MultiNMRFit: A software to fit 1D and pseudo-2D NMR spectra

Pierre Millard[1,2,*], Loïc Le Grégam[1,2], Svetlana Dubiley[1], Valeria Gabrielli[1], Thomas Gosselin-Monplaisir[1], Guy Lippens[1], Cyril Charlier[1,*]

[1]Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRA, INSA, Toulouse, 31077, France

[2]MetaToul-MetaboHUB, National Infrastructure of Metabolomics and Fluxomics, Toulouse, 31077, France

Summary. Nuclear Magnetic Resonance (NMR) is widely used for quantitative analysis of metabolic systems. Accurate extraction of NMR signal parameters – such as chemical shift, intensity, coupling constants, and linewidth – is essential for obtaining information on the structure, concentration, and isotopic composition of metabolites. We present MultiNMRFit, an open-source software designed for high-throughput analysis of one-dimensional NMR spectra, whether acquired individually or as pseudo-2D experiments. MultiNMRFit extracts signal parameters (e.g. intensity, area, chemical shift, and coupling constants) by fitting the experimental spectra using built-in or user-defined signal models that account for multiplicity, providing high flexibility along with robust and reproducible results. The software is accessible both as a Python library and via a graphical user interface, enabling intuitive use by end-users with or without computational expertise. We demonstrate the robustness and flexibility of MultiNMRFit on ^{1}H, ^{13}C, and ^{31}P NMR datasets collected in metabolomics and isotope labeling studies.

Availability and Implementation. MultiNMRFit is implemented in Python 3 and was tested on Unix, Windows, and MacOS platforms. The source code and the documentation are freely distributed under GPL3 license at https://github.com/NMRTeamTBI/MultiNMRFit/.

Supplementary data. Supplementary data are available online.

Contact. Cyril Charlier (charlier@insa-toulouse.fr) and Pierre Millard (millard@insa-toulouse.fr)

1. Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful tool for investigating metabolism. Key applications include identifying and quantifying metabolites in metabolomics, determining the isotopic content of metabolites in ^{13}C-fluxomics, and identifying and characterizing protein-metabolite interactions (Koczula *et al.*, 2016; Moco, 2022; Nargund *et al.*, 2013; Diether *et al.*, 2019). Extracting meaningful information from NMR spectra requires precise characterization of each signal of interest, including their position (chemical shift), intensity (and area) and shape (coupling constants and linewidth) (Edison *et al.*, 2021). Signal parameters can be estimated by fitting spectra using various mathematical models that relate these parameters to the observed signals. This fitting step is often achieved with in-house scripts, which, while functional and flexible, may lack the robustness and reproducibility necessary for large-scale or complex NMR studies of metabolic systems. Commercial software, such as TopSpin (Bruker), Chenomx (Chenomx Inc), and Mnova (Mestrelab), offer user-friendly alternatives, but function as proprietary "black boxes" and are limited to predefined signal models. To enhance accessibility and reproducibility, open-source libraries for R, MATLAB, and Python have been developed, including MetaboLab (Ludwig and Günther, 2011), MetaboDecon1D (Häckl *et al.*, 2021), BATMAN (Hao *et al.*, 2012, 2014), AQuA (Röhnisch *et al.*, 2021), and FitNMR (Dudley *et al.*, 2020). However, they are designed for use via scripting and often lack features essential for users without computational expertise. Such missing features include statistical methods for calculating confidence intervals of extracted parameters, graphical user interfaces, and logging systems to ensure reproducibility by saving calculation parameters and process information along with the results. Another key challenge is that most existing tools treat each peak as an independent signal (Smith, 2017). However, the NMR signal of a single nucleus often consists of multiple peaks arising from homo- and heteronuclear scalar couplings, which are intrinsic properties crucial for identifying metabolites and quantifying their isotopic content. While tools like Dmfit for solid and liquid NMR data (Massiot *et al.*, 2002) and, more recently, the R package rnmrfit (Sokolenko *et al.*, 2019) account for multiplicity, most current software does not, which limits their effectiveness in extracting and utilizing this information.

Consequently, detailed analysis of NMR spectra remains a low throughput task, typically requiring expert users, which can limit its scalability for large datasets such as those generated in real-time NMR experiments. The development of user-friendly, open-source software that incorporates signal multiplicity and supports flexible implementation of tailored signal models would significantly enhance the scope, efficiency, quality, and reproducibility of NMR-based metabolic studies.

To address these challenges, we have developed MultiNMRFit, a versatile Python software for fitting one dimensional NMR spectra, whether acquired individually or as pseudo-2D datasets in time-course

experiments. MultiNMRFit is generic and can be applied to any nucleus (as shown here for ^{1}H, ^{13}C, and ^{31}P spectra), any type of 1D or pseudo-2D NMR experiment, and any type of signal. It includes built-in signal models for common multiplets and is easily extendable with user-defined models to process less conventional signals. MultiNMRFit is accessible both as a Python library and via a graphical user interface. This software is intended for both expert and non-expert NMR users, regardless of their computational expertise. We demonstrate the robustness of MultiNMRFit on synthetic data and through its application to real-time metabolomics and isotope labeling experiments.

2. Methods and implementation

The general workflow of MultiNMRFit is shown in Figure 1. MultiNMRFit is a Python-based library that operates within common web browsers via a Streamlit-powered graphical user interface. Users can load 1D and pseudo-2D Bruker spectra pre-processed beforehand (*i.e.* referenced, phase-corrected and baseline-corrected). Spectra can also be provided as text files containing chemical shift values and corresponding intensities. MultiNMRFit performs automated peak picking within a specified spectral region, and users can add additional peaks manually. One or several peaks can then be grouped into a signal, and a model must be selected for each signal. As detailed in the next section, each model contains equations to simulate the corresponding signal based on parameters. MultiNMRFit automatically suggests a list of models based on the number of peaks (e.g., *doublet of doublet*s and *quartet* for signals containing four peaks). Note that a given peak can be included in multiple signals to account for overlapping peaks. The initial values and bounds of each signal's parameters are automatically estimated from the peak list to ensure reliable convergence and can be refined by users if needed. For example, the initial chemical shift and intensity are taken directly from the peak list, while coupling constants are estimated from the difference in chemical shifts between peaks. Bounds can be defined as either absolute or relative values, and their default values (contained in the model files) can be adjusted by users to better reflect their experimental spectra. For spectrum fitting, parameters p are estimated by minimizing the cost function $c(p)$, which is defined as the sum of squared residuals between the experimental intensities (a_i) and the values simulated by the model $y_i(p)$:

$$c(p) = \sum_i (a_i - y_i(p))^2$$

The cost function is minimized using the L-BFGS-B method (Byrd *et al.*, 1995), with an optional initial refinement step performed by the differential evolution algorithm (Storn and Price, 1997). Standard deviations on parameters are estimated from the covariance matrix. For batch processing of spectra, users can select a spectrum as reference to automatically initialize signals and parameters values for the next spectra, thereby automating the processing of multiple spectra for high-throughput analysis. Finally, MultiNMRFit provides a range of interactive plots for visual inspection of fitting results. The

results can be exported as *csv* files containing *(i)* parameter values (with their confidence intervals) and *(ii)* chemical shifts and intensities of experimental and fitted spectra.

The complete process is automatically saved as a pickle file, which can be reloaded in MultiNMRFit to continue previous work or reanalyze calculation results. This file can also be shared, enabling collaborative analysis across multiple users and ensuring reproducibility. Detailed usage of MultiNMRFit can be found in the online documentation (https://multinmrfit.readthedocs.io).

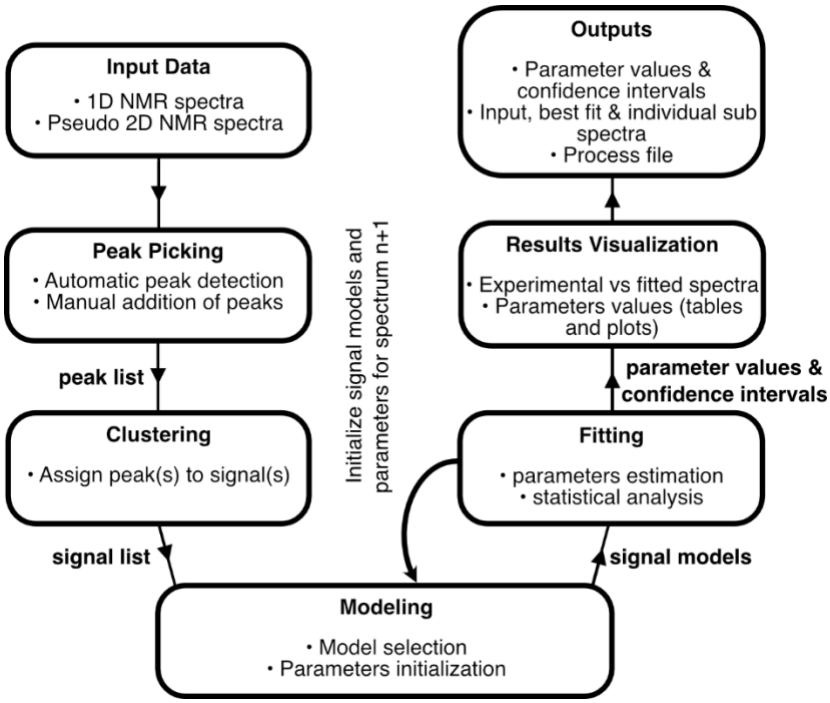


Figure 1. Flow chart of MultiNMRFit.

3. Signal models

A signal model consists of *(i)* equations that describe signal intensity as a function of chemical shift (used for simulations) and *(ii)* a list of parameters with their default values and bounds (used for optimization). Various mathematical functions, including Lorentzian, Gaussian, and Voigt functions, can be used to simulate the NMR spectrum of a given signal. For example, the intensity y_i of a doublet can be described using the following mixed Gaussian-Lorentzian function:

$$y_i = I/2 \cdot \Big(a \cdot \Big(e^{-(x_i - x_s + J/2)^2/(2 \cdot {l_w}^2)} + e^{-(x_i - x_s - J/2)^2/(2 \cdot {l_w}^2)}\Big) + (1-a) \cdot \big(1 + (x_i - x_s + J/2)/{l_w}^2\big) + \big(1 + (x_i - x_s - J/2)/{l_w}^2\big)\Big)$$

with parameters x_s (chemical shift of the signal), I (total signal intensity), J (scalar coupling constant), l_w (linewidth at half maximum), and a (fraction of Gaussian part of the signal). An NMR spectrum containing multiple signals can be simulated by summing the spectra of individual signals.

By default, MultiNMRFit provides mixed Gaussian-Lorentzian models for common multiplets, including singlet, doublet, triplet, doublet of doublets, and quadruplet. Advanced users can create custom models to handle less conventional signals, such as signals with other coupling patterns or described using alternative mathematical functions. A model template is included with MultiNMRFit for this purpose, and a detailed tutorial on model construction is available in the documentation.

4. Validation with synthetic datasets

We validated MultiNMRFit using synthetic datasets (Supplementary data). Briefly, we generated spectra containing two signals (a triplet and a quadruplet) with varying degrees of overlap and noise, and we used MultiNMRFit to retrieve the parameters of each signal. Results confirm the capability of the software to handle complex spectral data and deliver robust parameter estimation even under challenging conditions (such as low signal-to-noise ratio or significant signal overlaps, Supplementary Figures S1 and S2). Moreover, fitting multiplets rather than individual peaks increased the accuracy and precision of the estimated parameters (Supplementary Figure S3).

5. Application to metabolomics datasets

We used MultiNMRFit to monitor the dynamic conversion of glucose through the initial steps of glycolysis. Glucose was added to a buffer solution containing sodium trimethylsilyl propionate (TSP, used as internal standard), ATP and the first three enzymes of glycolysis: glucokinase, glucose-6-phosphate isomerase, and phosphofructokinase (Figure 2A, Supplementary data). The time-course signals of glucose (GLC), glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), fructose-1,6-bisphosphate (FBP), ATP and ADP were monitored in real-time using a pseudo-2D experiment consisting of 1024 1D ^{1}H spectra. All signal parameters were successfully extracted, including for overlapping signals of ADP and ATP (Figure 2B). The integrals of signals specific to each metabolite were normalized to the TSP signal (Figure 2C). A progressive decrease in GLC signal was observed, accompanied by a transient accumulation of G6P and F6P, which were eventually converted into FBP. The high precision and resolution on the estimated dynamics illustrate the value of MultiNMRFit for high-throughput processing of large datasets, with the relative concentrations of the six metabolites of interest quantified in less than three hours for the 1024 spectra. Simultaneously the ^{31}P spectra were acquired along the way (Cox *et al.*, 2019) and were equally fitted using MultiNMRFit (Figure S4),

To demonstrate the applicability of MultiNMRFit for other nuclei, we fitted ^{13}C NMR spectra obtained from a separate experiment in which we monitored the conversion of uniformly ^{13}C-labeled GLC to G6P by hexokinase using a ^{13}C detected pseudo-2D experiment (Supplementary data). In this experiment, both GLC and G6P give rise to a doublet of triplets. Here again, the spectra could be fitted

accurately, and the concentration dynamics of the two ^{13}C-labeled reactants could be obtained from real-time data (Supplementary Figure S5).

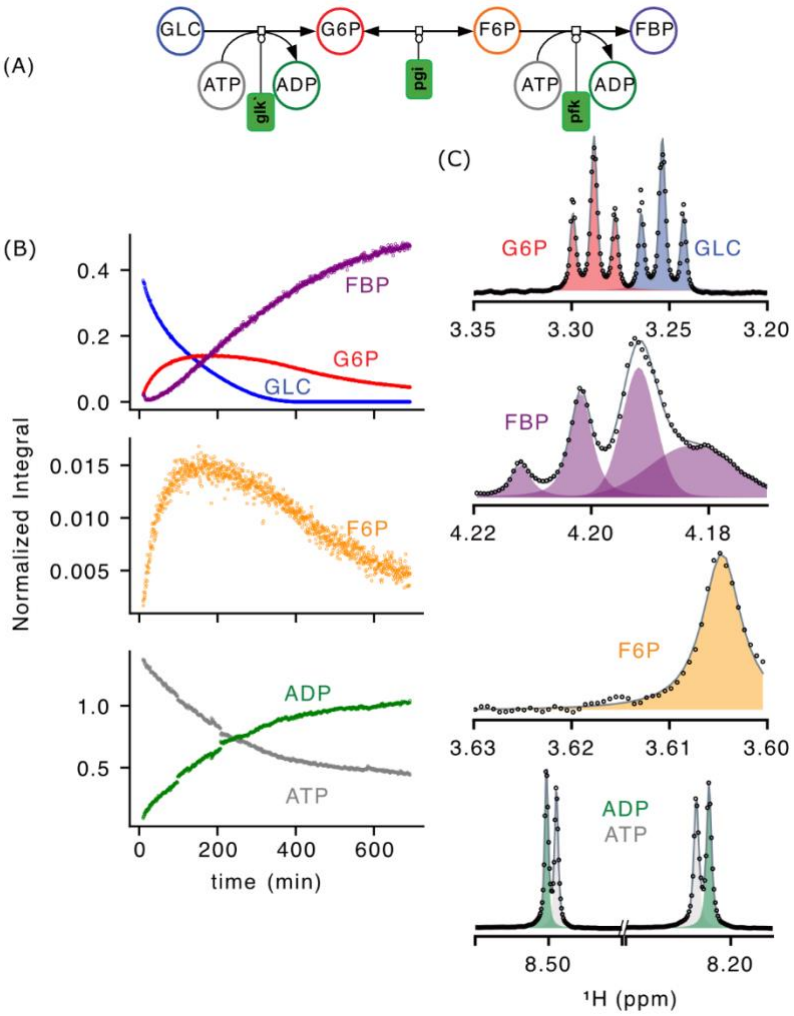


Figure 2. Applications of MultiNMRFit to multi-enzymatic reactions. (A) Scheme of the biochemical pathway monitored in a real-time NMR experiment. **(B)** Dynamic of integrals of glucose (GLC, blue), glucose-6-phosphate (G6P, red), fructose-6-phosphate (F6P, orange), fructose-1,6-bisphosphate (FBP, purple), adenosine triphosphate (ATP, gray), and adenosine diphosphate (ADP, green). Integrals were normalized with the TSP signal. **(C)** Examples of signals processed with MultiNMRFit for each molecule of interest, where individual signals are shown by the shaded areas and the experimental and fitted spectra are represented by dots and lines, respectively.

6. Application to isotopic studies of metabolism

MultiNMRFit can also exploit the multiplicity resulting from the presence of isotopes in metabolites to quantify their isotopic content – a key capability for applications such as detecting counterfeit chemicals, identifying metabolic pathways, and quantifying metabolic fluxes (Ribeiro *et al.*, 2018; Christensen *et al.*, 2002).

To demonstrate this, we conducted a ^{13}C-isotope labeling experiment by growing *Escherichia coli* on a mixture of glucose (15 mM, 80% 1-$^{13}C_1$-glucose + 20% U-$^{13}C_6$-glucose) and acetate (8 mM, 100% 2-$^{13}C_1$-acetate). Samples were collected each hour to quantify the dynamics of all isotopic forms of acetate

177 (Supplementary data). In the ^{1}H NMR spectra, the methyl protons of acetate may give rise to nine peaks
178 corresponding to the four isotopic species of this compound: a singlet for $^{12}C_2$-acetate, a doublet with a
179 small (6 Hz) ^{2}J coupling constant for 1-$^{13}C_1$-acetate, a doublet with a large (127 Hz) ^{1}J coupling constant
180 for 2-$^{13}C_1$ acetate, and a doublet of doublet for U-$^{13}C_2$-acetate (Figure 3 and S6). The results showed
181 that the total acetate concentration remained stable throughout the experiment (8.5±0.1 mM), which
182 could be interpreted as an acetate flux close to zero (Figure 3C). However, while the provided isotopic
183 form (2-$^{13}C_1$-acetate) initially represents 99% of the total acetate pool, its concentration decreased over
184 time (from 8.5 to 6.5 mM). Concurrently, the concentration of the other labeled forms (derived from
185 glucose) increased (from 0.1 to 1.9 mM). These results support previous findings that acetate is
186 continuously exchanged between cells and their environment, even in the absence of net acetate
187 accumulation in the medium (Enjalbert *et al.*, 2017).
188 As an additional test case, we performed a ^{13}C-labeling experiment to generate a standard sample with
189 known and predictable labeling patterns of metabolites, which enables validation of analytical and
190 processing methods (Millard *et al.*, 2014). We analyzed the labeling pattern of aspartate using a ^{1}H-^{13}C
191 HSQC experiment, which provides information on four isotopic species (Figure 3D). Quantification
192 results were in close agreement with the predicted labeling pattern of aspartate (differences below 1%),
193 demonstrating that MultiNMRFit can accurately quantify the isotopic content of metabolites from ^{13}C-
194 NMR spectra.
195 These examples illustrate the applicability of MultiNMRFit for extracting accurate quantitative isotopic
196 information on metabolites from ^{1}H and ^{13}C NMR spectra.

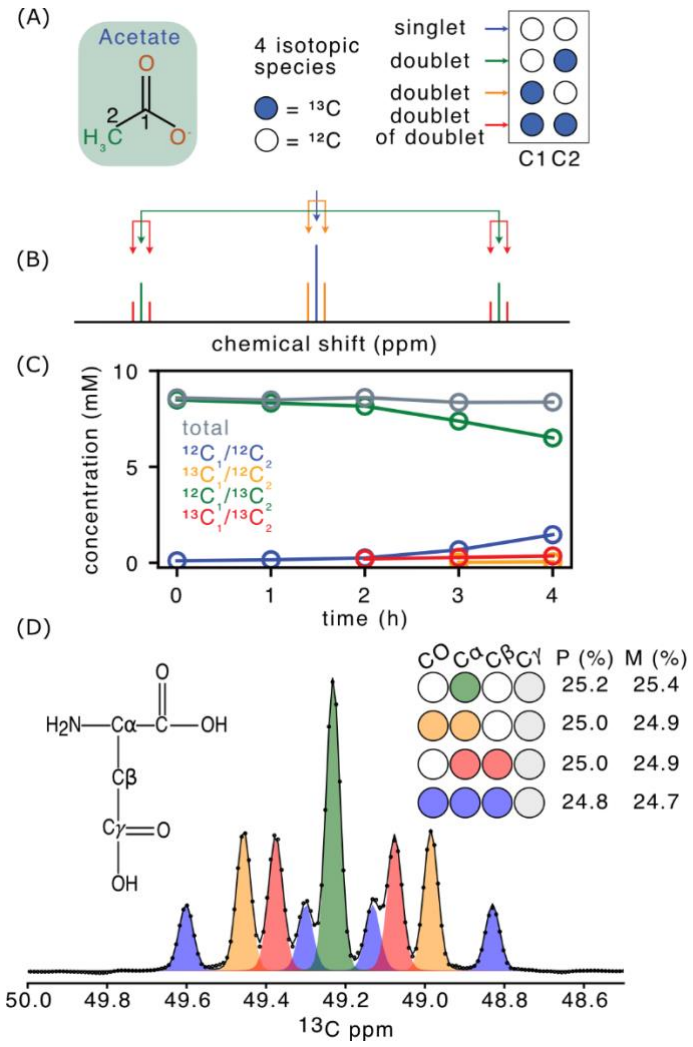


Figure 3. Applications of MultiNMRFit to isotope labeling studies. (A) Isotopic forms of acetate. **(B)** Theoretical ^{1}H spectrum of the methyl protons of acetate, with a singlet for $^{12}C_2$-acetate (blue), a doublet for 1-$^{13}C_1$-acetate (orange), a doublet for 2-$^{13}C_1$ acetate (green), and a doublet of doublets for U-$^{13}C_2$-acetate (red). **(C)** Time-course concentrations of the four isotopic forms of acetate (using the color code shown in panel B) and of the total acetate concentration (in grey). **(D)** Experimental spectrum of ^{13}C-labeled aspartate focused on the Cα signal processed with MultiNMRFit, with the experimental shown as dots, the fitted spectrum as a line, and individual signals as shaded areas. On the right side the different labeling scheme are displayed. The table corresponds to the predicted (P) and the measured (M) fractions of each isotopic species.

Conclusion

We present MultiNMRFit, an open-source Python software designed to extract detailed quantitative parameters from NMR spectra. MultiNMRFit is applicable to spectra acquired from any nucleus, whether as 1D or pseudo-2D spectra, and can handle any type of multiplicity through a flexible system for implementing NMR signal models. MultiNMRFit is accessible to non-expert users through a graphical user interface and can also be used as a Python package, enabling seamless integration into metabolomics and fluxomics workflows. Furthermore, MultiNMRFit holds significant potential for

other NMR applications, including protein studies, the pharmaceutical industry, biotechnology, and chemistry.

Acknowledgements

We thank Marin Martin Mawah and Pauline Rouan for testing previous versions of the software and for insightful discussions. We thank Dr. Edern Cahoreau and Lindsay Peyriga (MetaToul, Toulouse metabolomics & fluxomics facilities, www.metatoul.fr) for access to NMR facilities. MetaToul is part of MetaboHUB, the French National Infrastructure for Metabolomics and Fluxomics (www.metabohub.fr).

Author contributions

PM & **CC** Software, Project administration, Supervision, Conceptualization, Writing - review & editing; **LL** Software**; SD, VG** & **TM** Resources; **GL** Conceptualization, Writing - review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Funding

The PhD thesis of Loïc Le Grégam was funded by the ANR (grant MetaboHUB-ANR-11-INBS-0010), and the PhD thesis of Thomas Gosselin-Monplaisir was funded by the MICA department of INRAE and the Région Occitanie (grant COCA-COLI). This work was also supported by the ANR and the FWF (grant FUNCEMM, ANR-23-CE44-0038). Thomas Gosselin-Monplaisir also benefited from a grant managed by the Agence Nationale de la Recherche under the Investissements d'Avenir Programme (grant ANR-18-EURE-0021).

References

Byrd,R.H. *et al.* (1995) A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.*, **16**, 1190–1208.

Christensen,B. *et al.* (2002) Analysis of flux estimates based on^{13} C-labelling experiments. *European Journal of Biochemistry*, **269**, 2795–2800.

Cox,N. *et al.* (2019) Integrated pH Measurement during Reaction Monitoring with Dual-Reception^{1} H–^{31} P NMR Spectroscopy. *Anal. Chem.*, **91**, 3959–3963.

Diether,M. *et al.* (2019) Systematic mapping of protein-metabolite interactions in central metabolism of *Escherichia coli*. *Molecular Systems Biology*, **15**, e9008.

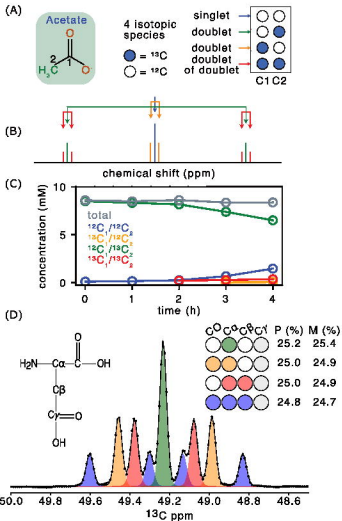
Dudley,J.A. *et al.* (2020) Resolving overlapped signals with automated FitNMR analytical peak modeling. *Journal of Magnetic Resonance*, **318**, 106773.

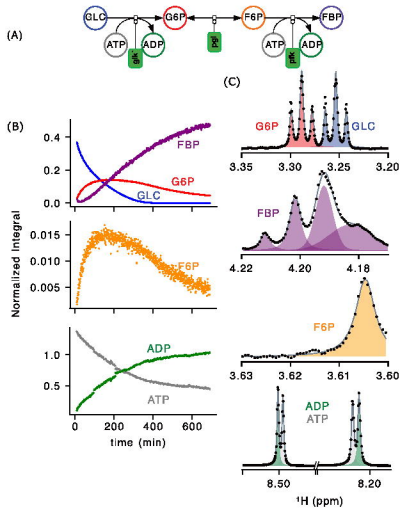
Edison,A.S. *et al.* (2021) NMR: Unique Strengths That Enhance Modern Metabolomics Research. *Anal. Chem.*, **93**, 478–499.

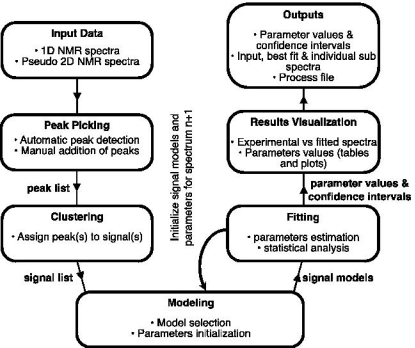
Enjalbert,B. *et al.* (2017) Acetate fluxes in Escherichia coli are determined by the thermodynamic control of the Pta-AckA pathway. *Sci Rep*, **7**, 42135.

Häckl,M. *et al.* (2021) An R-Package for the Deconvolution and Integration of 1D NMR Data: MetaboDecon1D. *Metabolites*, **11**, 452.

Hao,J. *et al.* (2012) BATMAN—an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics*, **28**, 2088–2090.
Hao,J. *et al.* (2014) Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nat Protoc*, **9**, 1416–1427.
Koczula,K.M. *et al.* (2016) Metabolic plasticity in CLL: adaptation to the hypoxic niche. *Leukemia*, **30**, 65–73.
Ludwig,C. and Günther,U.L. (2011) MetaboLab - advanced NMR data processing and analysis for metabolomics. *BMC Bioinformatics*, **12**, 366.
Massiot,D. *et al.* (2002) Modelling one- and two-dimensional solid-state NMR spectra. *Magnetic Reson in Chemistry*, **40**, 70–76.
Millard,P. *et al.* (2014) Isotopic Studies of Metabolic Systems by Mass Spectrometry: Using Pascal's Triangle To Produce Biological Standards with Fully Controlled Labeling Patterns. *Anal. Chem.*, **86**, 10288–10295.
Moco,S. (2022) Studying Metabolism by NMR-Based Metabolomics. *Front. Mol. Biosci.*, **9**, 882487.
Nargund,S. *et al.* (2013) Nuclear Magnetic Resonance Methods for Metabolic Fluxomics. In, Alper,H.S. (ed), *Systems Metabolic Engineering*, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 335–351.
Ribeiro,M.V.D.M. *et al.* (2018) Authenticity assessment of anabolic androgenic steroids in counterfeit drugs by ^{1}H NMR. *Anal. Methods*, **10**, 1140–1150.
Röhnisch,H.E. *et al.* (2021) Improved Automated Quantification Algorithm (AQuA) and Its Application to NMR-Based Metabolomics of EDTA-Containing Plasma. *Anal. Chem.*, **93**, 8729–8738.
Smith,A.A. (2017) INFOS: spectrum fitting software for NMR analysis. *J Biomol NMR*, **67**, 77–94.
Sokolenko,S. *et al.* (2019) Robust 1D NMR lineshape fitting using real and imaginary data in the frequency domain. *Journal of Magnetic Resonance*, **298**, 91–100.
Storn,R. and Price,K. (1997) Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, **11**, 341–359.


(A)
Acetate
4 isotopic species
= 13C
= 12C
singlet
doublet
doublet
doublet of doublet
C1 C2
(B)
chemical shift (ppm)
(C)
concentration (mM)
total
time (h)
(D)
P (%)
M (%)
25.2 25.4
25.0 24.9
25.0 24.9
24.8 24.7
13C ppm


(A)
GLC
G6P
F6P
FBP
ATP
ADP
hk
pgi
pfk
(B)
FBP
G6P
GLC
F6P
ADP
ATP
Normalized Integral
time (min)
(C)
G6P
GLC
FBP
F6P
ADP
ATP
¹H (ppm)


Input Data
• 1D NMR spectra
• Pseudo 2D NMR spectra
Peak Picking
• Automatic peak detection
• Manual addition of peaks
peak list
Clustering
• Assign peak(s) to signal(s)
signal list
Modeling
• Model selection
• Parameters initialization
Initialize signal models and parameters for spectrum n+1
signal models
Fitting
• parameters estimation
• statistical analysis
parameter values & confidence intervals
Results Visualization
• Experimental vs fitted spectra
• Parameters values (tables and plots)
Outputs
• Parameter values & confidence intervals
• Input, best fit & individual sub spectra
• Process file