Journal of Magnetic Resonance 298 (2019) 35–47

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

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

A field-invariant method for quantitative analysis with benchtop NMR

Yevgen Matviychuk, Jet Yeo, Daniel J. Holland *

University of Canterbury, Private Bag 4800, Cristchurch 8140, New Zealand

ARTICLE INFO

Article history:
Received 23 August 2018
Revised 26 November 2018
Accepted 26 November 2018
Available online 27 November 2018

Keywords:
Quantitative NMR spectroscopy
Medium-field NMR
Benchtop instruments
Mixture analysis

ABSTRACT

Recently developed benchtop instruments have the potential of bringing the benefits of NMR spectroscopy to the wide variety of industrial applications. Unfortunately, their low spectral resolution poses significant challenges for traditional quantification approach. Here we present a novel model-based method designed to overcome these challenges. By defining our models in terms of quantum mechanical properties of the underlying spin system, we make our approach invariant to the spectrometer field strength and especially suitable for analyzing benchtop data. Our experimental results on prepared samples and natural fruit juices confirm the applicability of our method for quantitative analysis of medium-field ^{1}H NMR spectra. The developed method succeeds in accurately separating the spectra of glucose anomers and even monitoring their interconversion in non-deuterated water. Furthermore, the compositions of unbuffered natural fruit juices estimated using data from 43 MHz to 400 MHz spectrometers are in good agreement with each other and with the reference values from nutrition databases.

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a popular non-destructive technique in analytical chemistry where it provides a fast and convenient alternative to standard chemometric methods for mixture analysis [1–4]. Recent developments in compact NMR instrumentation make these techniques readily available for use in industrial applications and promise great benefits for users in diverse fields from agriculture and petrochemistry to forensics and education. Even though state-of-the art benchtop spectrometers deliver field homogeneity and provide functionality of more expensive superconducting magnets, their advance is hindered by the lack of effective algorithmic methods for robust quantitative analysis of lower-field NMR data. The goal of this work is to lift this obstacle by developing a novel model-based quantification method specifically designed for processing benchtop NMR data.

A typical 1D NMR spectrum in the simplest one-pulse experiment exhibits a collection of peaks; different chemical species produce distinct peak patterns owing to differences in magnetic shielding and, as a result, chemical shifts. Their line intensities are proportional to the amounts of related nuclei, and thus the mole fraction of a chemical specie in a mixture, which is the underlying assumption in quantitative NMR. Peak integration is a widely accepted intuitive technique for estimating these intensities, which, however, has notable drawbacks. First, it requires certain data preprocessing, such as phase and baseline correction, which may be cumbersome for large and complex datasets. Second, it is limited by spectral resolution and achieves the best results when separate non-overlapping peaks can be identified for each component. Proton NMR – the most common mode of quantitative experiment due to its high signal to noise ratio and fast acquisition – is notorious for peak splitting of coupled spins and peaks overlap, even in molecules of only moderate complexity (e.g. monosaccharides). These factors are further exacerbated at the lower field strengths of benchtop instruments where the coupling distortions become so severe that they completely change the shapes of the peaks causing multiple resonances to overlap and introducing broad baselines. Consequently, the established peak integration method often becomes inadequate for analysis of complex chemical mixtures with benchtop spectrometers. Although several algorithms have been proposed recently to integrate even overlapping peaks [5–8], accurate quantitative analysis of heavily crowded spectral regions remains a significant challenge (see Fig. 1).

Here we address these issues and develop an effective quantification approach capable of analyzing data at any field strength, including those acquired on benchtop instruments. Instead of peak integration, we base our method on the ideas of parametric modeling. We view each NMR signal as an instance of a model with tunable parameters, which transforms the quantification problem to that of parameter estimation. For example, the sought intensity of a peak can be found by fitting a Lorentzian function with specific position and width to the experimental spectrum – a method that in general is less susceptible to noise and peak overlap than the

* Corresponding author.
E-mail address: daniel.holland@canterbury.ac.nz (D.J. Holland).

https://doi.org/10.1016/j.jmr.2018.11.010
1090-7807/

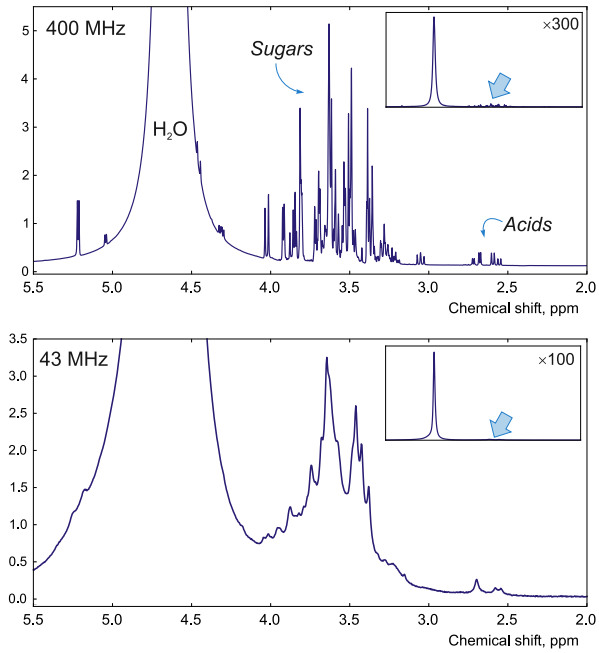

Fig. 1. Examples of ^{1}H spectra of an apple juice sample acquired with high-field superconducting magnet (top) and a benchtop system (bottom). Peaks overlap and baseline distortions prevent accurate quantitative analysis of these data with the traditional peak integration approach.

traditional integration. This or similar principles underly many developed algorithms, such as AMARES [9], QUEST [10], AQSES [11], Indirect Hard Modeling [6], and others [12–15]. Furthermore, the ideas of model-based representation is the core of modern automated techniques and software packages for NMR data analysis, including Global Spectral Deconvolution in Mnova [16], CRAFT [17], BATMAN [18,19], BQuant [20], and Chenomx among the most popular ones. These high-throughput tools have been found especially effective in metabolomics for processing high-resolution spectra of biofluids consisting of hundreds of chemicals. To address the problem of specie identification in complex biological samples, Tiainen et. al. [21] propose to use quantum mechanical (QM) approach to model NMR spectra of separate components based on their available molecular structures. These ideas have been realized in ChemAdder package for analysis of biofluids [22] and form the basis for our method as well. Importantly, QM models are inherently field-invariant; they take into account and concisely describe any higher-order coupling effects often seen in ^{1}H spectra at lower field strength. This allows us to easily generate NMR fingerprints for chemical species and use them to quantitatively analyze even low-resolution benchtop data.

We demonstrate our approach by analyzing the composition of sugar mixtures as food screening is one of the potential niche applications for benchtop NMR systems. Glucose, fructose, and sucrose – the three sugars most commonly occurring in fruit juices, wine, soft drinks, and other beverages – have different chemical properties and sweetness. Their relative concentrations (in addition to acids) largely determine the taste of a product. Thus, knowledge of the exact composition of sugars in the mixture is important in the food industry, where it provides a means for process monitoring (e.g. fermentation), quality control, or even detection of adulteration (e.g. undisclosed sweetening of fruit juices or dilution of them with apple juice [23,24]). High-field NMR spectroscopy has been found particularly effective in addressing these problems, but they are often deemed too challenging for low-resolution benchtop instruments. The presence of multiple isomers of sugar species in relatively low concentrations, whose very complex overlapping spectra are further obstructed by the strong solvent peak are the main obstacles that prevent effective and accurate quantification.

In the next section, we present the main idea of the model-based approach for NMR quantification. We describe the details of the abstract model for an NMR signal of chemical mixtures and derive the least squares solution for the amounts of its constituents. Then we discuss four possible ways for modeling reference signals of mixture components followed by specific examples of models for glucose, fructose, and sucrose in Section 5.1. Finally, the quantification results of gravimetrically prepared mixtures of sugars and samples of several apple, citrus and tropical fruit juices are presented in Section 5.2.

2. Theory

This section provides a theoretical background for our quantification method of chemical mixtures. In this work, the components of the studied mixture are known – the expected chemical species are identified by the user and supplied as input information; their relative concentrations are to be estimated. Instead of relying on integration of well-separated peaks, for each of the K mixture constituents, we define an entire ideal reference NMR signal based on the known molecular structure and information found in spectral databases. We assume that the observed signal is a superposition of these ideal models; the associated weighting coefficients are determined by the amount of each component in the mixture and can be found with the standard regression analysis.

2.1. A mathematical model of FID signals

In our work, we view an ideal NMR signal of a K-component chemical mixture as a superposition of corresponding signature signals $\mathrm{u}_k, k = 1, \ldots, K$, weighted by the amount of the kth chemical in the solution, c_k:

$$\mathbf{x} = e^{i\varphi_0} \sum_{k=1}^{K} c_k \mathrm{u}_k(\theta_k, \tau). \tag{1}$$

Here, u_k may correspond to either free induction decay (FID) signals or the resulting spectra, and thus the above equation provides a generalized view for both time and frequency-domain methods. The sets of model parameters θ_k determine the appearance of each reference signal; they are related to the nature of particular compounds and may include, for example, chemical shifts of the peaks, their relative intensities and widths, as well as values of J-coupling constants if relevant. We provide more detail about defining model signals $\mathrm{u}_k(\theta_k, \tau)$ later. Additionally, the global phase shift φ_0 and the ringdown delay τ bear the meaning of the zero- and first-order phasing terms, respectively. In general, the resulting model $\mathbf{x}$ is a complex-valued vector of length N.

Given an experimental NMR signal $\mathbf{y}$, we aim to estimate a suitable set of parameters by fitting the model to the data in the least-squares sense:

$$\min_{\{\theta_k\}_{k=1}^{K}, \tau, \varphi_0, \mathbf{c}} \left\| \mathbf{y} - \mathbf{Z}\mathbf{c}e^{i\varphi_0} \right\|, \tag{2}$$

where in a compact matrix notation, $\mathbf{Z} \in \mathbb{C}^{N \times K}$ is a model matrix with columns $\mathrm{Z}_{:,k} = \mathrm{u}_k(\theta_k, \tau)$ for $k = 1, \ldots, K$, $\mathbf{c} = [c_1, \ldots, c_K]^{\mathrm{T}}$ is a column vector of intensities, and $\|\cdot\|$ denotes the Euclidean norm. We note that $\mathbf{Z}$ implicitly depends on all θ_k and τ. Given a fixed model matrix $\widehat{\mathbf{Z}}$ evaluated with some values of $\theta_1, \ldots, \theta_K, \varphi_0$, and τ, Eq. (2) leads to the ordinary least squares problem with respect to the intensities c_k. Following derivations in [25], we obtain a real-valued estimate for the vector of component intensities:

$$\widehat{\mathbf{c}} = \mathrm{Re}\left(\widehat{\mathbf{Z}}^{\mathrm{H}}\widehat{\mathbf{Z}}\right)^{-1}\mathrm{Re}\left(\widehat{\mathbf{Z}}^{\mathrm{H}}\mathbf{y}e^{-i\widehat{\varphi}_0}\right), \tag{3}$$

where $[\cdot]^{\mathrm{H}}$ denotes the conjugate transpose of a matrix and $\widehat{\varphi}_0$ is an estimate of the global phase shift [26,27],

$$\widehat{\varphi}_0 = \frac{1}{2}\angle\left\{\left[\widehat{\mathbf{Z}}^{\mathrm{H}}\mathbf{y}\right]^{\mathrm{T}}\left[\widehat{\mathbf{Z}}^{\mathrm{H}}\widehat{\mathbf{Z}}\right]^{-1}\left[\widehat{\mathbf{Z}}^{\mathrm{H}}\mathbf{y}\right]\right\}, \tag{4}$$

where $\angle[\cdot]$ denotes the argument of a complex number.

The results of Eqs. (3) and (4) were obtained under the assumption that suitable values of the underlying model parameters $\{\theta_k\}_{k=1}^{K}$ and τ are known. In practice, they may need to be estimated as minimizers of the original model fitting problem in Eq. (2). After substituting the found estimates $\widehat{\mathbf{c}}$ and $\widehat{\varphi}_0$ and rearranging its terms, this problem reduces to unconstrained maximization of the variable projection functional,

$$\mathcal{G}(\theta_1,\ldots,\theta_K,\tau) = \widehat{\mathbf{c}}^{\mathrm{T}}\left[\widehat{\mathbf{Z}}^{\mathrm{H}}\widehat{\mathbf{Z}}\right]\widehat{\mathbf{c}}. \tag{5}$$

With these assumptions, one may define the estimated intensities with their associated 95% confidence intervals as $c_k \pm 1.96\,\widetilde{c}_k$ for $k = 1,\ldots,K$, where $\widetilde{c}_k$ is the corresponding standard deviation found as:

$$\widetilde{c}_k = \sigma\sqrt{\left[\mathrm{Re}\left(\widehat{\mathbf{Z}}^{\mathrm{H}}\widehat{\mathbf{Z}}\right)^{-1}\right]_{kk}}, \tag{6}$$

and $[\cdot]_{kk}$ denotes the kth diagonal entry of a matrix in the brackets, and σ is the standard deviation of noise.

Finally, we note that the solution presented above can be viewed as a special case in a more general Bayesian framework [28–30,25], which potentially allows us to incorporate prior information about parameters into the model and estimate the uncertainty of found results. Specifically, Eq. (3) defines the maximum likelihood estimator of component intensities under the assumption that the additive noise in $\mathbf{y}$ is circularly symmetric zero-mean Gaussian with variance σ^2. This assumption on the noise distribution is supported by the principle of maximum entropy [28] and the Central Limit Theorem as it is likely to be contributed by a large number of independent additive sources (e.g. thermal noise). Furthermore, since the Gaussian distribution is preserved by the Fourier transform, the same optimality conditions hold for models defined in the time and frequency domains leading to same intensity estimates in both cases.

3. Reference models of mixture components

In this section, we provide more detail regarding the definition of individual signal components u_k in Eq. (1). We start by reviewing strengths and weaknesses of the classical Lorentzian peaks model. We then simplify its parameter space by incorporating information about the underlying molecular structure and effectively wrap the basic superposition model with a higher level of abstraction. Finally, we propose a hierarchical data organization to further reduce the number of free parameters and facilitate the model fitting, which becomes crucial in processing lower field data on benchtop instruments.

3.1. A base model of NMR signals and its limitations

To simplify the discussion, from now on, we drop the index k when referring to signals u_k, but it is assumed that each summand in Eq. (1) is modeled separately with a distinct set of parameters θ_k. Each u_k eventually forms a column of $\mathbf{Z}$ and thus represents an elemental quantified entity; our primary goal is to estimate the corresponding intensity c_k.

In general, a reference component for a given specie u comprises a set of resonances (e.g. Lorentzian peaks) defined by their positions, intensities, and decay rates [29,30,11,31]. Simply by adding more peaks, this base model can potentially be extended to spectra of any complexity, however such a straightforward approach is often impractical. Since each of P peaks is specified separately, the growing dimension of the parameter space rapidly makes model fitting intractable. Indeed, adjusting the model to match the data observed at different pH, concentrations, or other experimental conditions that alter the appearance of the spectrum, may in general require retuning of $3 \cdot P$ parameters.

Further limitations of the base model are revealed when trying to adapt it to different field strengths, particularly lower fields of $B_0 \approx 1$ T common in non-cryogenic benchtop magnets. We note that an observed NMR signal is a result of quantum transitions undergone by the spin system [32]. Specifically, for a system of n coupled spins of quantum number 1/2 (such as protons), there are $P = n \cdot 2^{n-1}$ allowed quantum transitions, but due to resolution limits, fewer peaks are observed in a typical NMR spectrum. In spectra obtained at high fields, multiple close resonances effectively merge into single peaks, which allows us to use an approximate model with much lower number of components. For example, the ^{1}H spectrum of propanol acquired at 400 MHz (see Fig. 2) can be very accurately modeled as a combination of only $P = 12$ aggregate peaks instead of all $7 \cdot 2^6 = 448$ possible transitions.

With decreasing field strength, chemical shift differences approach values of coupling constants causing groups of transition peaks to split; this completely destroys the familiar appearance of ^{1}H multiplet patterns as evident in the bottom panel of Fig. 2. Nevertheless, one may still view such spectra as collections of many overlapping peaks, as done, for example, in indirect hard modeling [5,6]. These aggregate peaks however can rarely be modeled with

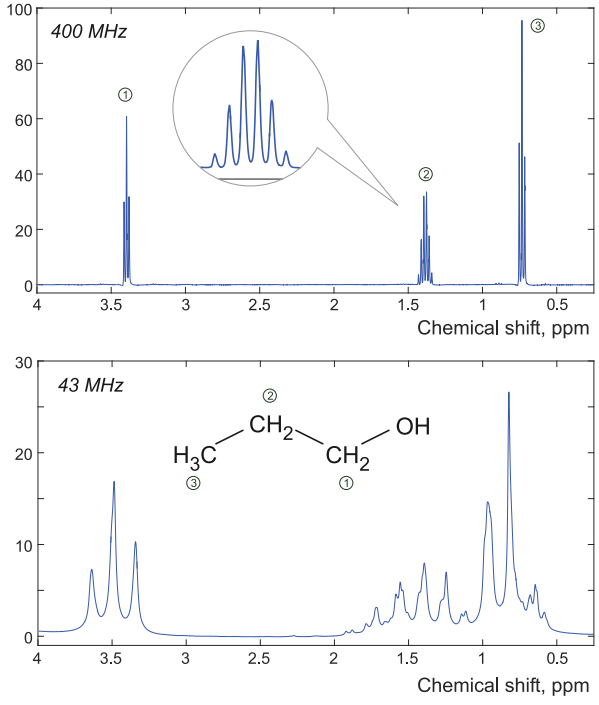

Fig. 2. ^{1}H NMR spectra of propanol acquired at two different field strength. Note the severe distortions of multiplets related to coupled nuclei observed on a benchtop instrument (bottom). Even though its complexity increases, the spectrum is uniquely determined by the same three chemical shifts and two coupling constants.

simple Lorentzian functions, and their varying shapes need to be taken into account, which further inflates the parameter space. The relation of the newly arising resonances to the underlying molecular structure is far less intuitive making it harder, or impossible, to explicitly assign any meaning to the parameters. At high field too, this task rapidly becomes cumbersome even for moderately large molecules (e.g. sugars). We address these issues in the next subsection with a higher-level quantum mechanical model formulation.

3.2. A quantum mechanical model of a specie

Even though an NMR spectrum of a chemical specie is not fixed – its peaks can move as a response to changing experimental conditions – it is necessarily determined by the underlying molecular structure. Specifically, all P transition peaks are defined by quantum mechanical (QM) properties of the spin system, and in general can be expressed as some non-explicit function:

$$\{\omega_p, b_p, \alpha_p\}_{p=1}^{P} = f^{\mathrm{QM}}(\boldsymbol{\delta}, \boldsymbol{J}, \boldsymbol{r}). \tag{7}$$

This function takes into account shielding effects experienced by different nuclei, and as a result – their individual chemical shifts $\boldsymbol{\delta}$, the set of mutual J-coupling constants $\boldsymbol{J}$, and possibly a relaxation model with rates $\boldsymbol{r}$ that defines the resulting peak widths [21,33,34]; here, as earlier, bold symbols denote *arrays* of parameters, with each entry corresponding to a specific nucleus in $\boldsymbol{\delta}$ and $\mathbf{r}$, or pairs of nuclei in $\boldsymbol{J}$. In this work, we use a simple isotropic damping model for relaxation and assume that all nuclei in the same spin system relax with equal exponential rates leading to peaks of the same width.

Typically, due to splitting of coupled resonances, the number of peaks observed in the spectrum is much higher than the number of spin chemical shifts $\boldsymbol{\delta}$ and J-coupling constants. Thus, adopting the QM formulation allows us to reduce the number of free parameters to fit. Furthermore, since a QM model does not make any assumptions about the experimental conditions but rather describes a molecule itself, the same specification of a chemical specie is suitable for modeling NMR experiments with any types of pulse sequences run at any field strength. In the above example, the locations and intensities of the twelve peaks of propanol can be uniquely determined by only three chemical shifts and two coupling constants; the same set of five parameters determine the more complicated spectrum in the bottom panel in Fig. 2, which can be obtained simply by changing the field strength in the QM model f^{QM} of Eq. (7). This field-invariance is crucial in our method for processing benchtop data.

In general, evaluating f^{QM} requires diagonalization of the Hamiltonian operator of the spin system; further detail of its implementation can be found in Appendix A and in [35,36]. While the applicability of this direct approach is limited to smaller spin systems of no more than 12–13 coupled spins, fast temporal propagation algorithms have been developed recently to overcome the computational hurdle [34,37,38]. Given values of chemical shifts, coupling constants, and a relaxation model, they gradually evolve the resulting FID signal in time and effectively realize a function $\mathrm{u} = f^{\mathrm{tQM}}(\boldsymbol{\delta}, \boldsymbol{J}, \boldsymbol{r})$ instead. Such techniques are capable of handling much larger spin systems (more than 40 spins [37], or even 200 spins if state-space approximations are used [33]), however we found their running times suboptimal when working with smaller molecules (e.g. sugars), in which cases the diagonalization method is preferred. Furthermore, in contrast to Eq. (7), temporal propagation methods do not explicitly provide parameters of separate transition peaks, $\{\omega_p, b_p, \alpha_p\}$, which can be valuable in certain applications.

Finally, it is worth mentioning that experimentally obtained spectra of pure components can be used as reference signals u as well [39]. This approach is especially appealing for modeling very large molecules with many overlapping peaks in their spectra, whose complexity prevents quantum mechanical simulation or for specifying chemical species with unknown structure. Unfortunately, even if the spectra of all pure components are available, they are only applicable for analyzing data acquired at the same field strength and do not easily accommodate slight changes in the spectra due to variations in experimental conditions. Extensive preprocessing needed to compile such a reference library (mainly denoising, baseline, phase correction, and suppression of solvent peaks) further complicates this approach, especially on low-field instruments. We expect the resulting quantification accuracy to be lower in this case, and thus the more flexible parametric modeling is preferable.

We summarize the discussed approaches for modeling reference signals of individual chemical species in Table 1 along with the associated parameters. The resulting signals u eventually become the columns of the model matrix $\mathbf{Z}$ to represent separate quantified constituents of an analyzed chemical mixture.

3.3. A unified model of a mixture component

Undoubtedly, accurately matching reference models with measured data is the most computationally demanding step of the entire quantification procedure: minimizing the variable projection functional $\mathcal{G}$ given by Eq. (5) entails solving the complex multidimensional non-convex optimization problem. For example, to properly fit a QM model of an isomer of glucose, one needs to estimate seven chemical shifts and possibly also seven J-coupling values; other monosaccharides require finding similar numbers of parameters. Even though good initial values are usually available in numerous NMR databases or as a result of previous experiments, they may require adjustment to account for possible variations of experimental conditions in each particular case. When multiple resonances overlap in a small spectral region, simultaneous fitting of all model parameters becomes practically impossible if the spec-

Table 1
Different methods of generating reference model signals u.

Input parameters	Modeled FID
1. Base model	
– global shift $\widetilde{\omega}$ and rate $\widetilde{\alpha}$	$\widetilde{\mathrm{u}} = e^{\left[i\widetilde{\omega}-\widetilde{\alpha}\right]t}$
– peaks positions $\boldsymbol{\omega}$	$\mathrm{u} = \widetilde{\mathrm{u}}\sum_p b_p e^{\left[i\omega_p-\alpha_p\right]\mathbf{t}}$
– peaks intensities $\boldsymbol{b}$	
– relaxation rates $\boldsymbol{\alpha}$	
2. QM model (diagonalization)	
– global shift $\widetilde{\omega}$ and rate $\widetilde{\alpha}$	$\widetilde{\mathrm{u}} = e^{\left[i\widetilde{\omega}-\widetilde{\alpha}\right]t}$
– chemical shifts $\boldsymbol{\delta}$	$\boldsymbol{\omega}, \boldsymbol{b}, \boldsymbol{\alpha} = f^{\mathrm{QM}}(\boldsymbol{\delta}, \boldsymbol{J}, \boldsymbol{r})$
– coupling constants $\boldsymbol{J}$	$\mathrm{u} = \widetilde{\mathrm{u}}\sum_p b_p e^{\left[i\omega_p-\alpha_p\right]\mathbf{t}}$
– relaxation rates $\boldsymbol{r}$	
3. QM model (temporal propagation)	
– global shift $\widetilde{\omega}$ and rate $\widetilde{\alpha}$	$\widetilde{\mathrm{u}} = e^{\left[i\widetilde{\omega}-\widetilde{\alpha}\right]t}$
– chemical shifts $\boldsymbol{\delta}$	$\mathrm{u} = \widetilde{\mathrm{u}} f^{\mathrm{tQM}}(\boldsymbol{\delta}, \boldsymbol{J}, \boldsymbol{r})$
– coupling constants $\boldsymbol{J}$	
– relaxation rates $\boldsymbol{r}$	
4. Experimental data	
– global shift $\widetilde{\omega}$ and rate $\widetilde{\alpha}$	$\widetilde{\mathrm{u}} = e^{\left[i\widetilde{\omega}-\widetilde{\alpha}\right]t}$
	$\mathrm{u} = \widetilde{\mathrm{u}} f^{\mathrm{exp}}(\cdot)$
5. Group of species in fixed proportion	
– global shift $\widetilde{\omega}$ and rate $\widetilde{\alpha}$	$\widetilde{\mathrm{u}} = e^{\left[i\widetilde{\omega}-\widetilde{\alpha}\right]t}$
	$\mathrm{u} = \widetilde{\mathrm{u}}\sum_j c_j \mathrm{u}_j$

tral resolution is insufficient (e.g. when analyzing sugar mixtures with benchtop instruments).

Fortunately, we observe that, at least for some species, modeling deviations in individual peaks *independently* is often not necessary. Our analysis of sugar solutions with pH ranging from 1.2 to 10 (see Section 5.1.2), as well as multiple experiments with natural fruit juices, reveal that peaks of glucose, fructose, and sucrose do not move notably relative to each other. However, their location *as a group* on the chemical shift scale may change (particularly when there is no unambiguous reference to set the 0 ppm mark). This suggests that when fitting signal models to datasets of certain chemical mixtures, to account for the move of the entire group of peaks, one may simply offset all chemical shifts in the QM model by the same amount $\widetilde{\omega}$. Since such shift deviations are negligible comparing to the absolute resonance frequencies, they do not affect corresponding coupling patterns. Therefore, computationally demanding QM simulations can be performed only once for some values in the middle of the range, and then the entire resulting spectrum is shifted accordingly to fit the observations. Furthermore, line-broadening caused by the loss of magnetic field homogeneity is likely to affect all peaks in the spectrum as well, and thus can be taken into account by controlling the widths of all peaks simultaneously with a single higher-level parameter. If a reference model is generated with any of the methods discussed above, moving its peaks by $\widetilde{\omega}$ and applying Lorentzian broadening of rate $\widetilde{\alpha}$ is easily achieved by multiplying the modeled FID with a decaying harmonic signal, $\widetilde{u} = e^{\left[i\widetilde{\omega}-\widetilde{\alpha}\right]t}$.

By extension, one may even combine models of separate chemicals and adjust them simultaneously if desired (see item 5 in Table 1). For example, if in a measured spectrum, all peaks of both anomers of glucose are shifted by the same amount with respect to their database values, this deviation is corrected by adjusting a single parameter $\widetilde{\omega}$ instead of fourteen chemical shifts. Such aggregation tremendously reduces the effective dimension of the parameter space and facilitates model fitting. Nevertheless, we note that other molecules or their specific functional groups may show higher deviation of chemical shifts and *J*-couplings as a response to changing conditions. Thus, we reserve the option to fine tune these parameters individually, if necessary, after suitable values of the global $\widetilde{\omega}$ and $\widetilde{\alpha}$ have been found. Such adjustments are useful, for example, when analyzing the content of organic acids using high resolution spectra of fruit juices (see Section 5.2.3).

To summarize, there are at least four distinct ways to specify a chemical specie as a model NMR signal dependent on a set of input parameters. Importantly, all these types of reference signals are equivalent from the point of view of the model for a chemical mixture given by Eq. (1); the choice of a suitable representation mainly depends on the nature of a particular chemical specie and a given application.

4. Materials and methods

In this work, we use samples of two kinds: controlled mixtures of sugars and samples of natural fruit juices. Solid mixtures of glucose, fructose, and sucrose were prepared gravimetrically and dissolved in 100% deionized H_2O. Glucose and sucrose were purchased from Thermo Fisher Scientific Australia/New Zealand, and fructose was purchased from Ajax Finechem; the purities of components specified by the manufacturers are $> 99.0\%$. Mettler Toledo AX205 balance with instrument accuracy of 0.1 mg (provided in the calibration protocol of the manufacturer) was used to prepare ≈4–5 g of each solid mixture; water was added to obtain 15 mL of the resulting solutions (which, depending on the ratio of sugars, corresponds to molarity of ≈ 1 M).

Fruit juice samples were prepared fresh and filtered using a Millex syringe filter unit with 0.22 μm pore size to reduce turbidity. No pasteurization or pH equilibration was performed to avoid altering the natural composition of the samples. For the same reason, no deuterated water was added to the samples.

The measurements were performed on two NMR spectrometers: a medium-field Magritek Spinsolve benchtop system operating at a ^{1}H frequency of 43.6 MHz and on a 400 MHz Agilent 400MR spectrometer equipped with a OneNMR probe. On the benchtop system, ^{1}H FID signals were acquired with 32768 points and dwell time of 200 μs using a one-pulse sequence with pulse angle of 90°; datasets with variable numbers of scans were acquired with repetition time of up to 60 s. Likewise, on the high-field instrument, we acquired 16384 time points with dwell time of 312.5 μs and pulse angle of 45°. Furthermore, a coaxial capillary insert containing D_2O and DSS was used to establish a lock signal for the high-field spectrometer and provide a chemical shift reference.

Our proposed method, including the evaluation and fitting of QM models, was implemented in Python 3.5. Model parameters were optimized using the SciPy implementation of the L-BFGS-B algorithm with basin-hopping [40]. Given found estimates of components' intensities $\widehat{c}_k$, we compute and report their respective mole fractions as $\widehat{\chi}_k = \frac{\widehat{c}_k}{\sum_k \widehat{c}_k}$. Comparative quantification of high-field datasets of controlled mixtures was performed in Mest ReNova (Mnova) ver. 12.0.2. The global automatic phase correction and multi-point baseline correction algorithms were used followed by manual adjustment where necessary. The amounts of glucose and sucrose were determined by integration of peaks of the anomeric protons (located at 5.2 ppm and 5.4 ppm respectively); the amount of fructose was calculated as a difference based on the integral in the 3.15–4.25 ppm range.

5. Results and discussion

5.1. Quantitative analysis of sugar mixtures

To demonstrate the practicality of our proposed model-based approach, we perform quantitative analysis of glucose, fructose, and sucrose mixtures in water. High-field NMR spectroscopy has been proven an excellent technique for analyzing compositions of various food samples [1,2] as well as in numerous conformational studies of carbohydrates [41–43]. However, relatively high complexity of NMR spectra of most sugars largely precludes the traditional peak integration analysis especially of data observed on medium-field benchtop instruments. Moreover, the water peak located closely to the spectra of sugars significantly distorts the baseline and further complicates the analysis. Since the proposed model-based approach is less sensitive to peak overlaps, this important problem is a perfect case study to test our method.

In addition to the relatively high number of coupled protons (6–7 in most monosaccharides), another complicating consideration is that reducing sugars, of which glucose and fructose are two common examples, exist in solutions in several tautomeric forms. For example, α- and β-glucopyranose, under normal conditions account for approximately 37.5% and 62.5% of glucose in aqueous solutions, respectively [42]. For fructose, four tautomers, namely α- and β-fructopyranose and α- and β-fructofuranose, can be observed in solution in considerable amounts (2.67%, 68.23%, 6.24%, 22.35%, respectively in tautomeric equilibrium at 20 °C [43]). We disregard the intermediate open aldehyde form, which typically accounts for only 0.05% of fructose in solution. The tautomeric differences in the molecular structure affect magnetic shielding of even distant protons and thus alter their

chemical shifts and *J*-coupling constants. As a result, these different isomers are distinguished by NMR spectroscopy and thus need to be modeled separately with their own reference models.

In what follows, we briefly describe the specifics of our proposed algorithmic approach for quantification of sugar mixtures and build reference models for isomers of glucose, fructose, and sucrose.

5.1.1. Implementation of reference models of sugars

To apply our quantification method to mixtures of sugars, we need to define reference signals for each isomer, which we then match with the observed data to determine their relative fractions in solution. For this task, we use the quantum mechanical description of the spin systems as outlined in Section 3.2 and specifically rely on its property of field-invariance.

For the purpose of analyzing ^{1}H NMR datasets, we model each isomer of the considered sugars as a spin system of 7 coupled protons. We regard a molecule of sucrose as being composed of two uncoupled moieties, glucose and fructose, present in equal amounts and model them with separate QM formulations. Thus, for each spin system we need to estimate 7 chemical shifts and 6–7 *J*-coupling constants to completely specify the model of an isomer. These parameters have been previously assigned using high-field NMR data and are reported in the literature [41–43]. However, differences in experimental conditions in these sources (e.g. presence and amount of D_2O in solutions added to provide a lock signal for the spectrometer) may notably affect the reported chemical shifts and are difficult to account for consistently. To avoid these inaccuracies and confirm the parameter values, we acquire spectra of ≈ 0.5 M pure sugar solutions in 100% deionized water on a 400 MHz spectrometer, generate QM model signals for each case using Eq. (7) and fine-tune their parameters. The found values are listed for each studied isomer in Appendix B and agree with the assignments reported elsewhere [41–45]. The top panel in Fig. 3 compares the generated models of glucose isomers with experimental spectra of an equilibrated solution of pure glucose measured on a high-field spectrometer. We observe a good match between the models and experimental spectrum with an rms residual of 0.067, which is 30 times lower than the average height of the glucose peaks (≈ 2.0). Similarly, we obtain models for the four most abundant isomers of fructose (Fig. 4) and the two components of sucrose (Fig. 5), which also closely match the experimental spectra.

To validate the obtained models for analyzing lower-field data, we collect ^{1}H spectra of the same three samples of pure sugars with a benchtop instrument. Using the proton chemical shifts and *J*-coupling constants measured at high field, we generate signal models corresponding to the 43 MHz operating frequency and fit them to the data by adjusting only the position of the entire group of peaks. Bottom panels in Figs. 3–5 demonstrate the observed benchtop data and the fitted models. Notably, resonances attributed to different protons can no longer be separated in these spectra. Instead, we observe an ensemble of closely spaced and highly overlapping peaks corresponding to different quantum transitions of these large spin systems. To simplify fitting of low resolution data, the widths of all peaks of each specie are assumed to be equal; we find $\alpha = 1.0\ s^{-1}$ for fructose, $2.3\ s^{-1}$ for glucose, and $3.4\ s^{-1}$ for sucrose. When analyzing mixtures of three sugars, we use these values to initialize the fitting, but adjust them simultaneously with a single top level parameter $\widetilde{\alpha}$.

Benchtop NMR spectra typically have lower SNR than traditional high field data and are often thought to have poorer lineshape. However, the lineshape imperfections seen in the benchtop NMR data in Figs. 3–5 are not due to inhomogeneity of the magnetic field but rather arise from the different quantum

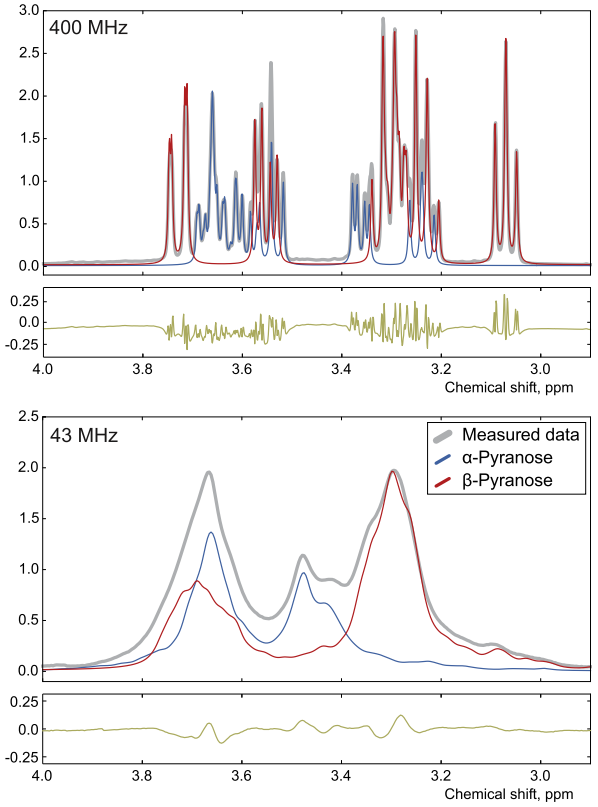

Fig. 3. Modeling the reference spectra of glucose at 400 MHz (top) and 43 MHz (bottom). The water peak fitted with our model was subtracted from the spectrum in this figure for clarity. Panels under each plot show the residual signals after model fitting.

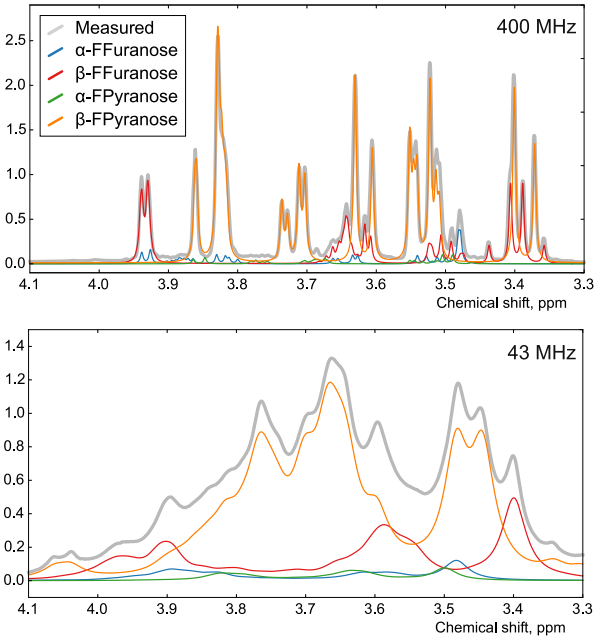

Fig. 4. Modeling the reference spectra of fructose at 400 MHz (top) and 43 MHz (bottom). The water peak fitted with our model was subtracted from the spectrum in this figure for clarity.

transitions of these large spin systems. To illustrate this, Fig. 6 displays the modeled resonances for the isomers of glucose. Clearly, assymetry of peaks and small bumps near the baseline observed

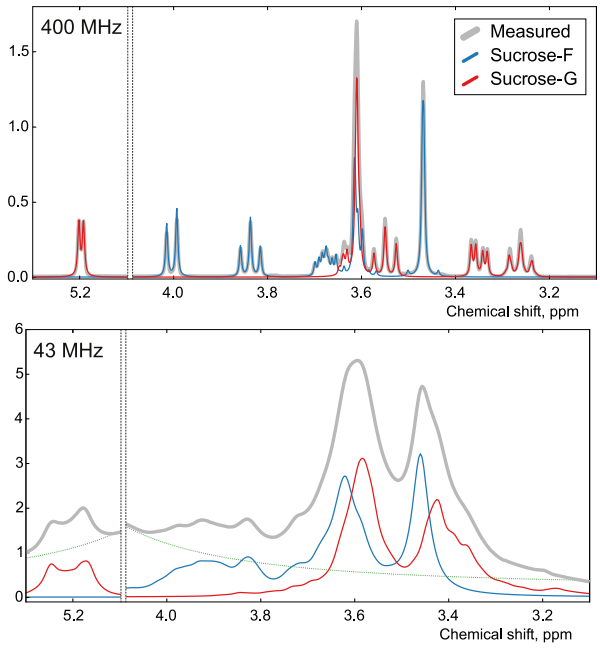

Fig. 5. Modeling the reference spectra of sucrose at 400 MHz (top) and 43 MHz (bottom). The green dashed line in the 43 MHz spectrum corresponds to the fitted water peak. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

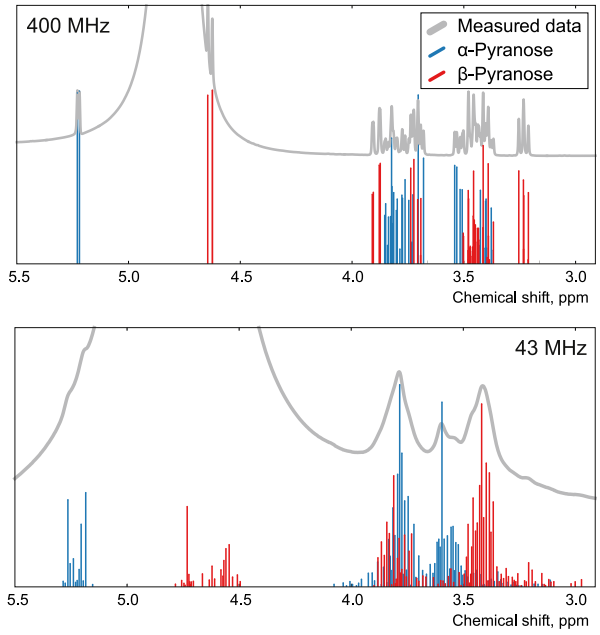

Fig. 6. Separate quantum transition peaks comprising spectra of glucose at high (top) and medium (bottom) field strength. Note that lineshape imperfections observed in the benchtop data, such as peak asymmetries and small bumps near the baseline, are well explained by superposition of multiple resonances.

in the experimental spectra are a result of superposition of the transition peaks. They are well explained and fitted by our QM models transferred to the medium field, which achieves rms of the residual of only 0.036 – two orders of magnitude lower than the dynamic range of the glucose peaks (≈ 2.0) and four orders of magnitude lower than the height of the neighbouring water peak (262.5, not shown in Fig. 3).

Finally, having a model for each sugar isomer, we use our proposed method to quantify their relative concentrations in solutions measured with high and medium field strength spectrometers. The obtained results are compared in Table 2 with their reference values found in the literature ([42] for glucose and [43] for fructose) or the ground truth (for sucrose). Our method achieves error of no more than 1% in relative concentrations of isomers with medium-field and less than 0.5% with high-field data. The reported 95% confidence intervals are computed with Eq. (6), where we estimate σ based on the difference between the measured data and the fitted model. Thus, in addition to the effect of noise, the estimated uncertainties include the effect of slight model misfit due to possible errors in model parameters. Notably, our estimates obtained at both field strengths have comparable confidence bounds. We discuss the implications of different error sources on the final accuracy of estimation in the next section and extend our analysis to quantification of mixtures of sugars.

5.1.2. Outline of our computational approach

To apply the principles of model-based quantification to the analysis of sugar mixtures with benchtop NMR, we rely on the field invariance property of QM models described in Section 3.2. We assume that chemical shifts and *J*-couplings obtained using well resolved high-field data remain unchanged in different samples. Thus, models of spin systems defined with these values can be transferred to lower field strengths and immediately used to analyze benchtop data without any fine-tuning. Indeed, our analysis of sugar solutions with pH varying from 1.2 to 10 (see Fig. 7) reveals that deviations in positions of individual peaks occur only at extreme pH. Experiments with multiple samples of natural fruit juices in Section 5.2.3, whose pH range from 2.2 for lime juice to 4.6 for mango juice, corroborate the assumption that peaks of the studied sugars do not shift noticeably with respect to each other, and their spectra are effectively fixed. However, we allow changes in widths and positions of all peaks simultaneously when fitting models for each particular experiment.

Furthermore, to estimate possible errors due to model misfit, we run the following numerical simulation. Using the values listed in Table 4 in Appendix B, we generate synthetic signals for the mixture of glucose isomers taken in equal amounts at 400 and 43 MHz ^{1}H spectrometer frequencies. We then apply our quantification method to these synthetic signals. To simulate the uncertainty in the experimental parameters, we randomly deviate all chemical shifts in the model with respect to the "true" values used to generate the data. Interestingly, the same deviation in chemical shift leads to much higher discrepancy between the model and the data at higher operating frequency than at 43 MHz as shown in Fig. 8, which agrees with similar recent findings in [46]. Consequently, all other conditions being equal, model-based quantification of benchtop data is more robust to inaccuracies in chemical shifts of the peaks (e.g. as a result of changing pH) than its higher resolution 400 MHz counterpart. Model errors such as these will give

Table 2
Relative equilibrium concentrations of sugar isomers in aqueous solutions found with our method using data from 400 MHz spectrometer to 43 MHz benchtop instrument; the estimates are reported along with their 95% confidence intervals.

	Reference	400 MHz	43 MHz
α-GP	0.375	0.373±0.007	0.379±0.006
β-GP	0.625	0.627±0.009	0.621±0.007
α-FF	0.062	0.051±0.005	0.053±0.008
β-FF	0.224	0.228±0.007	0.241±0.010
α-FP	0.027	0.023±0.007	0.019±0.009
β-FP	0.682	0.698±0.011	0.688±0.016
Su-F	0.5	0.498±0.006	0.502±0.005
Su-G	0.5	0.502±0.006	0.498±0.005

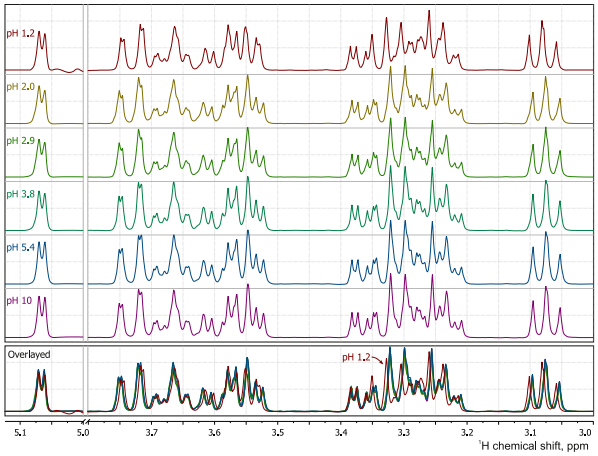

Fig. 7. Spectra of 0.3 M glucose samples with varying pH; the chemical shift scale is positioned to match the anomeric doublet at 5.07 ppm. Note that changes in the spectra due to moving peaks occur only at the extreme pH of 1.2. In the pH range of most fruit juices (2.2 . . . 4.6), positions of the glucose peaks can be considered fixed. Similar results are observed for fructose and sucrose as well.

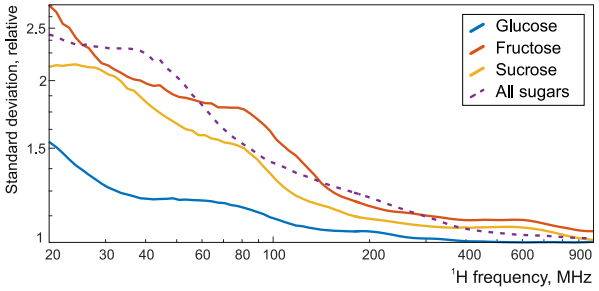

Fig. 9. Average deviation of intensity estimates due to peak overlap expected with our model-based quantification approach at varying field strength. The solid lines correspond to estimating relative concentrations of isomers of glucose and fructose as well as the two components of sucrose; the dashed line is for the case of analyzing mixtures of the three sugars (while holding ratios of their isomers fixed in the models).

a lower limit on the estimated error that cannot be removed even with improvement in the SNR of the measured data, but can be mitigated by additional fitting of underlying model parameters where possible.

On the other hand, the uncertainty in the estimates inevitably increases at lower field strength due to significantly reduced spectral resolution. The effects of peak overlap on the estimation accuracy can be conveniently quantified with Eq. (6) assuming fixed variance of noise $\sigma^2 = 1$. Using datasets simulated at various field strengths, we plot in Fig. 9 the average estimated values of deviations of component intensities, $\widetilde{c} = \frac{1}{K}\sum_k \widetilde{c}_k$, for mixtures of isomers of glucose and fructose and also for the two components of sucrose modeled separately. As expected, the analysis of mixtures of four fructose isomers is a more challenging problem with higher associated uncertainty than the quantification of binary mixtures of glucose isomers or sucrose components. In the latter cases, the non-overlapping peaks of anomeric protons in spectra of glucose located further downfield from the main group of peaks allow us to achieve more accurate quantification. Furthermore, the expected estimation error due to peak overlap increases almost twofold at 43 MHz comparing to the corresponding 400 MHz cases. Importantly, these simulations assume that the true chemical shifts and *J*-couplings are known. With experimental data, their estimation inevitably contributes to additional errors. Since high-field models are more sensitive to the accuracy of parameters, their resulting uncertainty is comparable to that of the values obtained at 43 MHz, as we observe in Table 2.

In most applications, the combined amount of all isomers of a specific sugar is of primary interest. Since the equilibrium concentrations of different isomers c_j are known (see Table 2), we define a (group) reference model for each sugar as a sum of models of its isomers u_j included in fixed proportion, $\mathrm{u} = \sum_j c_j \mathrm{u}_j$. Similarly, we model two moieties of sucrose separately and then combine them in equal ratios. Thus, when processing complex mixtures, such as fruit juices, we effectively reduce the number of degrees of freedom, which facilitates model fitting, and directly quantify the total amount of each sugar. As shown in Fig. 9, we may expect average uncertainties in quantifying mixtures of the three sugars to be comparable to the analysis of fructose isomers or the sucrose moieties.

Finally, to model the peak of water, we use a single Lorentzian centered at ≈ 4.75 ppm, but without an intention to quantify it. Instead, we restrict fitting to a narrow region of interest that includes only spectra of sugars (namely ≈3.0–4.0 ppm and also 5.2–5.5 ppm in high-field data to include the peaks of anomeric protons). Thus only the shoulders of the water peak are fitted, which serves the purpose of baseline correction for the analysis of much weaker sugar components.

5.2. Quantification results

In this subsection, we present results of quantitative analysis of sugar mixtures with our method using benchtop NMR data. First we look at the ability of our method to distinguish between

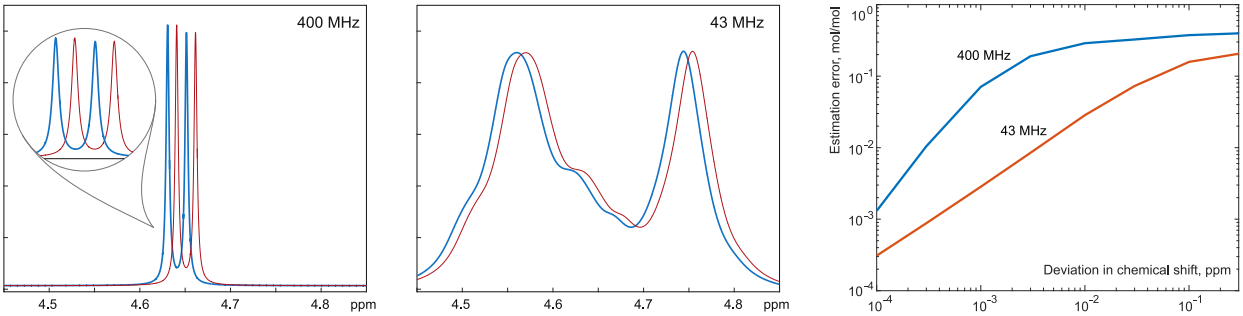

Fig. 8. Differences between peaks of the anomeric proton in modeled spectra of β-Glucopyranose as a result of a chemical shift deviation by 0.01 ppm simulated for 400 MHz (left) and 43 MHz (middle) spectrometer frequencies. Right: The estimation error as a function of deviation in chemical shifts between the model and the data; chemical shifts of all protons in the models are offset by random amounts from their values used to generate simulated data.

different isomers of glucose. Then we process data of mixtures with known compositions, and finally look at the analysis of natural fruit juice samples.

5.2.1. Monitoring interconversion of glucose in H_2O

The ability of NMR spectroscopy to distinguish different structural isomers makes it an important tool for studying kinetics of interconversion reactions of reducing sugars [47,48]. For example, α-glucopyranose, upon dissolving in water, undergoes conversion into its β- form until the equilibrium at approximately 37.5 : 62.5 in the respective mole fractions is reached [49]. The rate of reaction, determined by fitting a mono-exponential model, is highly dependent on experimental conditions, primarily the temperature. Due to the difference of optical properties of the glucose isomers, their interconversion has been traditionally studied with polarimetric methods. However, NMR provides a convenient alternative potentially applicable in more general cases for studying structural changes in other species.

In high-field NMR spectroscopy, the amounts of glucose isomers are typically quantified by integrating the peaks of the anomeric proton that occur at 5.2 and 4.6 ppm for α- and β- forms respectively (see Fig. 6). Unfortunately, they are often obstructed by the much stronger peak of water at ≈ 4.75 ppm, which distorts the baseline and complicates integration. The use of deuterated solvent alleviates this problem, but notably slows down the reaction, which may be undesirable. Likewise, water suppression techniques inevitably introduce scaling into the acquired spectrum reducing the accuracy of quantification and thus should be avoided if possible. The spectrum resolution is further reduced at lower field strengths of benchtop NMR; separate anomeric peaks can no longer be clearly observed and integrated, which makes signal modeling methods the only viable option for quantification.

In our experiment, we prepare a 0.5 M solution of glucose in 100% deionized H_2O, immediately place it into a sample tube, and transfer it to the benchtop instrument. A series of spectra was acquired at 45 s intervals over the course of several hours. The temperature inside the bore of the magnet was kept at approximately 30 °C.

The left panel in Fig. 10 shows an overlay of the acquired spectra. In each spectral plane, we estimate relative concentrations of the glucose isomers with our method and plot the found results with respect to the reaction time in the right panel of the figure. The rate constant estimated as the parameter of the fitted mono-exponential law was found to be $5.5 \times 10^{-4}\ s^{-1}$, which is within the range of reported reference values [49]. While SNR in this dataset is relatively high (the ratio of the height of glucose peaks to the standard deviation of noise is ≈ 10), the close proximity of the significantly more intense water peak notably distorts the baseline and complicates quantification. Despite these challenges, our method achieves stable reconstruction with low deviation of estimates from the underlying exponential model of the experiment.

5.2.2. Quantification of controlled mixtures

In the next series of experiments we extend our method to quantification of ternary mixtures of glucose, fructose, and sucrose in varying relative concentrations in pure deionized H_2O. We collect data for each sample on both medium and high-field spectrometers. In our analysis, we use the model signals designed in Section 5.1.1 and only adjust their global position and peak width simultaneously for all three sugars. The relative concentrations of different isomers of fructose and glucose are fixed at their reference values reported in Section 5.1.1. To account for the baseline distortions caused by the shoulders of the water peak, we fit its position and width, but we do not quantify the water amount. Finally, we also process the high-field data using peak integration. Peak integration is not possible for the low-field data due to extensive peak overlap.

The results of quantification are summarized in Fig. 11 and compared to the relative concentrations estimated gravimetrically; their deviations averaged across the samples for each species are listed in Table 3. The uncertainties in the gravimetric mole fractions are determined by means of error propagation based on the accuracy of laboratory balances and are expected to be no more than 10^{-4} mol/mol. The values estimated with our method are equipped with respective 95% confidence intervals tighter than 0.01 mol/mol for the benchtop and 0.01–0.02 mol/mol for the high-field data (or 1% and 1–2% respectively). Notably, the results of our model-based analysis of the benchtop data are comparable and often better then those obtained with the high-field spectrometer. Slightly higher than expected average deviations observed in Table 3 for both methods can be attributed primarily to an imperfect fit of model signals. Specifically, accurate estimation of chemical shifts and J-coupling constants of fructose is notably challenging even with well-resolved high-field data due to significant overlaps of the spectra of its multiple isomers. Furthermore, closely located peaks of coupled protons – for example, H_{6a} and H_{6b} in sucrose (see Table 4 in Appendix B) – are extremely sensitive to misspecification of their parameter values. These imperfect input models inevitably contribute to the final quantification error. Thus, Eq. (6) is regarded here as a lower bound on the confidence intervals of quantification. More comprehensive analysis of the contribution of uncertainty in specific model parameters to the final accuracy of estimation can be performed in a probabilistic setting and is a topic of ongoing research [25]. Finally, the results of peak integration are strongly affected by the quality of baseline correction in the vicinity of the water peak and also the necessity to estimate the integral of fructose as a difference due to the peak overlap.

5.2.3. Composition analysis of fruit juices

Finally, in this section we analyze compositions of natural fruit juices with our model-based approach. Several samples of juices prepared from fresh fruit are measured on a high-field 400 MHz spectrometer as well as a benchtop instrument, approximately one week later. An example of acquired spectra on both instruments along with models fitted to them are shown in Fig. 12.

We estimate the amounts of glucose, sucrose, and fructose as described in Section 5.1.2 and plot the relative mole fractions on ternary diagrams in Fig. 13. Each circle represents a separate sample; their radii are proportional to the total amounts of sugars estimated approximately with respect to the absolute intensity of the acquired signals. Furthermore, in addition to the three sugars, we quantify the ratio of predominant acids (malic and citric) and indicate it with pie charts for each sample.

One can immediately note the strong agreement between data from the high-field and benchtop instruments: for example differences among apples of several varieties are consistent in both datasets. Similarly, citrus fruit and grape juices occur in the corresponding regions of the diagrams with similar total amounts of sugars. While the spectra of sugars remain stable with varying pH, in our study, we observe that chemical shifts of peaks of citric and malic acids deviate notably among multiple samples of juices. In addition to their typically lower concentration, this increases the resulting uncertainty. Nevertheless, the estimated mole fractions of the two predominant acids are consistent between the spectrometers, except in a few cases.

Furthermore, a sample of bottled apple juice measured twice approximately six months apart (bottom left corner of the diagrams) shows a decrease in the amount of sucrose as the juice fermented. Similar changes occur with the yellow kiwifruit juice: this

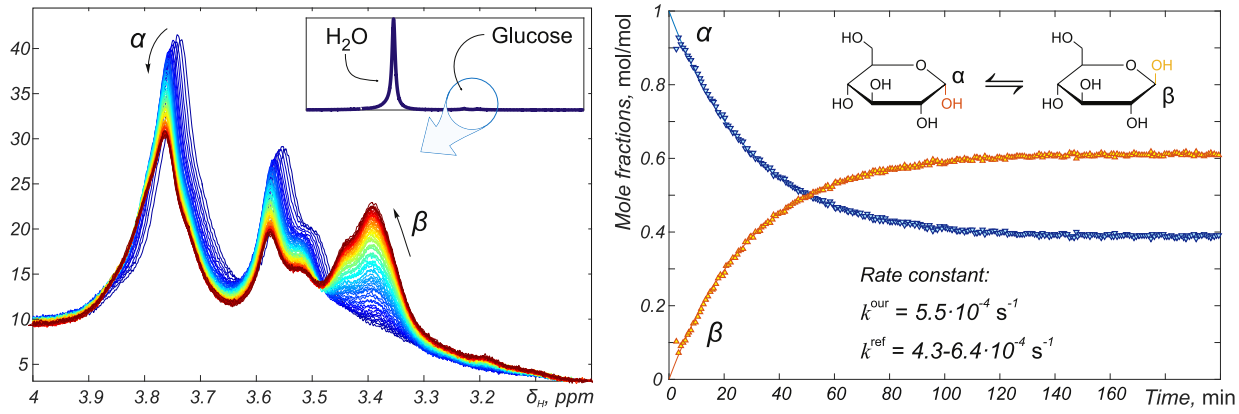

Fig. 10. Monitoring glucose interconversion with benchtop NMR. Left: The acquired data array overlayed in a single plot. Right: The result of exponential fitting to the quantified mole fractions of glucose isomers. The notable deviation of the first point from the fitted model is supposedly caused by equilibration of temperature immediately upon placing the sample into the spectrometer.

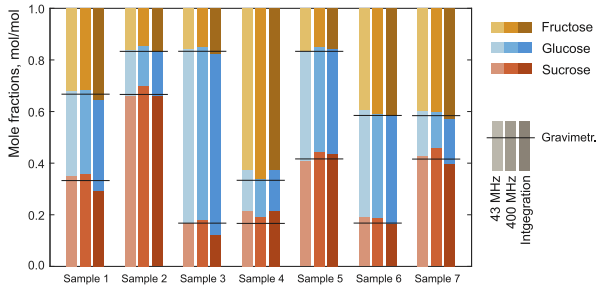

Fig. 11. Quantitative analysis of controlled sugar mixtures. For each sample, from left to right, results of our method with data acquired on a benchtop (43 MHz) and a high-field (400 MHz) spectrometers, and results of the conventional peak integration of the high-field spectra. Horizontal bars indicate gravimetrically estimated values of mole fractions.

sample was measured on the high-field spectrometer immediately after preparation but only two weeks later with the benchtop instrument. This suggests the potential of portable benchtop instruments for accurate on-line monitoring of fermentation reactions, which are of great importance in the food industry.

Finally, in the central panel of Fig. 13, we show a similar diagram for composition of some fruit juices using data from the Nutrition Database of the US Department of Agriculture [50]. These reference values support the findings of our analysis.

6. Conclusions

This paper presents a novel method for quantitative analysis of low-resolution NMR data acquired on benchtop instruments. Our model is formulated in terms of quantum mechanical parameters of the underlying spin systems and is inherently field-invariant: the same values of chemical shifts and *J*-coupling constants easily evaluated with well resolved high-field spectra faithfully represent data acquired at medium-field strength. Furthermore, possible

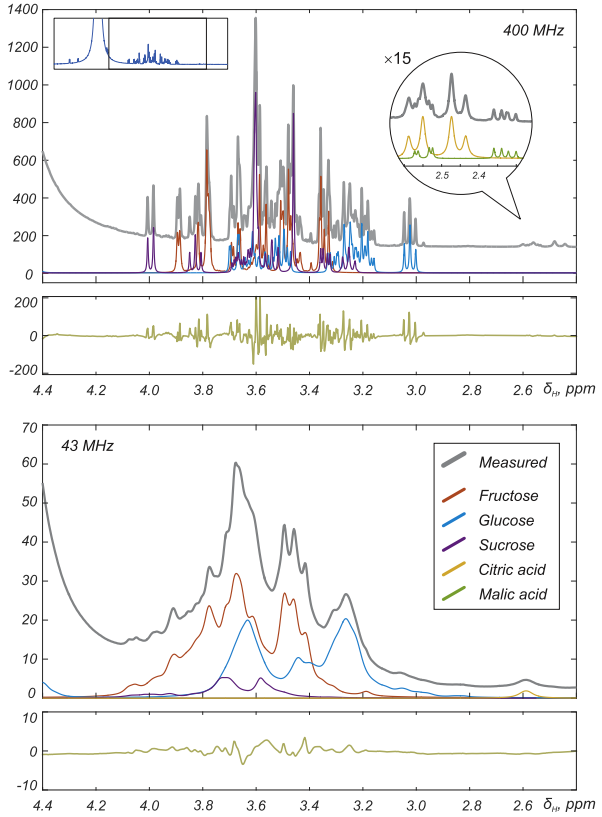

Fig. 12. Examples of spectra of a juice (yellow kiwi) acquired at different field strengths along with fitted models of sugars and two acids. The residual signals between the models and the data are shown beneath each graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Average deviations of estimated mole fractions from gravimetric values (expressed in percents of total amounts) in the analysis of ternary sugar mixtures quantified with our proposed model-based approach and with the traditional peak integration.

		Glucose	Fructose	Sucrose	Avg.
Model-based	43 MHz	1.70	0.55	1.62	1.29
	400 MHz	2.71	1.48	1.28	1.83
Integration		2.52	1.17	1.38	1.69

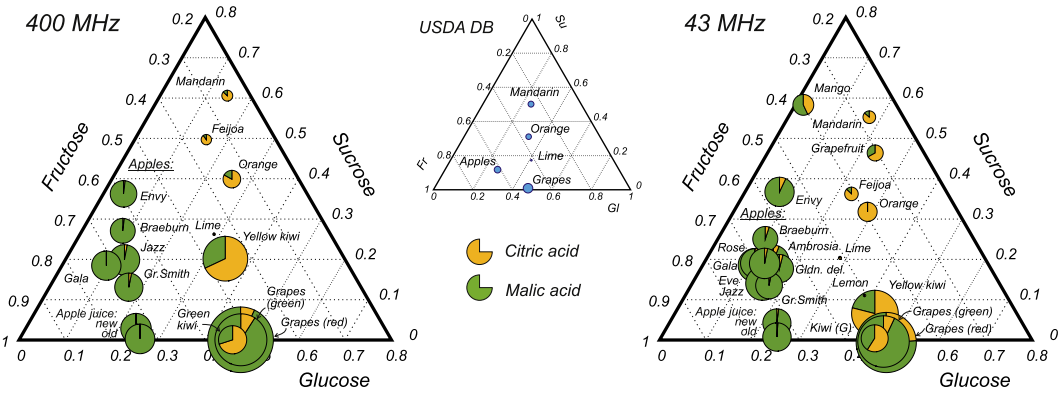

Fig. 13. Results of analyzing compositions of different natural fruit juices with high-field (left) and benchtop (right) spectrometers. The position of each circle on the ternary diagrams encodes the relative mole fraction of the three sugars, and its radius is proportional to the total amount of sugars. The estimated fractions of citric and malic acids are indicated with pie charts for each sample. Reference values for some juices from the Nutrient Database of the US Department of Agriculture are given in the middle [50].

chemical shift variations due to differences in experimental conditions are typically less pronounced in benchtop data, which makes our method particularly appealing in these cases.

Our quantitative analysis of unbuffered sugar mixtures in water – a challenging problem even for high-field NMR – yields accuracy of up to 0.02 *mol*/*mol* for concentrations of major sugars measured with a benchtop instrument. It is comparable to the conventional approach of peak integration, which however, requires spectra of sufficient resolution and is inapplicable for the benchtop data. A formal study of contribution of uncertainties in model parameters to the overall accuracy of quantification is a topic of ongoing research.

Acknowledgements

Financial support was provided by the New Zealand Ministry of Business, Innovation and Employment.

Appendix A

This section provides necessary detail about the functional relation f^{QM} between the free QM parameters of a spin system – spin chemical shifts $\boldsymbol{\delta}$ and J-couplings $\boldsymbol{J}$ – and the positions and intensities of the resulting peaks in the spectrum as defined in Eq. (7). With a notable exception of the simplest AB system of two coupled spins, f^{QM} cannot be expressed explicitly in closed form. Instead, we use the matrix formalism for representation of angular momentum operators and diagonalize the Hamiltonian of the spin system directly [36]. The Hamiltonian matrix $\mathbf{H}$ is defined as a sum of the chemical shift (Zeeman) $\mathbf{H}_Z$ and coupling terms $\mathbf{H}_J$ as:

$$\begin{aligned}\mathbf{H} &= \mathbf{H}_Z + \mathbf{H}_J \\ &= -2\pi\gamma B_0 \sum_s [1-\delta_s]\mathbf{I}_z(s) \\ &\quad +2\pi \sum_{s,s'>s} \boldsymbol{J}_{s,s'} \left[\mathbf{I}_x(s)\mathbf{I}_x(s') + \mathbf{I}_y(s)\mathbf{I}_y(s') + \mathbf{I}_z(s)\mathbf{I}_z(s')\right],\end{aligned} \tag{8}$$

where the indexes s and s' run over all spins in the system and $\mathbf{I}_x, \mathbf{I}_y, \mathbf{I}_z$ denote their respective Cartesian spin operators; γ is a gyromagnetic ratio and B_0 is the magnetic field strength in T [35]. For a system of n spins, we define the above operators as Kronecker tensor products of series of n $\mathbb{C}^{2\times 2}$ matrices:

$$\mathbf{I}_x(1) = \sigma_x \otimes \sigma_0 \ldots \sigma_0,$$

$$\mathbf{I}_x(2) = \sigma_0 \otimes \sigma_x \ldots \sigma_0,$$

$$\mathbf{I}_x(s) = \sigma_0 \otimes \sigma_0 \ldots \sigma_x \ldots \sigma_0,$$

with similar equations for $\mathbf{I}_y$ and $\mathbf{I}_z$, where σ_0 and σ_x, σ_y, and σ_z are the identity operator and the three Pauli matrices respectively:

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},$$

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

That is, for each $s = 1, \ldots, n$, in the sequence of Kronecker products $\sigma_0 \otimes \sigma_0 \otimes \ldots \otimes \sigma_0$, the sth term is substituted with a certain Pauli matrix resulting in the corresponding $2^n \times 2^n$ operator, $\mathbf{I}_x(s), \mathbf{I}_y(s)$, or $\mathbf{I}_z(s)$.

Each eigenpair $(\mathbf{h}_q, \lambda_q)$ of the diagonalized Hamiltonian $\mathbf{H} = \sum_q \lambda_q \mathbf{h}_q \mathbf{h}_q^{\mathrm{T}}$ corresponds to a certain quantum state from $\alpha\alpha\ldots\alpha$ to $\beta\beta\ldots\beta$ in the decreasing order of eigenvalues λ_q for $q = 1, \ldots, 2^n$. If the transition from state q' to state q'' is allowed, i.e. if an odd number of spins flip their state, the frequency of the corresponding resonance is found as $\omega_p = 2\pi(\gamma B_0 + \lambda_{q'} - \lambda_{q''})$, and the intensity of the peak is $b_p = \left[\mathbf{h}_{q'}^{\mathrm{T}} \mathbf{h}_{q''}\right]^2$.

Appendix B

See Table 4 and Fig. 14.

Table 4
Chemical shifts with respect to DSS (top row in normal font) and three-bond *J*-coupling constants (bottom row, in *italic* font) used to generate model signals for the isomers of studied sugars: α- and β-Glucopyranose (GP), α- and β-Fructofuranose (FF), α- and β-Fructopyranose (FP), and the two moeities of Sucrose, Su-F and Su-G.

	H_1 $^3J_{1-2}$	H_{1a} $^3J_{1a-1b}$	H_{1b}	H_2 $^3J_{2-3}$	H_3 $^3J_{3-4}$	H_4 $^3J_{4-5}$	H_5 $^3J_{5-6a}$	H_{6a} $^3J_{5-6b}$	H_{6b} $^3J_{6a-6b}$
α-GP	5.230			3.531	3.709	3.408	3.824	3.840	3.764
	3.817			*9.732*	*9.336*	*10.357*	*1.870*	*5.349*	*−12.129*
β-GP	4.641			3.239	3.483	3.401	3.459	3.896	3.722
	8.326			*9.158*	*9.389*	*9.604*	*2.001*	*5.962*	*−12.419*
α-FF		3.646	3.658		4.105	3.993	4.054	3.817	3.695
		−12.090			*4.220*	*7.244*	*3.043*	*5.508*	*−11.900*
β-FF		3.590	3.551		4.108	4.105	3.822	3.799	3.674
		−12.350			*8.500*	*7.572*	*2.976*	*6.368*	*−12.660*
α-FP		3.707	3.650		4.033	3.941	3.863	3.857	3.707
		−11.680			*6.729*	*2.954*	*1.574*	*2.550*	*−12.790*
β-FP		3.708	3.560		3.792	3.891	3.995	4.017	3.701
		−11.730			*9.978*	*3.554*	*1.121*	*1.753*	*−12.930*
Su-F		3.676	3.676		4.211	4.044	3.884	3.828	3.809
		−12.500			*8.751*	*8.469*	*2.723*	*6.952*	*−12.310*
Su-G	5.440			3.558	3.756	3.470	3.832	3.816	3.818
	3.350			*9.997*	*9.076*	*10.130*	*2.244*	*4.329*	*−14.440*

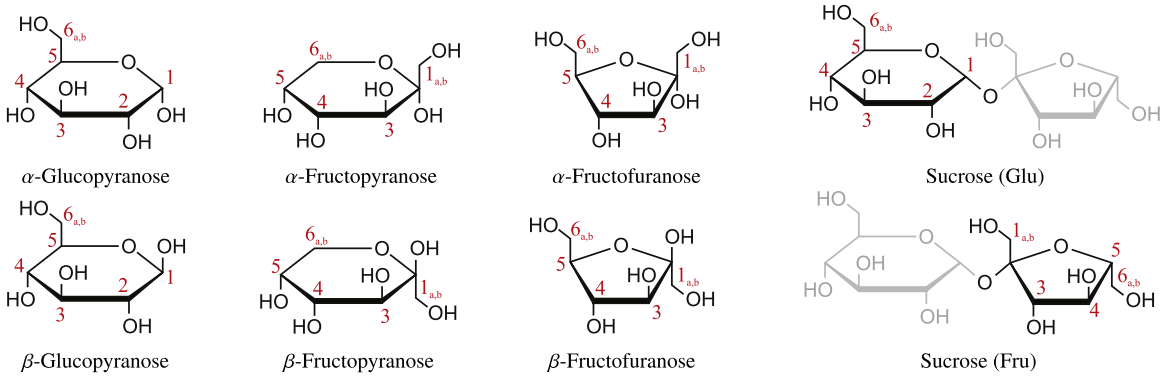

Fig. 14. Structural formulas of the studied isomers of sugars.

References

[1] L. Mannina, A.P. Sobolev, S. Viel, Liquid state 1H high field NMR in food analysis, Prog. Nucl. Magn. Reson. Spectrosc. 66 (2012) 1–39, https://doi.org/10.1016/j.pnmrs.2012.02.001.

[2] M.F. Marcone, S. Wang, W. Albabish, S. Nie, D. Somnarain, A. Hill, Diverse food-based applications of nuclear magnetic resonance (NMR) technology, Food Res. Int. 51 (2) (2013) 729–747, https://doi.org/10.1016/j.foodres.2012.12.046.

[3] C. Simmler, J.G. Napolitano, J.B. McAlpine, S.N. Chen, G.F. Pauli, Universal quantitative NMR analysis of complex natural samples, Curr. Opin. Biotech. 25 (2014) 51–59, https://doi.org/10.1016/j.copbio.2013.08.004, arXiv: NIHMS150003.

[4] G.F. Pauli, T. Gödecke, B.U. Jaki, D.C. Lankin, Quantitative 1 H NMR. Development and potential of an analytical method: an update, J. Nat. Prod. 75 (4) (2012) 834–851, https://doi.org/10.1021/np200993k, arXiv: NIHMS150003.

[5] F. Alsmeyer, H.-J. Koß, W. Marquardt, Indirect spectral hard modeling for the analysis of reactive and interacting mixtures, Appl. Spectrosc. 58 (8) (2004) 975–985, https://doi.org/10.1366/0003702041655368. <http://journals.sagepub.com/doi/10.1366/0003702041655368>.

[6] E. Kriesten, F. Alsmeyer, A. Bardow, W. Marquardt, Fully automated indirect hard modeling of mixture spectra, Chemometr. Intell. Laborat. Syst. 91 (2) (2008) 181–193, https://doi.org/10.1016/j.chemolab.2007.11.004.

[7] C. Cobas, F. Seoane, S. Sýkora, Global Spectral Deconvolution (GSD) of D-NMR spectra, Poster at SMASH Conference (d) (2008) 15706. https://doi.org/10.3247/sl2nmr08.011.

[8] T. Schoenberger, S. Menges, M.A. Bernstein, M. Pérez, F. Seoane, S. Sýkora, C. Cobas, Improving the performance of high-precision qNMR measurements by a double integration procedure in practical cases, Anal. Chem. 88 (7) (2016) 3836–3843, https://doi.org/10.1021/acs.analchem.5b04911.

[9] L. Vanhamme, A. van den Boogaart, S. Van Huffel, Improved method for accurate and efficient quantification of MRS data with use of prior knowledge, J. Magn. Reson. 129 (1) (1997) 35–43, https://doi.org/10.1006/jmre.1997.1244. <http://linkinghub.elsevier.com/retrieve/pii/S1090780797912441>.

[10] H. Ratiney, Y. Coenradie, S. Cavassila, D. Van Ormondt, D. Graveron-Demilly, Time-domain quantitation of 1H short echo-time signals: background accommodation, Magn. Reson. Mater. Phys. Biol. Med. 16 (6) (2004) 284–296, https://doi.org/10.1007/s10334-004-0037-9.

[11] J.B. Poullet, D.M. Sima, S. Van Huffel, MRS signal quantitation: a review of time- and frequency-domain methods, J. Magn. Reson. 195 (2) (2008) 134–144, https://doi.org/10.1016/j.jmr.2008.09.005.

[12] Y. Hiltunen, M. Ala-Korpela, J. Jokisaari, S. Eskelinen, K. Kiviniitty, M. Savolainen, Y.A. Kesäniemi, A lineshape fitting model for 1H NMR spectra of human blood plasma, Magn. Reson. Med. 21 (2) (1991) 222–232, https://doi.org/10.1002/mrm.1910210207.

[13] R. Laatikainen, M. Niemitz, U. Weber, J. Sundelin, T. Hassinen, J. Vepsäläinen, General strategies for total-lineshape-type spectral analysis of NMR spectra using integral-transform iterator, J. Magn. Reson. Ser. A 120 (1) (1996) 1–10, https://doi.org/10.1006/jmra.1996.0094. <http://www.sciencedirect.com/science/article/pii/S1064185896900947>.

[14] S.W. Provencher, Automatic quantitation of localized in vivo 1H spectra with LCModel, NMR Biomed. 14 (4) (2001) 260–264, https://doi.org/10.1002/nbm.698.

[15] J. Slotboom, C. Boesch, R. Kreis, Versatile frequency domain fitting using time domain models and prior knowledge, Magn. Reson. Med. 39 (6) (1998) 899–911, https://doi.org/10.1002/mrm.1910390607.

[16] C. Cobas, S. Sýkora, The bumpy road towards automatic global spectral deconvolution (GSD), in: Poster at 50th ENC Conference, Asilomar (CA, USA), March 29–April 4, 2009. https://doi.org/10.3247/SL3Nmr09.003.

[17] K. Krishnamurthy, CRAFT (complete Reduction to Amplitude Frequency Table) – robust and time-efficient Bayesian approach for quantitative mixture analysis by NMR, Mag. Res. Chem. 51 (12) (2013) 821–829, https://doi.org/10.1002/mrc.4022.

[18] J. Hao, W. Astle, M. De iorio, T.M.D. Ebbels, BATMAN – an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model, Bioinformatics 28 (15) (2012) 2088–2090, https://doi.org/10.1093/bioinformatics/bts308, arXiv: 1112.5794.

[19] J. Hao, M. Liebeke, W. Astle, M. De Iorio, J.G. Bundy, T.M.D. Ebbels, Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN, Nat. Protocols 9 (6) (2014) 1416–1427, https://doi.org/10.1038/nprot.2014.090. <http://www.ncbi.nlm.nih.gov/pubmed/24853927>.

[20] C. Zheng, S. Zhang, S. Ragg, D. Raftery, O. Vitek, Identification and quantification of metabolites in 1H NMR spectra by Bayesian model selection, Bioinformatics 27 (12) (2011) 1637–1644, https://doi.org/10.1093/bioinformatics/btr118.
[21] M. Tiainen, P. Soininen, R. Laatikainen, Quantitative quantum mechanical spectral analysis (qQMSA) of 1H NMR spectra of complex mixtures and biofluids, J. Magn. Reson. 242 (2014) 67–78, https://doi.org/10.1016/j.jmr.2014.02.008.
[22] R. Laatikainen, P. Laatikainen, E. Hakalehto, Quantitative quantum mechanical NMR analysis: the superior tool for analysis of biofluids, in: Proceedings of The 1st International Electronic Conference on Metabolomics, MDPI, Basel, Switzerland, 2016, p. C005. https://doi.org/10.3390/iecm-1-C005. URL <http://sciforum.net/conference/iecm-1/paper/3483>.
[23] C.W.P.d.S. Grandizoli, F.R. Campos, F. Simonelli, A. Barison, Grape juice quality control by means of 1H NMR spectroscopy and chemometric analyses, Quimica Nova 37 (7) (2014) 1227–1232.
[24] N. Ogrinc, I.J. Košir, J.E. Spangenberg, J. Kidrič, The application of NMR and MS methods for detection of adulteration of wine, fruit juices, and olive oil. A review, Anal. Bioanal. Chem. 376 (4) (2003) 424–430, https://doi.org/10.1007/s00216-003-1804-6.
[25] Y. Matviychuk, E. von Harbou, D.J. Holland, An experimental validation of a Bayesian model for quantification in NMR spectroscopy, J. Magn. Reson. 285 (2017) 86–100, https://doi.org/10.1016/j.jmr.2017.10.009. <http://linkinghub.elsevier.com/retrieve/pii/S1090780717302550>.
[26] M. Bydder, Solution of a complex least squares problem with constrained phase, Linear Algebra Appl. 433 (11–12) (2010) 1719–1721, https://doi.org/10.1016/j.laa.2010.07.011. <http://linkinghub.elsevier.com/retrieve/pii/S002437951000368X>.
[27] I. Markovsky, On the complex least squares problem with constrained phase, SIAM J. Matrix Anal. Appl. 32 (3) (2011) 987–992, https://doi.org/10.1137/110826497.
[28] G. Bretthorst, Bayesian analysis. I. Parameter estimation using quadrature NMR models, J. Magn. Reson. (1969) 88 (3) (1990) 533–551, https://doi.org/10.1016/0022-2364(90)90287-J. <http://linkinghub.elsevier.com/retrieve/pii/002223649090287J>.
[29] D.V. Rubtsov, J.L. Griffin, Time-domain Bayesian detection and estimation of noisy damped sinusoidal signals applied to NMR spectroscopy, J. Magn. Reson. 188 (2) (2007) 367–379, https://doi.org/10.1016/j.jmr.2007.08.008. <http://linkinghub.elsevier.com/retrieve/pii/S1090780707002467>.
[30] A.G. Wilson, Y. Wu, D.J. Holland, S. Nowozin, M.D. Mantle, L.F. Gladden, A. Blake, Bayesian Inference for NMR Spectroscopy with Applications to Chemical Quantification, arXiv preprint arXiv: 1402.3580.
[31] H. Yan, J.C. Gore, The relation of HSVD to LPSVD for fitting time-domain signals, J. Magn. Reson. (1969) 80 (2) (1988) 324–327, https://doi.org/10.1016/0022-2364(88)90303-4. <http://linkinghub.elsevier.com/retrieve/pii/0022236488903034>.
[32] A.D. Bain, D. Fletcher, P. Hazendonk, What is a transition?, Concepts Magn Reson. 10 (2) (1998) 85–98, https://doi.org/10.1002/(sici)1099-0534(1998)10:2<85::aid-cmr2>3.0.co;2-r.
[33] I. Kuprov, N. Wagner-Rundell, P.J. Hore, Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction, J. Magn. Reson. 189 (2) (2007) 241–250, https://doi.org/10.1016/j.jmr.2007.09.014, arXiv: 0701294.
[34] A.M. Castillo, L. Patiny, J. Wist, Fast and accurate algorithm for the simulation of NMR spectra of large spin systems, J. Magn. Reson. 209 (2) (2011) 123–130, https://doi.org/10.1016/j.jmr.2010.12.008.
[35] S.A. Smith, W.E. Palke, J.T. Gerig, The Hamiltonians of NMR. part I, Concepts Magn. Reson. 4 (2) (1992) 107–144, https://doi.org/10.1002/cmr.1820040202. <http://doi.wiley.com/10.1002/cmr.1820040202>.
[36] J. Canavagh, W.J. Fairbrother, A.G. Palmer II, M. Rance, N.J. Skelton, Theoretical Description of NMR Spectroscopy, Protein NMR Spectroscopy 2nd Edit.
[37] H.J. Hogben, M. Krzystyniak, G.T.P. Charnock, P.J. Hore, I. Kuprov, Spinach – a software library for simulation of spin dynamics in large spin systems, J. Magn. Reson. 208 (2) (2011) 179–194, https://doi.org/10.1016/j.jmr.2010.11.008, arXiv: NIHMS150003.
[38] L.J. Edwards, I. Kuprov, Parallel density matrix propagation in spin dynamics simulations, J. Chem. Phys. 136 (4) (2012) 044108, https://doi.org/10.1063/1.3679656, arXiv: 1109.4816.
[39] S. Ravanbakhsh, P. Liu, T.C. Bjorndahl, R. Mandal, J.R. Grant, M. Wilson, R. Eisner, I. Sinelnikov, X. Hu, C. Luchinat, R. Greiner, D.S. Wishart, Correction: accurate, fully-automated NMR spectral profiling for metabolomics, PLOS ONE 10 (7) (2015) e0132873, https://doi.org/10.1371/journal.pone.0132873. <http://linkinghub.elsevier.com/retrieve/pii/S002437951000368X>. https://plos.org/10.1371/journal.pone.0132873.
[40] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python, 2001. URL <http://www.scipy.org/>.
[41] A. De Bruyn, J. Van Loo, The identification by 1H- and 13C-n.m.r. spectroscopy of sucrose, 1-kestose, and neokestose in mixtures present in plant extracts, Carbohydr. Res. 211 (1) (1991) 131–136, https://doi.org/10.1016/0008-6215(91)84151-4.
[42] M.U. Roslund, P. Tähtinen, M. Niemitz, R. Sjöholm, Complete assignments of the 1H and 13C chemical shifts and JH,H coupling constants in NMR spectra of d-glucopyranose and all d-glucopyranosyl-d-glucopyranosides, Carbohyd. Res. 343 (1) (2008) 101–112, https://doi.org/10.1016/j.carres.2007.10.008.
[43] T. Barclay, M. Ginic-Markovic, M.R. Johnston, P. Cooper, N. Petrovsky, Observation of the keto tautomer of D-fructose in D2O using 1H NMR spectroscopy, Carbohydr. Res. 347 (1) (2012) 136–141. arXiv:NIHMS15000, https://doi.org/10.1016/j.carres.2011.11.003. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3254704/%5Cnhttp://ac.els-cdn.com/S0008621511005404/1-s2.0-S0008621511005404-main.pdf?_tid=ce8cdfd0-c385-11e5-9a62-00000aacb35d&acdnat=1453741696_aa16b829954b02bab334a51af2172cfa>.
[44] A. De Bruyn, M. Anteunls, G. Vereegge, A 1H-n.m.r. study of D-fructose in D2O. 41 (1975) 295–297.
[45] M. Jaseja, A.S. Perlin, P. Dais, Two-dimensional NMR spectral study of the tautomeric equilibria of D-fructose and related compounds, Magn. Reson. Chem. 28 (4) (1990) 283–289, https://doi.org/10.1002/mrc.1260280402.
[46] P.G. Takis, H. Schäfer, M. Spraul, C. Luchinat, Deconvoluting interrelationships between concentrations and chemical shifts in urine provides a powerful analysis tool, Nat. Commun. 8 (1) (2017) 1–11, https://doi.org/10.1038/s41467-017-01587-0.
[47] E.N. Drake, C.E. Brown, Application of nmr to biochemical kinetics. A laboratory experiment in physical biochemistry, J. Chem. Educ. 54 (2) (1977) 124, https://doi.org/10.1021/ed054p124. <http://pubs.acs.org/doi/abs/10.1021/ed054p124>.
[48] A.C. Dona, G. Pages, P.W. Kuchel, Kinetics of starch hydrolysis and glucose mutarotation studied by NMR chemical exchange saturation transfer (CEST), Carbohydr. Polym. 86 (4) (2011) 1525–1532, https://doi.org/10.1016/j.carbpol.2011.06.056. <http://linkinghub.elsevier.com/retrieve/pii/S0144861711005261>.
[49] M. Nelson, F.M. Beegle, Mutarotation of glucose and fructose, J. Am. Chem. Soc. (1919) 559–575, https://doi.org/10.1021/ja01461a007.
[50] United States Department of Agriculture, Agricultural Research Service, USDA Food Composition Databases, <http://ndb.nal.usda.gov/ndb/> (accessed: 2018-04-20).