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

Pierre Millard[1,2,*], Loïc Le Grégam[1,2], Svetlana Dubiley[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 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 without computational expertise. We demonstrate the robustness and flexibility of MultiNMRFit on 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 [1–4]. 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) [5]. 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, may lack the robustness and reproducibility necessary for large-scale or complex NMR studies of metabolic systems. Commercial software, such as TopSpin (Bruker) and Mnova (Mestrelab), provide alternative solutions, offering a user-friendly interface but functioning as proprietary "black boxes" and restricted to a limited set of signal models. Open-source R and Python packages (such as *MetaboDecon1D* [6] and *BATMAN* [7,8]) have also been developed to enhance accessibility and reproducibility. However, they are designed for programmatic use and do not include certain features essential for users without computational expertise. Such features include statistical methods for calculating confidence intervals for extracted parameters, a graphical user interface, and a logging system to save calculation parameters and process information along with the results. Another key challenge is that most existing tools treat each peak as an independent signal. However, NMR signals often consist of multiple peaks arising from homo- and heteronuclear scalar couplings, which are intrinsic properties crucial for identifying metabolites and quantifying their isotopic content. Considering signal multiplicity can also improve parameter estimation in cases of overlapping peaks. However, most current tools do not account for this complexity, which can limit their effectiveness in extracting and utilizing this information. Consequently, the detailed analysis of NMR spectra remains a low throughput task, typically performed by expert users, which can constrain the analysis of large datasets, such as those generated in real-time NMR experiments. The development of user-friendly, open-source software that incorporates signal multiplicity and enables flexible implementation of tailored signal models would significantly enhance the scope, efficiency, 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 can be applied to any nucleus (^{1}H, ^{13}C, ^{15}N, ^{31}P, *etc*.), 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, allowing biologists without computational expertise to use it effectively. 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 shifts and 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) and initializes parameters based on the peak list. Note that a given peak can be included in multiple signals to account for overlapping peaks. 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 [9], with an optional initial refinement step performed by the differential evolution algorithm [10]. 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)* parameters 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).

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 at chemical shift x_i can be described using the following Gaussian function:

$$y_i = I/2 \cdot \left(e^{-(x_i - x_0 + J/2)^2/(2 \cdot {l_w}^2)} + e^{-(x_i - x_0 - J/2)^2/(2 \cdot {l_w}^2)}\right)$$

with parameters x_0 (chemical shift of the signal), I (total signal intensity), J (coupling constant), and l_w (linewidth at half maximum). 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 tutorial on model construction is available in the documentation.

4. Validation and application to metabolomics and isotopic datasets

We validated MultiNMRFit using synthetic datasets (Supplementary data). 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).

Then, we used MultiNMRFit to monitor the dynamic conversion of glucose through the initial steps of glycolysis. Glucose was added to a buffer solution containing TSP (sodium trimethylsilyl propionate, 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 in 1024 1D ^{1}H spectra. All signal parameters were successfully extracted, including for overlapped 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.

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 [11,12]. 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 (Supplementary data). In the ^{1}H NMR spectra, the methyl protons of acetate may give raise to nine peaks corresponding to the four isotopic species of this compound: a singlet for $^{12}C_2$-acetate, a doublet with a small (6 Hz) ^{2}J coupling constant for 1-$^{13}C_1$-acetate, a doublet with a large (127 Hz) ^{1}J coupling constant for 2-$^{13}C_1$ acetate, and a doublet of doublet for U-$^{13}C_2$-acetate. The results showed that the total acetate concentration remained stable throughout the experiment (8.5±0.1 mM), which could be interpreted as an acetate flux close to zero (Figure 2D-G). However, while the provided isotopic form (2-$^{13}C_1$-acetate) initially represents 99% of the total acetate pool, its concentration decreased over time (from 8.5 to 6.5 mM). Concurrently, the concentration of the other labeled forms (derived from glucose) increased (from 0.1 to 1.9 mM). These results support previous findings [13] that acetate is continuously exchanged between cells and their environment, even in the absence of net acetate accumulation in the medium. This example illustrates the applicability of MultiNMRFit for extracting quantitative isotopic information on metabolites, providing meaningful biological insights.

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, Pauline Rouan and Dr. Valeria Gabrielli 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** & **TM** Resources; **GL** Conceptualization, Writing - review & editing.

Competing interests

The authors declare that they have no conflict of interest.

References

[1] K.M. Koczula, C. Ludwig, R. Hayden, L. Cronin, G. Pratt, H. Parry, D. Tennant, M. Drayson, C.M. Bunce, F.L. Khanim, U.L. Günther, Metabolic plasticity in CLL: adaptation to the hypoxic niche, Leukemia 30 (2016) 65–73. https://doi.org/10.1038/leu.2015.187.

[2] S. Moco, Studying Metabolism by NMR-Based Metabolomics, Front. Mol. Biosci. 9 (2022) 882487. https://doi.org/10.3389/fmolb.2022.882487.

[3] S. Nargund, M.E. Joffe, D. Tran, V. Tugarinov, G. Sriram, Nuclear Magnetic Resonance Methods for Metabolic Fluxomics, in: H.S. Alper (Ed.), Systems Metabolic Engineering, Humana Press, Totowa, NJ, 2013: pp. 335–351. https://doi.org/10.1007/978-1-62703-299-5_16.

[4] M. Diether, Y. Nikolaev, F.H. Allain, U. Sauer, Systematic mapping of protein□metabolite interactions in central metabolism of *Escherichia coli*, Molecular Systems Biology 15 (2019) e9008. https://doi.org/10.15252/msb.20199008.

[5] A.S. Edison, M. Colonna, G.J. Gouveia, N.R. Holderman, M.T. Judge, X. Shen, S. Zhang, NMR: Unique Strengths That Enhance Modern Metabolomics Research, Anal. Chem. 93 (2021) 478–499. https://doi.org/10.1021/acs.analchem.0c04414.

[6] M. Häckl, P. Tauber, F. Schweda, H.U. Zacharias, M. Altenbuchinger, P.J. Oefner, W. Gronwald, An R-Package for the Deconvolution and Integration of 1D NMR Data: MetaboDecon1D, Metabolites 11 (2021) 452. https://doi.org/10.3390/metabo11070452.

[7] 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 (2012) 2088–2090. https://doi.org/10.1093/bioinformatics/bts308.

[8] 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 Protoc 9 (2014) 1416–1427. https://doi.org/10.1038/nprot.2014.090.

[9] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A Limited Memory Algorithm for Bound Constrained Optimization, SIAM J. Sci. Comput. 16 (1995) 1190–1208. https://doi.org/10.1137/0916069.

[10] R. Storn, K. Price, Simple and Efficient Heuristic for global Optimization over Continuous Spaces, Journal of Global Optimization 11 (1997) 341–359. https://doi.org/10.1023/A:1008202821328.

[11] M.V.D.M. Ribeiro, N. Boralle, H.R. Pezza, L. Pezza, Authenticity assessment of anabolic androgenic steroids in counterfeit drugs by ^{1}H NMR, Anal. Methods 10 (2018) 1140–1150. https://doi.org/10.1039/C8AY00158H.

[12] B. Christensen, A. Karoly Gombert, J. Nielsen, Analysis of flux estimates based on ^{13}C□labelling experiments, European Journal of Biochemistry 269 (2002) 2795–2800. https://doi.org/10.1046/j.1432-1033.2002.02959.x.

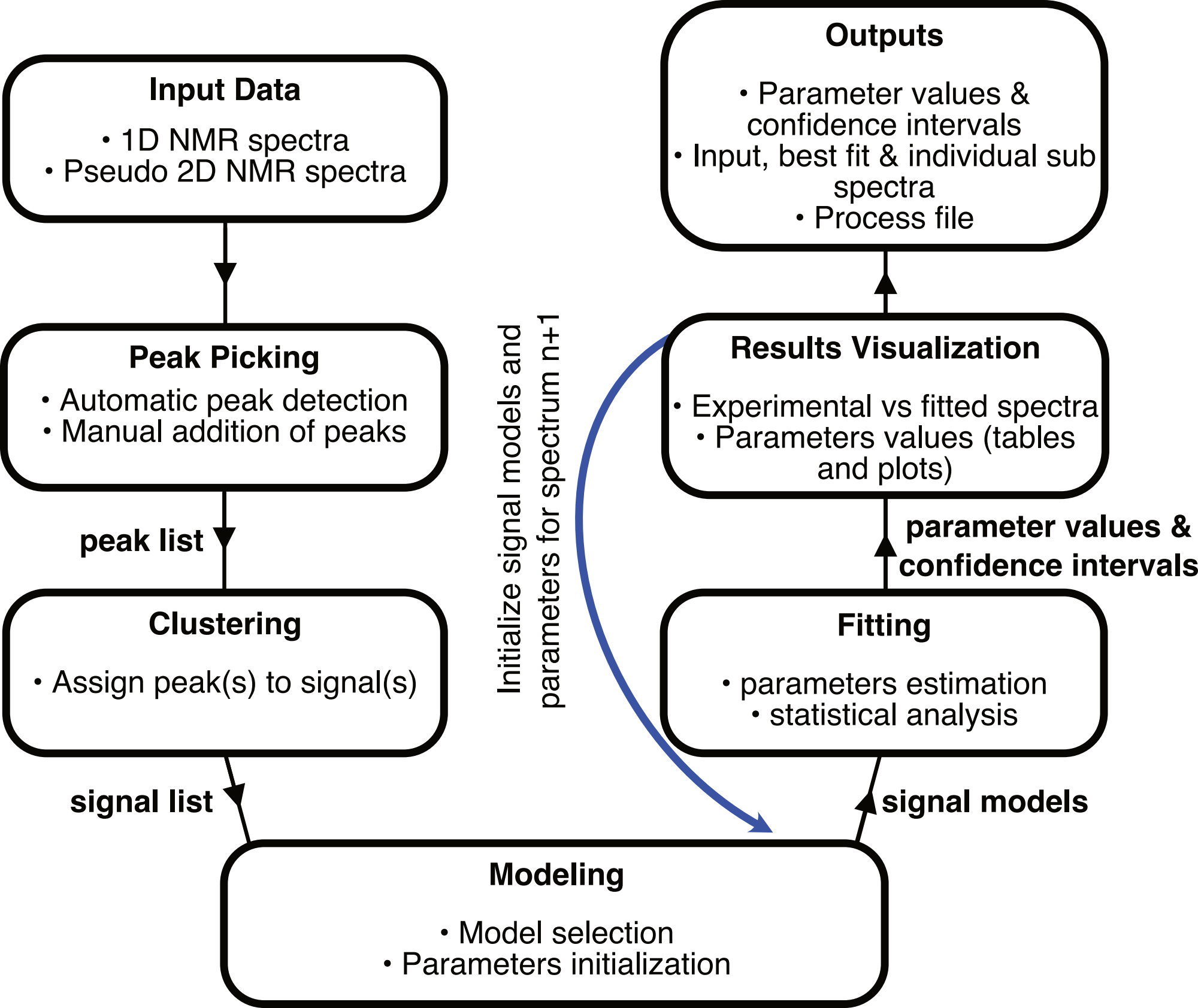
[13] B. Enjalbert, P. Millard, M. Dinclaux, J.-C. Portais, F. Létisse, Acetate fluxes in Escherichia coli are determined by the thermodynamic control of the Pta-AckA pathway, Sci Rep 7 (2017) 42135. https://doi.org/10.1038/srep42135.

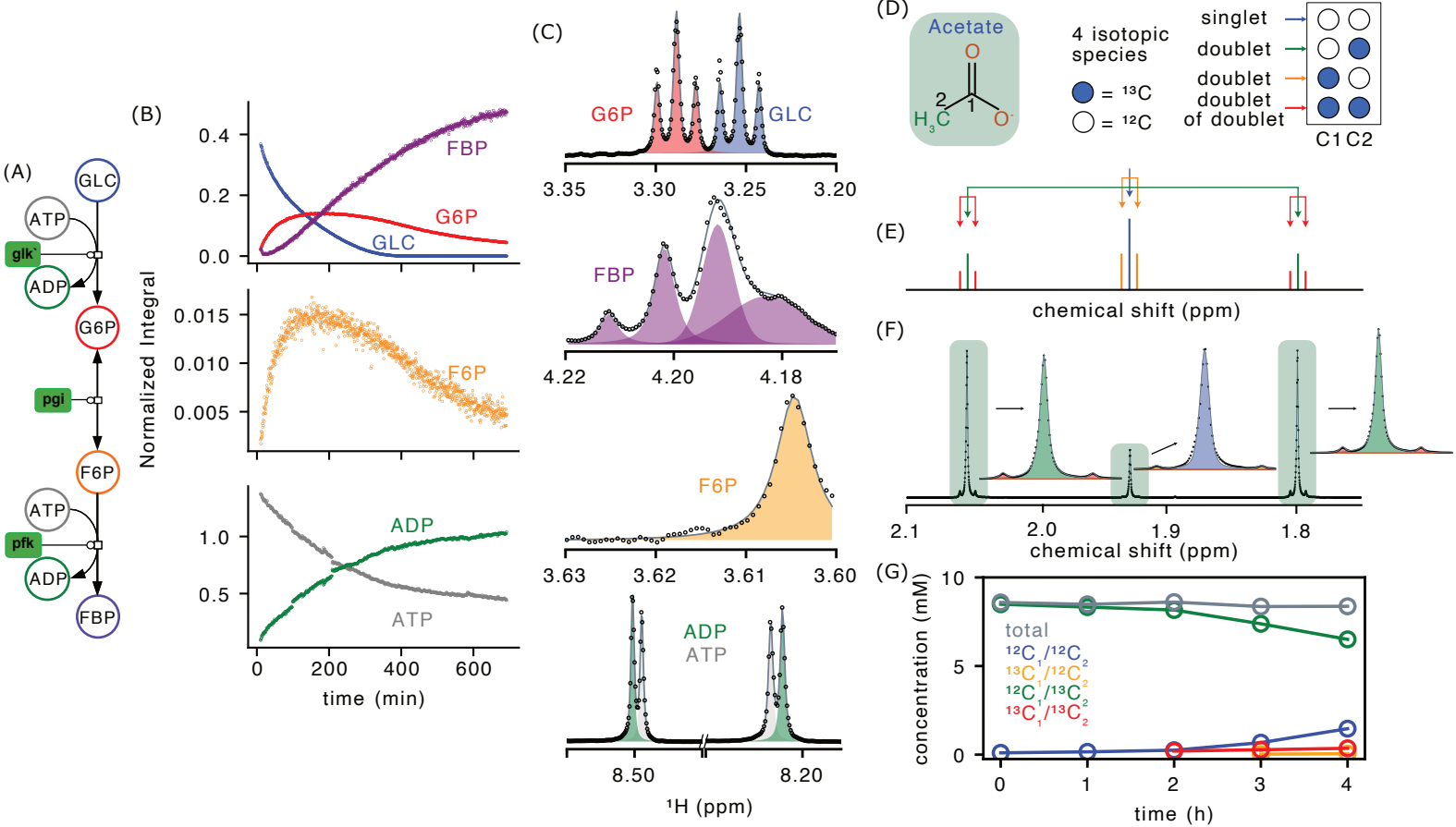
Figure Legends

Figure 1. Flow chart of MultiNMRFit.

Figure 2. Applications of MultiNMRFit to metabolomics and isotope labeling studies. (A) Scheme of the biochemical reactions 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. **(D)** Isotopic forms of acetate. **(E)** 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). **(F)** Example of spectrum processed with MultiNMRFit with the experimental spectrum shown as dots, the fitted spectrum as a line, and individual signals as shaded areas. **(G)** Time-course concentrations of the four isotopic forms of acetate (using the color code shown in panel E) and of the total acetate concentration (in grey).

233


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
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
Initialize signal models and parameters for spectrum n+1


(A)
GLC
ATP
glk
ADP
G6P
pgi
F6P
ATP
pfk
ADP
FBP
(B)
Normalized Integral
FBP
G6P
GLC
F6P
ADP
ATP
time (min)
(C)
G6P
GLC
FBP
F6P
ADP
ATP
¹H (ppm)
(D)
Acetate
4 isotopic species
= ¹³C
= ¹²C
singlet
doublet
doublet
doublet of doublet
C1 C2
(E)
chemical shift (ppm)
(F)
chemical shift (ppm)
(G)
concentration (mM)
total
¹²C₁/¹²C₂
¹³C₁/¹²C₂
¹²C₁/¹³C₂
¹³C₁/¹³C₂
time (h)