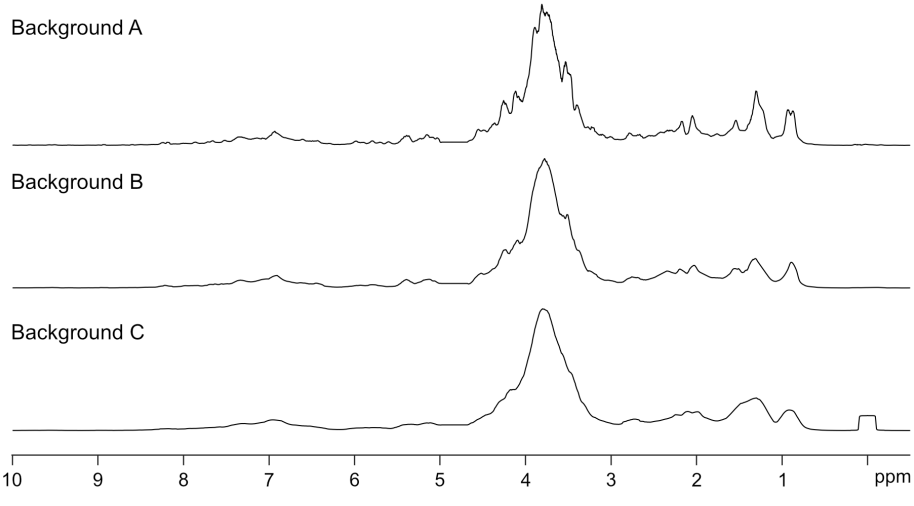


Fig. S6 The three spectral background models included in the simulations

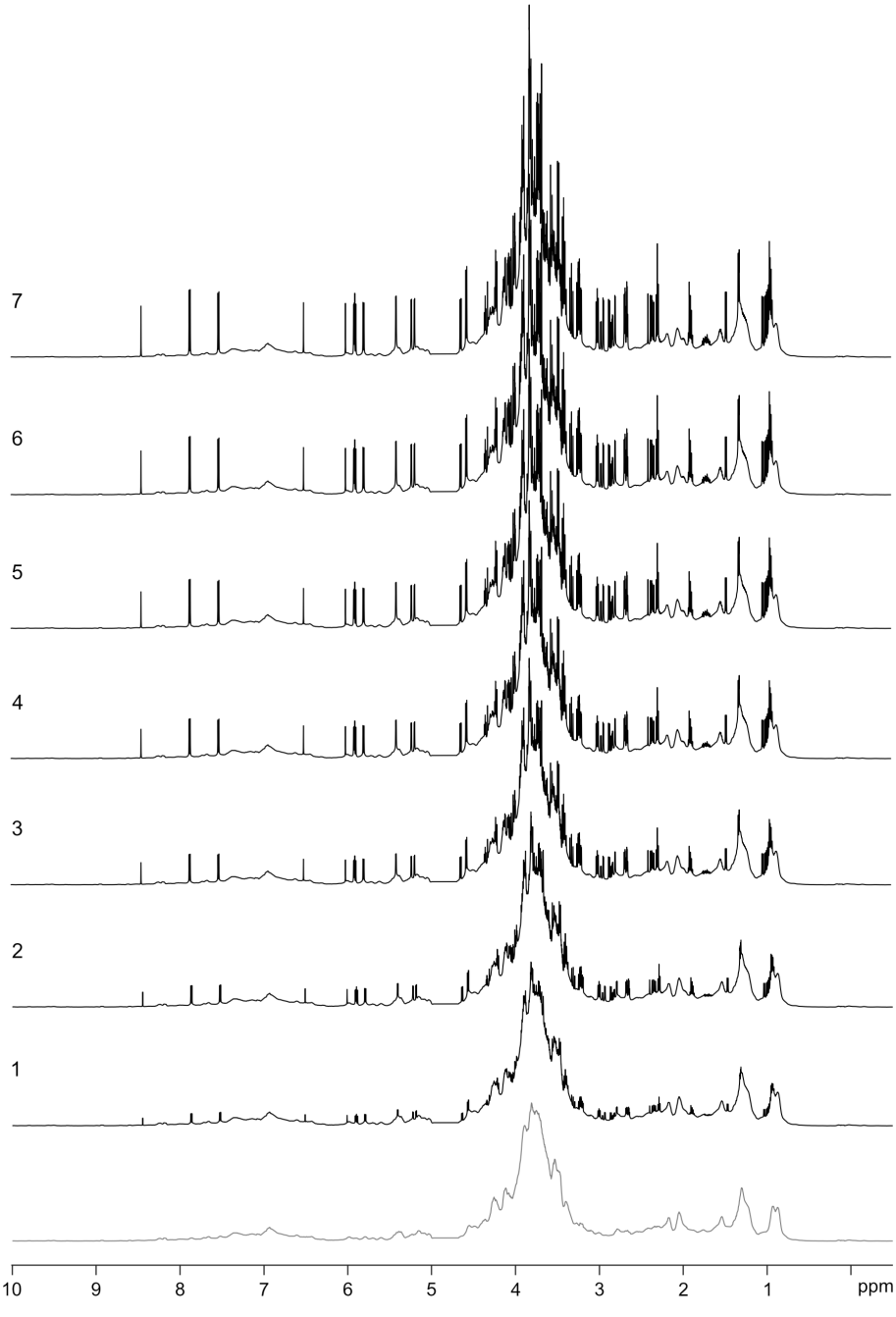


Fig. S7 The seven simulated spectra (black) based on background A (grey)

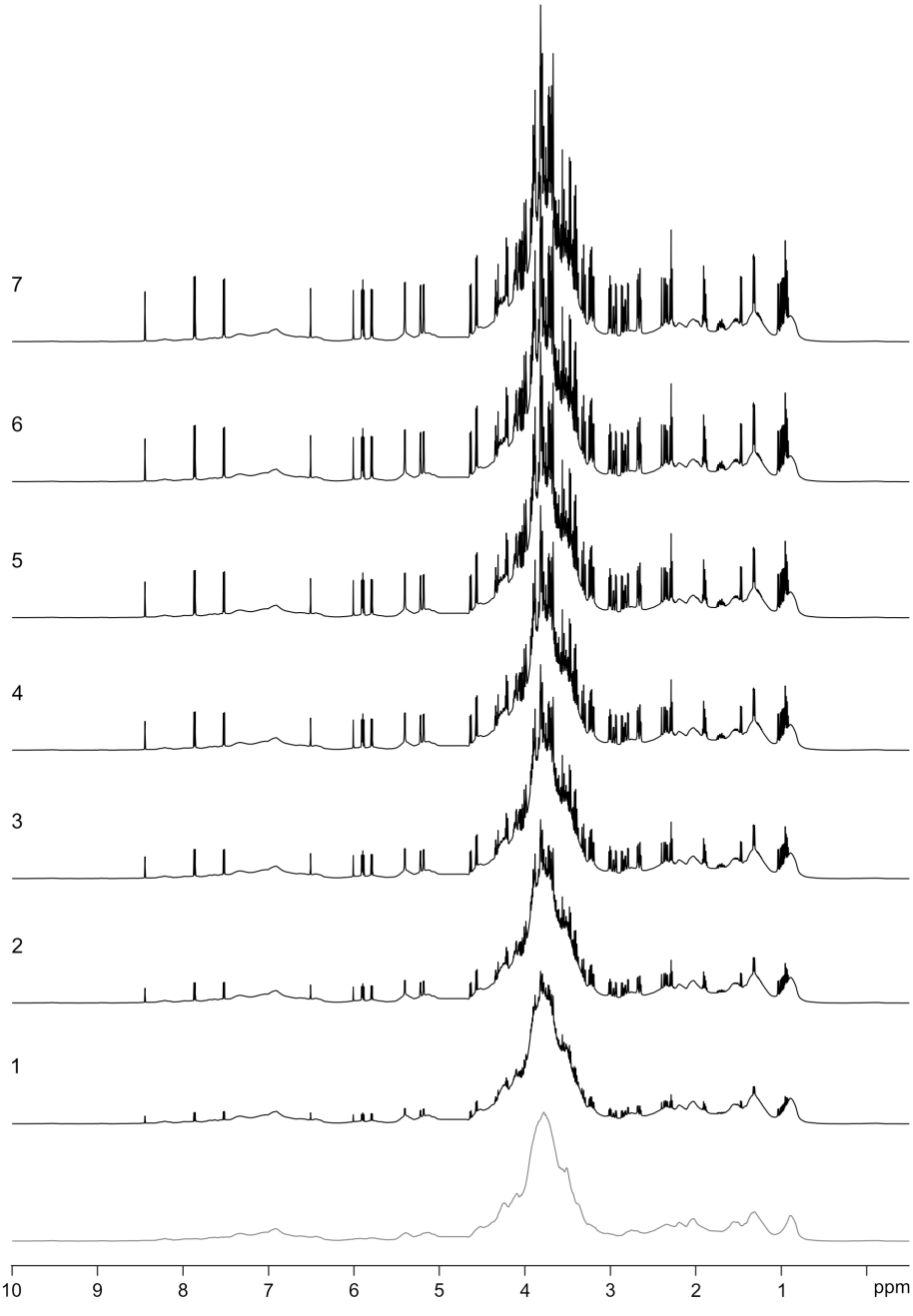


Fig. S8 The seven simulated spectra (black) based on background B (grey)

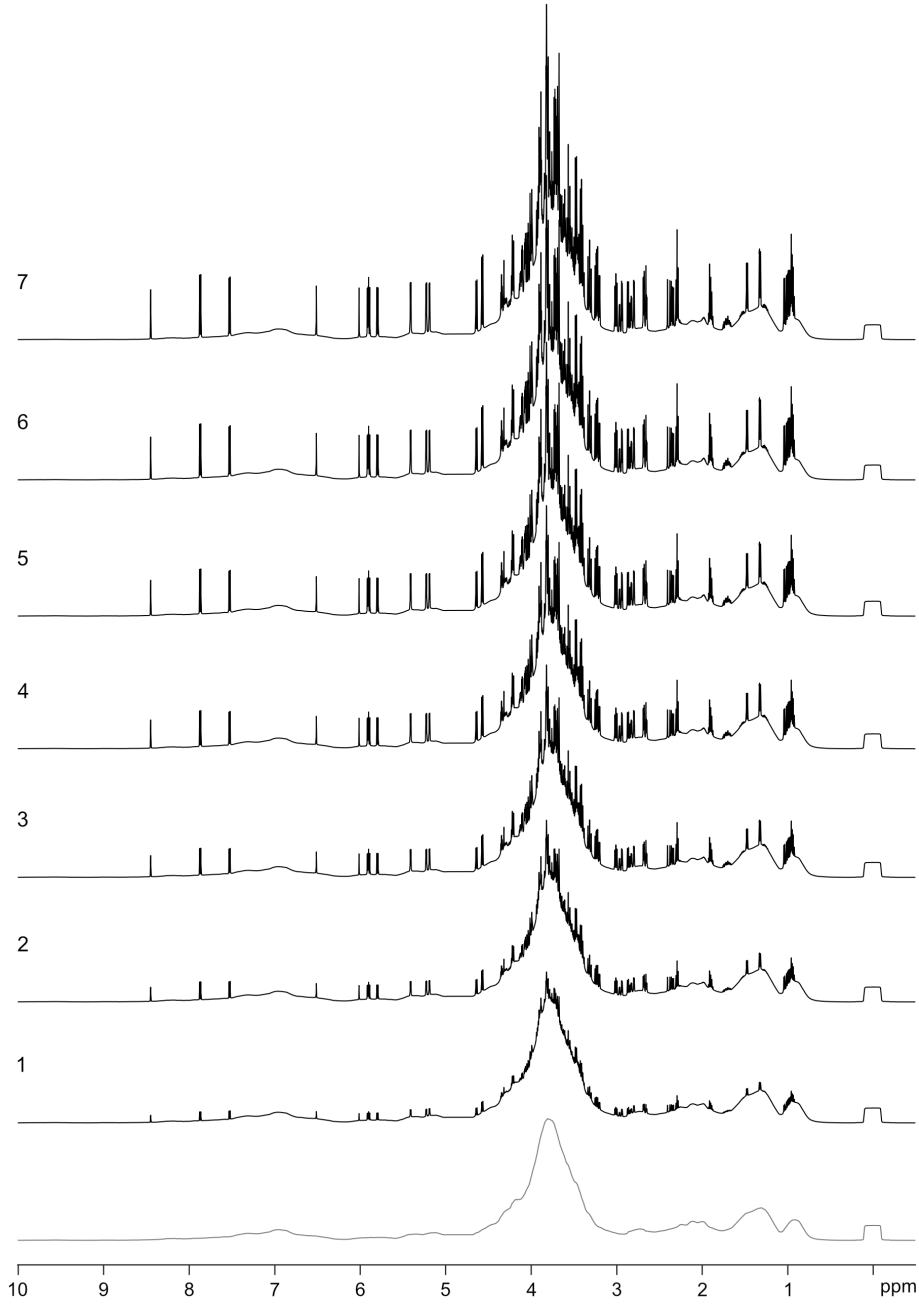
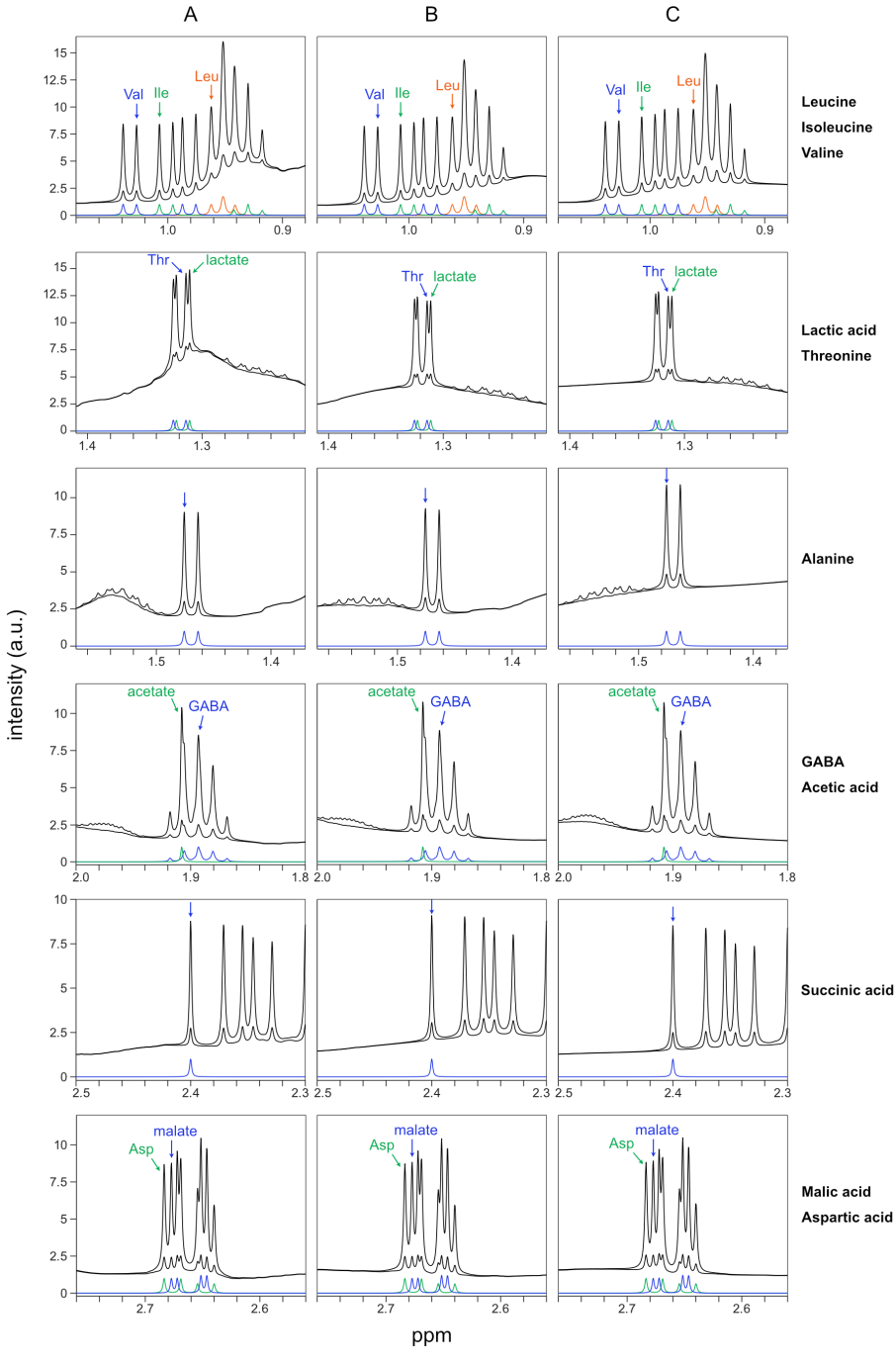
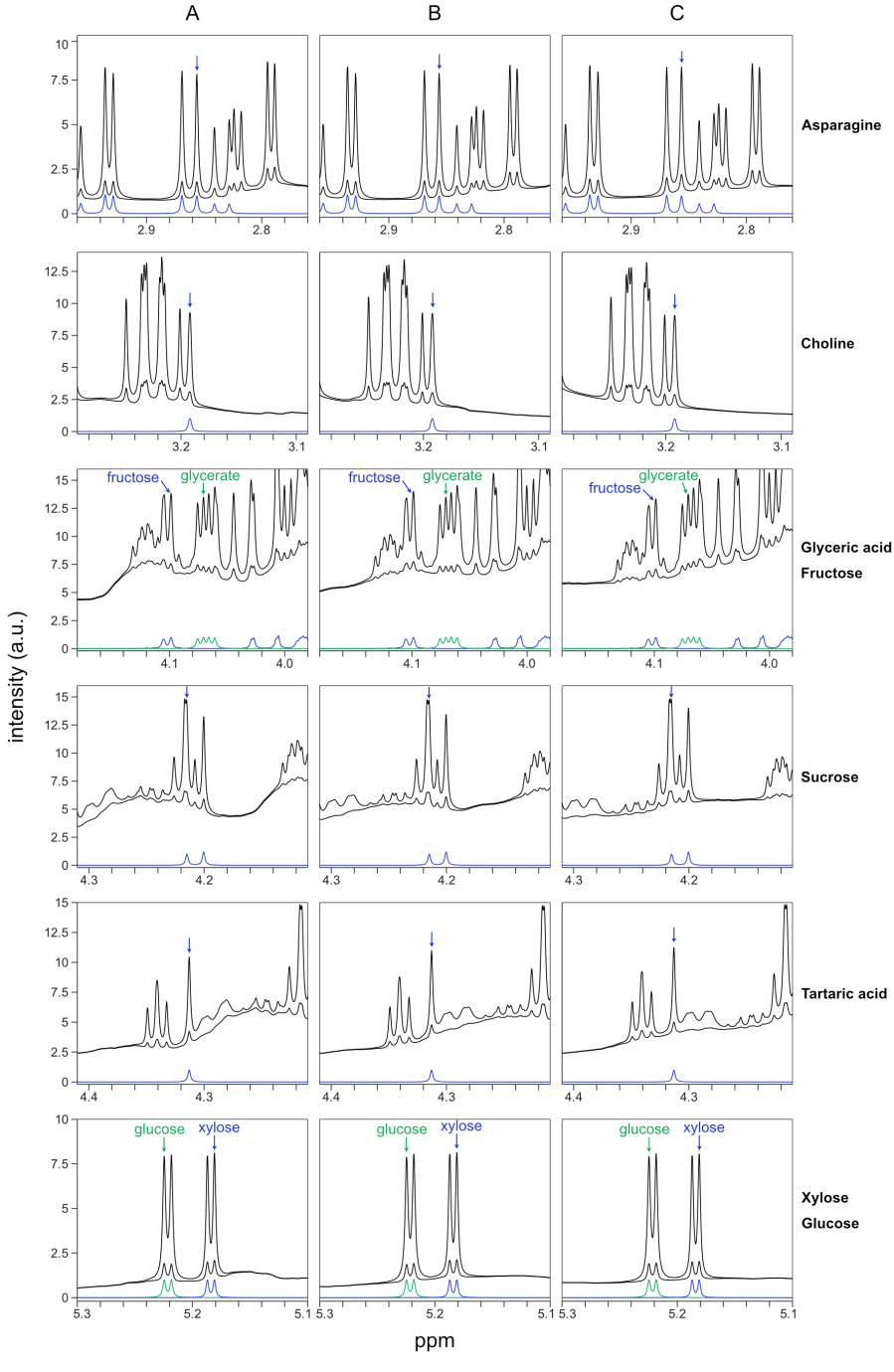


Fig. S9 The seven simulated spectra (black) based on background C (grey)


A
B
C
Val
Ile
Leu
Leucine
Isoleucine
Valine
Thr
lactate
Lactic acid
Threonine
Alanine
acetate
GABA
GABA
Acetic acid
Succinic acid
malate
Asp
Malic acid
Aspartic acid
intensity (a.u.)
ppm


A
B
C
Asparagine
Choline
fructose
glycerate
Glyceric acid
Fructose
Sucrose
Tartaric acid
glucose
xylose
Xylose
Glucose
intensity (a.u.)
ppm

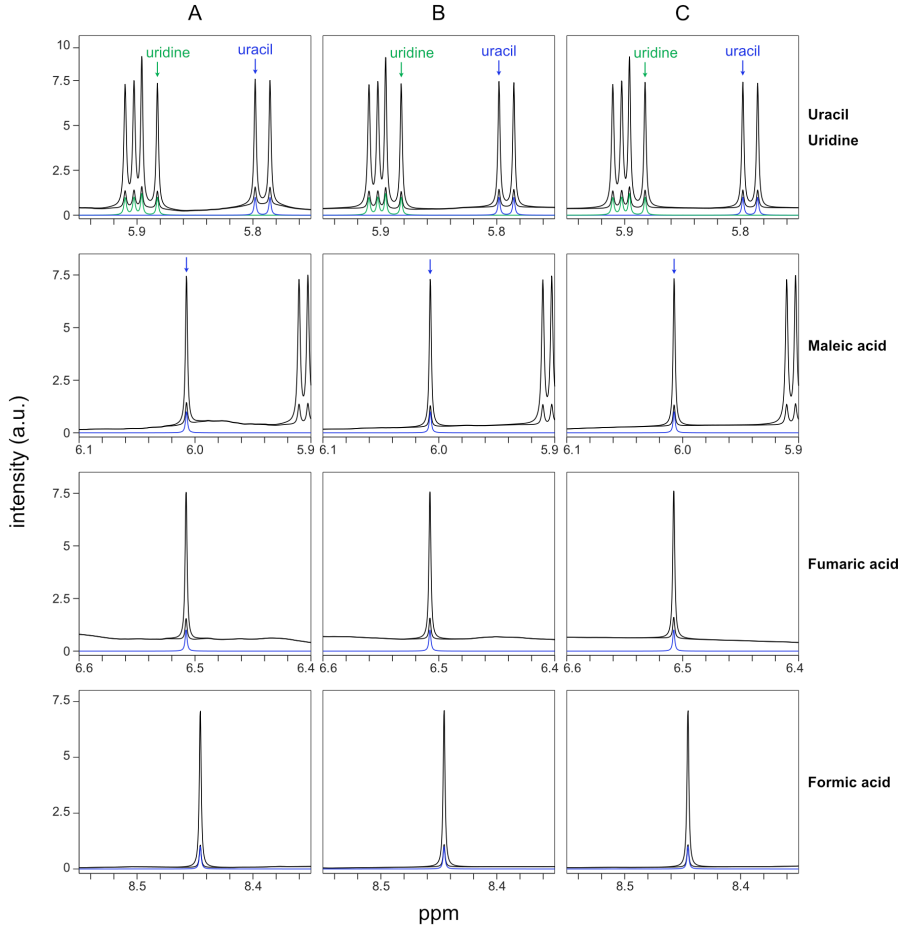


Fig. S10 The metabolite signals that were targeted in the analysis of the simulated spectra. Shown are the uncorrected simulated spectra with the lowest and highest proportion of narrow signal to spectral background for background A, B, and C (i.e. spectra 1 and 7 in Fig. S6-S8) (black). Shown are also the corresponding library spectra (coloured), normalised so that the intensities of the signals used in AQuA equal 1 a.u.

Table S2 The contribution (in percent) of the spectral background to the total target signal intensities in the simulated root exudate NMR spectra[a]

	Spectrum 1			Spectrum 7		
	A	**B**	**C**	**A**	**B**	**C**
Acetic acid	55.0	60.1	59.6	14.9	17.7	17.4
Alanine	66.8	69.3	79.3	22.3	24.4	35.4
Asparagine	44.3	45.9	54.5	10.2	10.8	14.6
Aspartic acid	57.3	58.2	59.7	16.1	16.6	17.5
Choline	67.0	66.0	64.3	22.3	21.7	20.5
Formic acid	6.3	8.7	7.8	1.0	1.3	1.2
Fructose	86.2	86.6	85.3	47.2	47.9	45.3
Fumaric acid	35.2	36.2	37.7	7.2	7.5	8.0
GABA	60.1	64.4	64.3	17.7	20.5	20.5
Glucose	47.8	46.3	47.5	11.6	11.0	11.4
Glyceric acid	86.0	86.0	86.5	46.7	46.7	47.8
Isoleucine	56.1	55.9	66.1	15.4	15.3	21.8
Lactic acid	86.2	78.7	80.3	47.1	34.5	36.8
Leucine	74.0	65.7	72.3	28.8	21.5	27.2
Maleic acid	30.6	22.4	24.4	5.9	4.0	4.4
Malic acid	56.5	57.5	58.9	15.6	16.2	17.0
Succinic acid	63.5	67.2	59.9	19.9	22.6	17.6
Sucrose	79.3	79.1	79.4	35.4	35.0	35.4
Tartaric acid	75.5	78.3	79.4	30.6	34.1	35.5
Threonine	85.6	78.7	80.3	46.0	34.5	36.8
Uracil	36.6	30.7	29.0	7.6	6.0	5.5
Uridine	25.7	24.2	28.5	4.7	4.4	5.4
Valine	55.7	52.5	62.3	15.2	13.6	19.1
Xylose	52.1	52.8	50.7	13.5	13.8	12.8

[a] Estimated by comparing the signal intensities in the simulated root exudate spectra, before correction with the airPLS algorithm, to the intensities in the corresponding narrow signal spectra. Included in the table are estimates for the simulated spectra with the lowest (1) and highest (7) proportion of narrow signals to spectral background, for each background (A, B, and C).

Table S3 Evaluation of the extended AQuA applied to 21 simulated NMR spectra[a]

		A				B				C			
	λ	**slope**	**intercept**	**R^2**	**diff %[b]**	**slope**	**intercept**	**R^2**	**diff %[b]**	**slope**	**intercept**	**R^2**	**diff %[b]**
Acetic acid	10^6	0.9370	-0.0931	0.9997	8.7	0.9017	-0.0087	1.0000	10.1	0.9013	-0.0162	1.0000	10.4
	10^7	0.9885	-0.0108	1.0000	1.5	0.9864	-0.0787	0.9998	3.6	0.9847	-0.1254	0.9995	4.8
	10^8	0.9929	0.0316	1.0000	0.5	0.9926	-0.0103	1.0000	1.0	0.9898	-0.0739	1.0000	3.1
Alanine	10^6	0.9856	-0.0028	0.9999	1.5	0.9686	-0.0560	1.0000	5.2	0.9709	-0.0280	1.0000	4.0
	10^7	0.9864	0.0500	1.0000	1.0	0.9800	-0.0718	1.0000	4.6	0.9872	0.0109	1.0000	1.1
	10^8	1.0029	0.4138	0.9996	15.3	1.0036	0.1328	0.9999	5.3	0.9962	0.0829	0.9995	3.1
Asparagine	10^6	0.9534	-0.0233	1.0000	5.4	0.9452	-0.0287	1.0000	6.6	0.9457	0.0196	1.0000	4.9
	10^7	0.9562	-0.0110	1.0000	4.8	0.9580	-0.0531	1.0000	6.1	0.9802	0.0146	0.9991	2.3
	10^8	0.9876	-0.0558	1.0000	3.3	0.9865	-0.0851	1.0000	4.4	0.9913	0.0974	0.9997	3.1
Aspartic acid	10^6	0.9548	-0.0584	0.9999	6.4	0.9454	-0.0097	1.0000	5.9	0.9443	0.0206	1.0000	5.0
	10^7	0.9787	0.1066	0.9999	1.9	0.9848	-0.0460	0.9995	2.8	0.9750	-0.0085	0.9997	2.5
	10^8	0.9913	0.2224	1.0000	7.0	0.9939	0.0690	1.0000	1.9	0.9876	0.0223	0.9999	1.5
Choline	10^6	0.9322	0.0167	1.0000	6.1	0.9393	-0.0458	1.0000	7.8	0.9389	-0.0204	1.0000	6.9
	10^7	0.9560	0.1190	1.0000	2.5	0.9781	-0.1462	0.9990	6.9	0.9568	-0.0531	0.9996	5.9
	10^8	0.9625	0.4813	0.9992	13.4	0.9935	0.2393	1.0000	8.0	0.9844	-0.0539	0.9999	3.3
Formic acid	10^6	0.9952	-0.0083	1.0000	0.8	0.9917	-0.0037	1.0000	1.0	0.9916	-0.0041	1.0000	1.0
	10^7	0.9978	-0.0028	1.0000	0.3	0.9985	-0.0036	1.0000	0.3	0.9981	-0.0033	1.0000	0.3
	10^8	0.9987	0.0095	1.0000	0.2	0.9992	0.0053	1.0000	0.1	0.9991	0.0027	1.0000	0.1
Fructose	10^6	0.9105	-0.1803	0.9931	14.6	0.8485	0.0302	0.9999	14.3	0.8471	-0.0124	0.9999	15.9
	10^7	0.8125	0.9849	0.9583	20.1	0.9362	0.0799	0.9998	3.7	0.9298	-0.0789	1.0000	9.8
	10^8	0.9223	1.7648	0.9998	55.0	0.9703	1.0166	0.9963	34.3	0.9376	-0.1367	0.9999	10.9
Fumaric acid	10^6	0.9906	-0.0264	1.0000	1.8	0.9908	-0.0058	1.0000	1.2	0.9910	-0.0023	1.0000	1.0
	10^7	0.9980	-0.0203	1.0000	0.9	0.9971	0.0018	1.0000	0.2	0.9977	0.0081	1.0000	0.2
	10^8	0.9979	0.0234	1.0000	0.6	1.0055	-0.0121	1.0000	0.4	0.9992	0.0189	1.0000	0.6
GABA	10^6	0.9128	-0.1571	0.9985	13.8	0.8431	-0.0094	1.0000	16.1	0.8442	-0.0060	1.0000	15.9
	10^7	0.9844	-0.0242	0.9999	2.4	0.9849	-0.1151	0.9992	5.6	0.9805	-0.1246	0.9987	5.9
	10^8	0.9919	0.0617	1.0000	1.5	0.9909	-0.0045	1.0000	1.0	0.9884	-0.0481	1.0000	2.9
Glucose	10^6	0.9798	-0.0242	0.9998	2.6	0.9679	-0.0374	1.0000	4.7	0.9674	-0.0128	1.0000	3.8
	10^7	0.9824	0.0689	1.0000	1.3	0.9829	-0.0465	1.0000	3.4	0.9828	-0.0056	1.0000	1.9
	10^8	0.9986	0.2701	0.9839	9.0	0.9868	0.0194	0.9988	2.1	0.9929	-0.0047	0.9999	1.0
Glyceric acid	10^6	0.8906	-0.0922	0.9997	14.0	0.8795	-0.2530	1.0000	21.2	0.8871	-0.0915	0.9999	14.8
	10^7	0.8247	0.4699	0.9815	14.7	0.8936	-0.3004	0.9998	21.3	0.9065	-0.0224	0.9994	9.7
	10^8	0.9014	1.1563	0.9997	32.0	0.9565	0.3315	0.9978	8.5	0.9174	-0.0180	0.9997	8.6
Isoleucine	10^6	0.9373	-0.0237	0.9998	6.9	0.9204	-0.0064	1.0000	8.2	0.9195	-0.0137	1.0000	8.6
	10^7	0.9225	-0.0142	1.0000	8.1	0.9355	-0.0218	1.0000	7.2	0.9333	-0.0265	1.0000	7.6
	10^8	0.9403	-0.0888	0.9985	9.2	0.9373	-0.0729	1.0000	8.9	0.9536	-0.0051	0.9991	4.4
Lactic acid	10^6	0.9017	0.7359	0.9864	16.8	0.9597	-0.0227	1.0000	4.9	0.9566	-0.0468	1.0000	6.0
	10^7	0.9918	1.4528	0.9983	47.4	0.9950	0.0414	0.9997	1.4	0.9860	-0.0060	0.9999	1.5
	10^8	0.9452	3.5842	0.9950	112.4	1.0034	0.5829	0.9993	20.0	0.9968	0.0978	0.9964	4.4
Leucine	10^6	0.8941	-0.2236	0.9973	17.8	0.8520	-0.0403	0.9998	16.6	0.8528	-0.0446	0.9999	16.6
	10^7	0.8877	0.2504	0.9998	6.2	0.8817	-0.1065	0.9998	15.5	0.8810	-0.0367	0.9997	13.0
	10^8	0.9171	1.1772	0.9815	34.1	0.8833	0.0747	0.9981	8.8	0.9583	-0.0278	0.9962	5.3
Maleic acid	10^6	0.9871	0.0310	1.0000	1.1	0.9894	-0.0013	1.0000	1.1	0.9890	-0.0062	1.0000	1.4
	10^7	1.0003	0.0997	0.9999	3.5	0.9969	-0.0017	1.0000	0.4	0.9972	-0.0091	1.0000	0.6
	10^8	0.9906	0.2288	0.9998	7.5	0.9985	0.0040	1.0000	0.1	1.0004	-0.0011	1.0000	0.1

Table S3 (continued) Evaluation of the extended AQuA applied to 21 simulated spectra[a]

		A				B				C			
	λ	**slope**	**intercept**	**R^2**	**diff %[b]**	**slope**	**intercept**	**R^2**	**diff %[b]**	**slope**	**intercept**	**R^2**	**diff %[b]**
Malic acid	10^6	0.9422	-0.1181	0.9999	9.7	0.9326	0.0038	1.0000	6.8	0.9310	0.0402	0.9999	5.7
	10^7	0.9772	0.0961	0.9999	1.6	0.9840	-0.0629	0.9991	3.4	0.9725	-0.0139	0.9994	2.9
	10^8	0.9912	0.2170	1.0000	6.7	0.9939	0.0745	1.0000	2.0	0.9880	0.0168	0.9999	1.5
Succinic acid	10^6	0.9941	-0.0044	0.9999	0.8	0.9885	-0.0041	1.0000	1.3	0.9888	-0.0041	1.0000	1.3
	10^7	0.9833	0.1030	1.0000	2.2	0.9869	-0.0021	1.0000	1.4	0.9866	0.0024	1.0000	1.3
	10^8	0.9833	0.1013	1.0000	2.1	0.9890	-0.0066	1.0000	1.3	0.9893	-0.0041	1.0000	1.2
Sucrose	10^6	0.9329	-0.0593	0.9967	8.5	0.8868	-0.0314	1.0000	12.3	0.8823	0.1062	0.9992	9.5
	10^7	0.9010	0.9155	0.9810	14.6	0.9910	-0.0534	0.9997	2.2	0.9619	0.1748	0.9996	3.2
	10^8	0.9887	1.2913	0.9999	33.7	0.9796	0.5537	0.9998	12.8	0.9709	0.2819	1.0000	4.8
Tartaric acid	10^6	0.9618	-0.0768	0.9983	6.2	0.9183	0.0690	1.0000	5.8	0.9249	0.0055	1.0000	7.4
	10^7	0.9508	0.0241	0.9982	4.6	0.9567	0.0345	1.0000	3.1	0.9632	0.0305	0.9999	2.8
	10^8	0.9575	0.2397	1.0000	4.7	0.9603	0.2986	0.9993	6.9	0.9779	0.2955	0.9996	8.9
Threonine	10^6	0.9100	0.5333	0.9891	13.0	0.9577	-0.0385	1.0000	5.7	0.9544	-0.0539	0.9999	6.5
	10^7	0.9930	1.2197	0.9984	39.8	0.9947	0.0219	0.9997	1.0	0.9861	-0.0118	0.9999	1.6
	10^8	0.9455	3.2740	0.9951	102.1	1.0035	0.5905	0.9993	20.2	0.9968	0.0827	0.9970	3.9
Uracil	10^6	0.9877	0.0102	0.9999	1.4	0.9669	-0.0115	1.0000	3.8	0.9673	-0.0062	1.0000	3.6
	10^7	0.9972	0.1559	0.9999	5.5	0.9949	-0.0209	1.0000	1.3	0.9947	-0.0006	1.0000	0.5
	10^8	0.9964	0.3030	1.0000	10.8	0.9980	0.0773	1.0000	2.7	0.9979	0.0077	1.0000	0.2
Uridine	10^6	0.9562	-0.0183	1.0000	5.0	0.9493	-0.0047	1.0000	5.3	0.9493	-0.0005	1.0000	5.1
	10^7	0.9932	0.0556	0.9999	1.4	0.9905	-0.0403	1.0000	2.4	0.9889	-0.0227	0.9999	1.8
	10^8	0.9904	0.1235	0.9999	3.6	0.9949	-0.0142	1.0000	1.0	0.9958	0.0082	1.0000	0.4
Valine	10^6	0.9671	-0.0035	1.0000	3.4	0.9614	-0.0210	1.0000	4.8	0.9618	-0.0052	1.0000	4.2
	10^7	0.9593	0.1596	1.0000	3.2	0.9676	-0.0187	1.0000	3.9	0.9650	0.0211	1.0000	2.7
	10^8	0.9527	0.0594	0.9999	3.3	0.9609	-0.0337	1.0000	5.1	0.9633	0.0236	0.9999	2.9
Xylose	10^6	0.9764	-0.0113	1.0000	2.7	0.9737	-0.0153	1.0000	3.3	0.9744	-0.0094	1.0000	3.0
	10^7	0.9777	0.1173	1.0000	2.3	0.9855	-0.0032	1.0000	1.6	0.9838	-0.0047	1.0000	1.8
	10^8	0.9873	0.4527	0.9668	14.4	0.9818	0.1459	0.9938	3.8	0.9894	0.0209	0.9998	1.3
Mean[c]	**10^6**	**0.9496**	**0.1055**	**0.9982**	**7.0**	**0.9366**	**0.0316**	**1.0000**	**7.2**	**0.9367**	**0.0240**	**0.9999**	**6.8**
	10^7	**0.9584**	**0.2722**	**0.9964**	**8.0**	**0.9726**	**0.0572**	**0.9998**	**4.3**	**0.9700**	**0.0340**	**0.9998**	**3.6**
	10^8	**0.9723**	**0.6513**	**0.9966**	**20.1**	**0.9818**	**0.1856**	**0.9993**	**6.7**	**0.9818**	**0.0597**	**0.9994**	**3.2**
Median[c]	**10^6**	**0.9541**	**0.0287**	**0.9999**	**6.2**	**0.9473**	**0.0181**	**1.0000**	**5.7**	**0.9475**	**0.0133**	**1.0000**	**5.4**
	10^7	**0.9806**	**0.1014**	**0.9999**	**2.8**	**0.9848**	**0.0437**	**1.0000**	**3.3**	**0.9817**	**0.0143**	**0.9999**	**2.4**
	10^8	**0.9882**	**0.2342**	**0.9999**	**8.3**	**0.9918**	**0.0746**	**1.0000**	**4.1**	**0.9894**	**0.0230**	**0.9999**	**3.0**

[a] Results from linear regression including seven different intensity scaling levels (1-7 arbitrary intensity units) of the simulated narrow signal spectrum (summed library spectra). The three spectral background models were evaluated separately. Predictor (x-axis): Signal heights in the simulated narrow signal spectra (1, 2, 3, 4, 5, 6, and 7 arbitrary intensity units). Response (y-axis): Signal heights in the simulated spectra with both broad and narrow signals, after correction with the airPLS algorithm. Three λ values (1×10^6, 1×10^7, and 1×10^8) were evaluated.

[b] Average difference (%) between the signal intensities in the simulated narrow signal spectrum and the signal heights in the corresponding simulated spectra with both broad and narrow signals, after correction with the airPLS algorithm. The formula used was $100 \times |\text{Intensity}_{\text{narrow spectrum}} - \text{Intensity}_{\text{simulated spectrum}}| / \text{Intensity}_{\text{narrow spectrum}}$

[c] Mean and median were calculated on the absolute values of the intercepts

4. Spike-in experiment

4.1 Methodology

A spike-in experiment was performed to verify the linearity and accuracy of the method. The metabolites were chosen to represent different chemical classes and their selected target signals have varying coupling patterns and appear in different spectral regions, with or without interference from background or other metabolites (Table S4 and Fig. S10). None of the metabolites were present in the pooled sample before the spike-in experiment, as evaluated using Chenomx Profiler, which means that the calculated concentrations ideally should equal the added amounts. This makes it possible to evaluate how well the proposed method can remove disturbing background while quantitatively retaining the metabolite signals.

Table S4 Summary of the target signal characteristics of the spiked metabolites

	Chemical class	NMR chemical shift (ppm)	Multiplicity	Interference broad signals/ background	Interference other metabolites
Asparagine	amino acid	2.86	dd	no	no
GABA	amino acid	1.89	m	no	yes
Tartaric acid	organic acid	4.31	s	yes	no
Threonine	amino acid	1.31	d	yes	yes
Xylose	sugar	5.18	d	yes	no

The metabolites were added in different concentrations, which is mainly because their target signals have different multiplicities and consist of different numbers of protons (Table S5). For example, the signal of threonine belongs to a doublet containing three protons whilst asparagine's signal is part of a doublet of doublets containing only one proton. This means that the target signal of asparagine is only around 2.5 times more intense than that of threonine, despite the ten times difference in concentration. The concentrations were furthermore chosen so that the target signals of all added metabolites should be over the limit of quantification (10×noise) in each sample.

Table S5 Approximate concentrations (µM) in the samples after spiking with 5 µl of each standard solution

	mult., no of H	Sample A	Sample B	Sample C	Sample D	Sample E
Asparagine	dd, 1H	400	200	1600	800	100
GABA	m, 2H	50	25	200	400	100
Tartaric acid	s, 2H	400	800	50	200	100
Threonine	d, 3H	10	80	40	20	160
Xylose	d, 0.35 H[a]	3200	800	400	200	1600

[a] The target signal belongs to the α-pyranose anomer, which makes up ca 35 % of the total xylose.

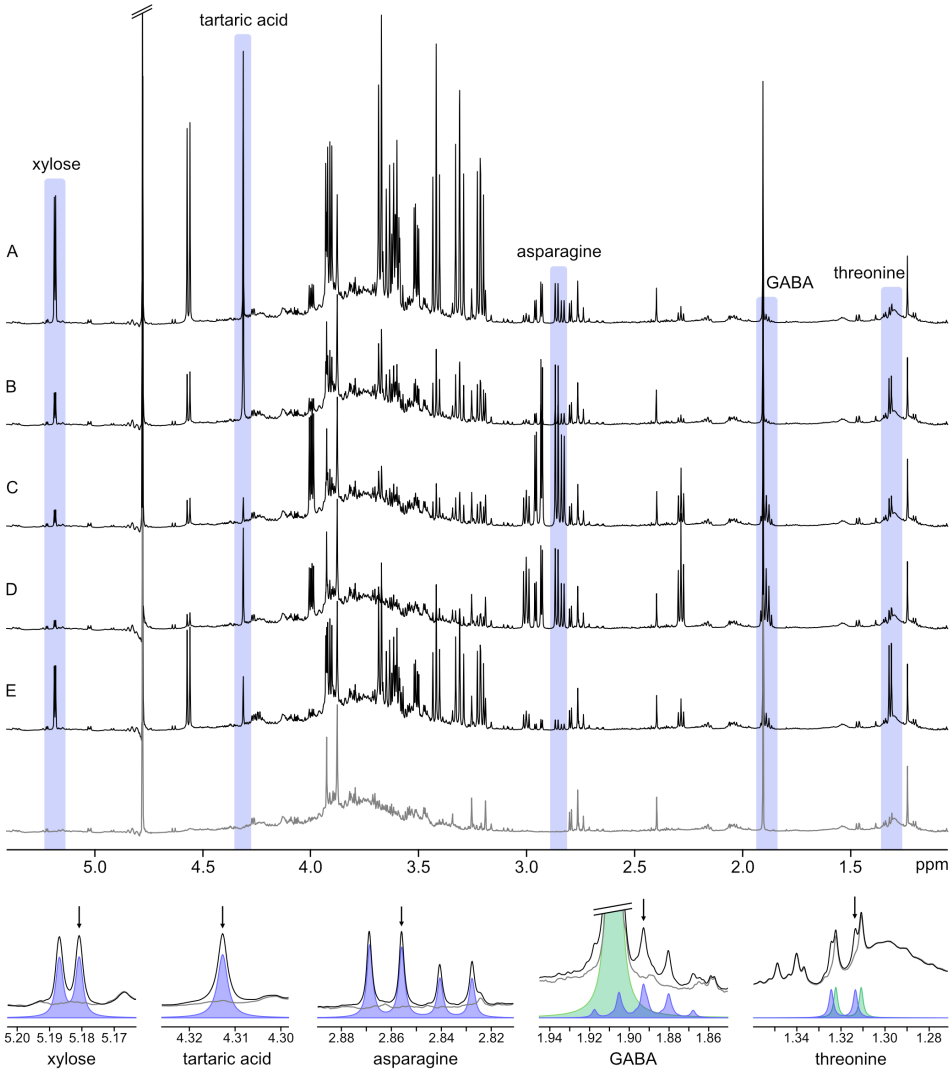


Fig. S11 Upper panel: NMR spectra of the pooled root exudate sample before (grey) and after (black) addition of the spike-in metabolites. Lower panel: Magnification of the target signal regions for the five spiked metabolites. Both the pooled root exudate without addition of spike-in metabolites (grey) and the spiked root exudate with the smallest amount of the respective metabolite added (black) are shown. The corresponding library signals are shown in blue. The target signals used for quantification are marked with arrows. The signals for GABA and threonine overlap with other metabolites (acetic acid and lactic acid, respectively); these are shown in green.

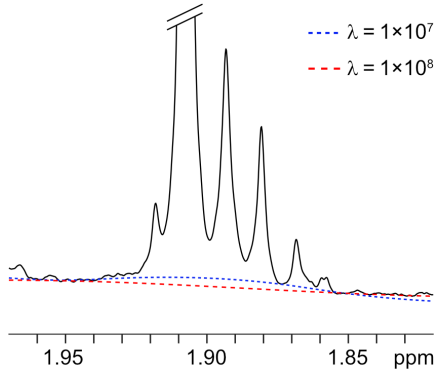


Fig. S12 The effect of two different λ values (1×10^7 and 1×10^8) on the baseline correction around the GABA signal used in AQuA

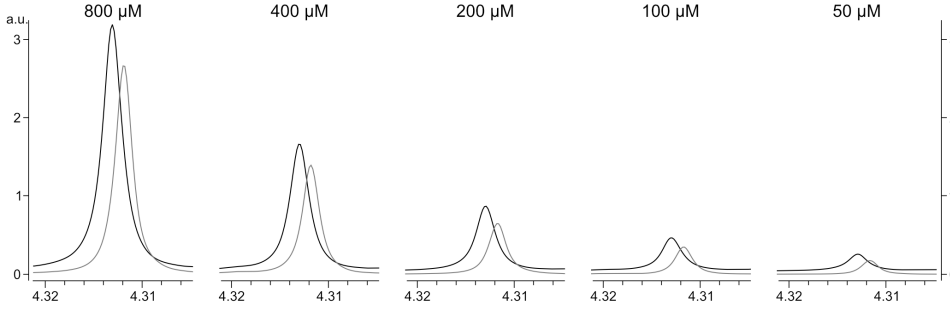


Fig. S13 The signal of tartaric acid in the spectra of the spiked root exudate samples (black) overlaid with the spectra from the corresponding spiked blank samples (grey). The intensities of all spectra have been normalised to the height of the internal standard signal. The pH differed slightly between the root exudates and the blank samples, hence the difference in chemical shift of the tartaric acid signal between the two sample sets.

4.2 Comparison of AQuA results with the actual concentrations

The calculated results from both the spiked root exudate samples and the spiked blank samples (Table S6) were compared with the actual concentrations (Table S7 and S8). Theoretically, the calculated concentrations should equal the actual concentrations for both sample sets, i.e. the slopes should be close to one. Because none of the metabolites was present in the sample to begin with, the intercepts should all be close to zero. In general, the calculated concentrations in the spiked root exudate samples agree better with the calculated concentrations in the corresponding blank samples (Table 1) than with the actual concentrations (Table S7). The difference is smaller for the spiked blank samples; the slopes are generally better when the results are compared with those for the spiked root exudates (Table 1), whereas the R^2 and intercept are comparable between Table 1 and Table S8.

Table S6 Calculated concentrations from the spike-in experiment

		Concentration (μM)			
	Sample	**Actual**	**Spiked sample[a]**	**Spiked blank[b]**	**% difference blank-sample[c]**
Asparagine	A	401	507	522	2.8
	B	1604	2057	2110	2.5
	C	200	251	257	2.4
	D	802	1021	1042	2.1
	E	100	128	133	3.8
GABA	A	50	51	55	8.0
	B	202	206	209	1.7
	C	25	25	27	6.4
	D	403	415	418	0.7
	E	101	104	107	3.3
Tartaric acid	A	400	504	438	15.2
	B	50	68	55	22.7
	C	800	988	845	16.9
	D	200	259	206	25.9
	E	100	131	110	18.7
Threonine	A	10	15	13	15.4
	B	41	56	51	9.0
	C	82	102	99	3.2
	D	20	27	26	2.2
	E	164	200	198	1.4
Xylose	A	3224	3048	3046	0.1
	B	403	398	398	0.1
	C	806	775	774	0.1
	D	202	207	197	5.2
	E	1612	1539	1535	0.2

[a] Concentrations for spiked root exudate samples calculated using an airPLS-extended AQuA including all metabolites listed in Table S1, plus tartaric acid and xylose. Used airPLS parameters: $\lambda_{default}=10^7$, $\lambda_{Thr}=10^5$-10^6, $\lambda_{GABA}=10^8$.

[b] Concentrations for spiked blank samples calculated using an AQuA including only the five spiked metabolites and lactic acid.

[c] Calculated as $100\times|C_{blank}-C_{sample}|/C_{blank}$

Table S7 Comparison of the actual spiked concentrations with the AQuA concentrations obtained for the spiked root exudate samples[a]

Metabolite	Max conc.[b]	R^2	Slope	Intercept	Rel. intercept[c]
Asparagine	1604	1.0000	1.28	-5.71	-0.356 %
GABA	403	1.0000	1.03	-0.868	-0.215 %
Tartaric acid	800	0.9999	1.22	10.2	1.28 %
Threonine	164	0.9994	1.20	3.65	2.23 %
Xylose	3224	1.0000	0.940	18.9	0.586 %

[a] Results from linear regression. Predictor (x-axis): Actual spiked concentrations. Response (y-axis): Concentrations for spiked root exudate samples calculated using an airPLS-extended AQuA including all metabolites listed in Table S1, plus tartaric acid and xylose. Used airPLS parameters: $\lambda_{default}=10^7$, $\lambda_{Thr}=10^5$-10^6, $\lambda_{GABA}=10^8$.

[b] Actual value (µM) for the spiked sample with the highest concentration

[c] Intercept as percent of the maximum concentration for the metabolite

Table S8 Comparison of the actual spiked concentrations with the AQuA concentrations obtained for the spiked blank samples[a]

Metabolite	Max conc.[b]	R^2	Slope	Intercept	Rel. intercept[c]
Asparagine	1604	1.0000	1.32	-5.35	-0.334 %
GABA	403	1.0000	1.03	2.36	0.586 %
Tartaric acid	800	0.9994	1.06	3.23	0.404 %
Threonine	164	0.9999	1.20	1.41	0.864 %
Xylose	3224	1.0000	0.941	14.4	0.447 %

[a] Results from linear regression. Predictor (x-axis): Actual spiked concentrations. Response (y-axis): Concentrations for spiked blank samples calculated using an AQuA including only the five spiked metabolites and lactic acid.

[b] Actual value (µM) for the spiked sample with the highest concentration

[c] Intercept as percent of the maximum concentration for the metabolite

One possible reason for the discrepancy between the calculated and actual concentrations is that minor errors in weighing, dilution, and pipetting will lead to the added concentrations differing slightly from the expected amounts. Another explanation is that the accuracy of the AQuA results depends on the level of agreement between the experimental data and the used database, in this case the Chenomx reference library. Before AQuA can calculate metabolite concentrations based on signal heights, it derives metabolite-specific calibration factors from a database. Any discrepancies between the experimental data and the database will be incorporated into the calibration factors and lead to consistent over- or underestimation of the concentration of affected metabolites. This effect is most easily observed in the spectra of the spiked blank samples. These spectra were quantified with AQuA directly, without any additional baseline correction, which means that slopes deviating from one in Table S8 are a result of either database discrepancies or experimental errors, and not necessarily poor algorithm performance. For asparagine and threonine, the calculated concentrations were

significantly higher than the actual concentrations (Table S8). For neither of these compounds was it possible to achieve a good fit for all signals; if the signals belonging to the alpha protons had been chosen to be used in AQuA, the calculated concentrations would have been lower. One explanation for the discrepancy is that the samples in this study were prepared in D_2O whereas the Chenomx spectral library is optimised for samples containing 90 % H_2O (Fig. S13). In general, the signal line widths in the reference library are wider than in the experimental spectra despite careful calibration of the DSS signal line width. Since AQuA calculates metabolite concentrations only based on signal heights, differences in line width can negatively affect accurate quantification.

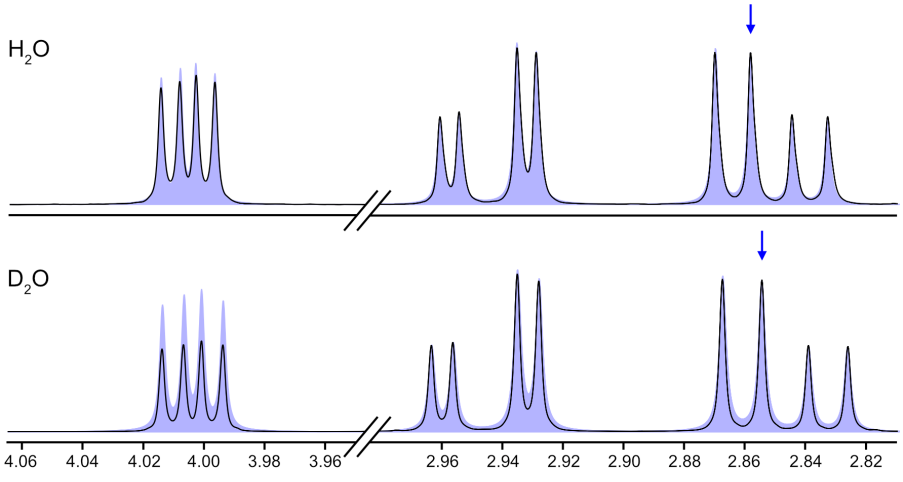


Fig. S14 Experimental spectrum of asparagine (black), dissolved in either 90% H_2O/10% D_2O or 100% D_2O, overlaid with the Chenomx library spectrum (blue) adjusted to the height of the signal that was used in AQuA (marked with arrow). The spectrum recorded in 90 % H_2O displays a better fit with the library spectrum for the alpha proton signal at 4 ppm.

References

Carr, H.Y. & Purcell, E.M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94 (3), 630–638. https://doi.org/10.1103/PhysRev.94.630

de Graaf, R.A., Prinsen, H., Giannini, C., Caprio, S. & Herzog, R.I. (2015). Quantification of ^{1}H NMR spectra from human plasma. *Metabolomics*, 11 (6), 1702–1707. https://doi.org/10.1007/s11306-015-0828-1

Meiboom, S. & Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29 (8), 688–691. https://doi.org/10.1063/1.1716296

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2024:49

In this thesis, two different strategies are investigated for reduction of signal overlap in NMR spectra of complex mixtures of small molecules. A band-selective NMR experiment was devised for suppression of intense, unwanted signals and improved dynamic range in both 1D and 2D NMR spectra. To enable absolute metabolite quantification in complex metabolomics 1D-^{1}H NMR spectra, an automated workflow was developed that computationally removes interference from broad background signals so that metabolites can be quantified.

Elin Alexandersson received her graduate education at the Department of Molecular Sciences, SLU, Uppsala. She received her B.Sc. degree in Molecular Biology at the University of Gothenburg and M.Sc. degree in Chemistry at Umeå University.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-038-5
ISBN (electronic version) 978-91-8046-039-2