

Spin System Modeling of Nuclear Magnetic Resonance Spectra for Applications in Metabolomics and Small Molecule Screening

Hesam Dashti, William M. Westler, Marco Tonelli, Jonathan R. Wedell,
John L. Markley,* Hamid R. Eghbalnia*

National Magnetic Resonance Facility at Madison and BioMagResBank, Department of
Biochemistry, University of Wisconsin-Madison, WI, USA

*John L. Markley; Tel: 608-262-3759, Email: jmarkley@wisc.edu

*Hamid R. Eghbalnia; Tel: 608-261-1167, Email: heghbaln@wisc.edu

Supporting Information

This addendum serves two purposes. It provides additional information concerning the functions and operations of the GISSMO software package. Additionally, it adds details relevant to the analyses presented in the manuscript. The material is organized into five sections. Section I describes the graphical user interface (GUI) for GISSMO and highlights several graphical dialog areas in which important information for the user is displayed. The functions available in the GUI were extensively tested through our derivation of spin system matrices for over 400 metabolites. Section II presents a flowchart for the operational aspects of the algorithms used in GISSMO. Section III discusses functions performed by the GISSMO interface. Section IV explains the output data from GISSMO. Section V presents statistics highlighting the fraction (and number) of metabolites that contain protons in the strongly coupled regime. Supplementary references are listed at the end.

I. Graphical User Interface for Spin Simulation

The graphical user interface (GUI) of GISSMO enables the efficient optimization of starting spin system matrices so that their simulations match experimental ^1H NMR spectra. Figure S1 shows this GUI and illustrates several components and related functionalities supported through the GUI.

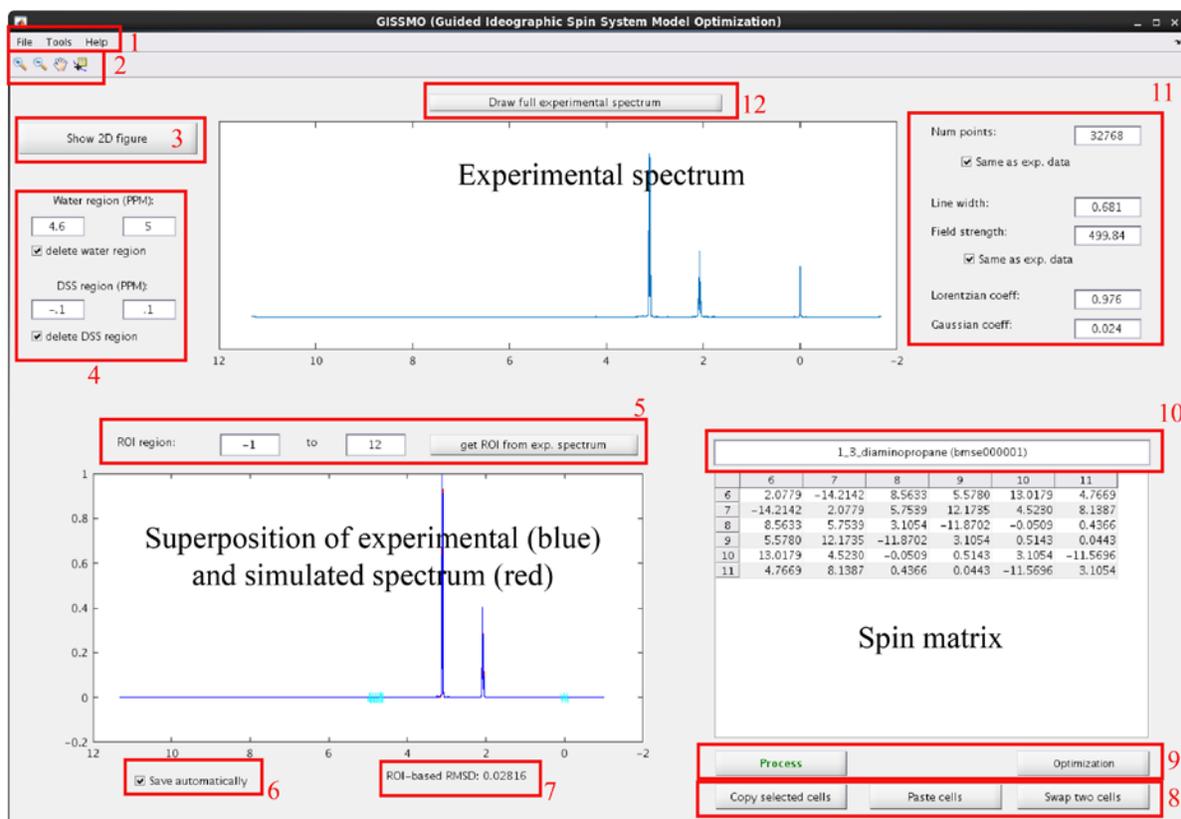


Figure S1. Screen shot of the spin simulation GUI. The example shown is the optimized spin system matrix of 1,3-diaminopropane. After loading a project, the box at the top shows the experimental spectrum to be used in the optimization, and the box at lower right shows the spin system matrix. This matrix can be adjusted manually or through automated optimization processes. The box at lower left displays the superposition of the simulated and experimental spectra for visual inspection of the optimization processes. Different features of the GUI are shown in red boxes (numbered counter-clockwise). **Box 1:** Tabs for creating, opening, or saving a project; 'Tools' tab provides processing options; 'Help' tab links to a manual for processing options. **Box 2:** Buttons (+ or -) for zooming spectra; (hand icon) for panning and moving spectra; and (right-most icon) for displaying chemical shift and amplitude of a point chosen on the spectrum. **Box 3:** When a project is loaded, this button loads a representation of the 2D structure of the compound. **Box 4:** Specifies adjustable regions for removing water and reference (e.g., DSS) signals from the experimental spectrum; removed regions are shown as cyan zig-zag lines on the superimposed spectra. **Box 5:** Specifies the region of interest (ROI). **Box 6:** Option for saving the workspace every time that the 'Process' button is pressed. **Box 7:** Shows the normalized root mean square differences (RMSD₁₀₀) between the experimental and simulated spectra. **Box 8:** Tools for changing values in the spin system matrix. **Box 9:** 'Process' button generates a simulated spectrum from information in the GUI; 'Optimization' button refines the match between the simulated and experimental spectra. **Box 10:** Shows compound name and ID. **Box 11:** Specifies 'points' (discrete evaluations of spectra) and the field strength for the spectral simulation. The line shape parameters (line width, Lorentzian and Gaussian coefficients) can be adjusted manually or optimized by tools in the GUI. **Box 12:** Provides access to the full experimental spectrum.

II. Flowchart for GISSMO

The GISSMO software is used to obtain an optimal spin system matrix. Optimal is defined in a sense that for a given experimental spectrum of a compound and its corresponding starting spin system matrix, GISSMO will seek a spin system matrix that simulates the best match (in the L_2 sense) to the experimental spectrum. The match is computed after applying a convolution with a Voigt line shape to the ideal spectrum. The parameters for the Voigt line shape are allowed to vary in order to obtain the best match. GISSMO implements a computationally efficient approach for simulating spectra. We compared the computational time and accuracy of several approaches, including the regular time evolution of the spin system matrices, as well as using diagonalization of the corresponding Hamiltonian matrix¹ of the compounds prior to evolving the spin system matrices. We took advantage of the sparse data structure to accelerate the computational time. Based on the results of this comparison (shown in Table S1), the diagonalization process was implemented as the core algorithm of GISSMO for simulating a spectrum from its corresponding spin system matrix. Note that this table shows the running-times for simulating a spectrum; with n cycles of optimization, the running time will increase by a factor of n . As indicated in Table S1, the process of simulating a spectrum with a spin system matrix of more than 10 spins is relatively time-consuming, especially considering the iterations needed for optimization. Since numerous compounds have more than 10 spins, we provide an option in the GUI to split the spin system matrix into smaller spin system matrices so as to accelerate the optimization processes. This option for splitting coupling matrices is explained in Section III below.

Table S1. Running time comparison. To compare the running times of the time evolution and Hamiltonian diagonalization methods, we simulated spectra of spin system matrices with up to 11 spins. The software ran on a Linux CentOS 6.5 workstation equipped with a CPU with eight i7 3.4 GHz cores and 10 G of memory.

Num. spins	Running time (sec)	
	Diagonalization	Time evolution
2	0.015	0.610
3	0.140	0.834
4	0.038	2.891
5	0.040	5.554
6	0.094	18.248
7	0.115	53.624
8	0.284	112.813
9	0.629	168.698
10	0.892	235.417
11	4.095	642.201

Our workflow is designed to accelerate the optimization processes (Figure S2). The default workflow uses the GUI's function for "splitting" spin system matrices when more than 10 spins are present, as based on the results in Table S1. The cutoff and the number of submatrices can be modified by the user. The optimization steps in this workflow are explained in Section III.

III. Using GISSMO

As noted above in Section I (Figure S1, Box 1), the menu bar in the GUI contains three tabs: File, Tools, and Help. In this section, we explain these tabs.

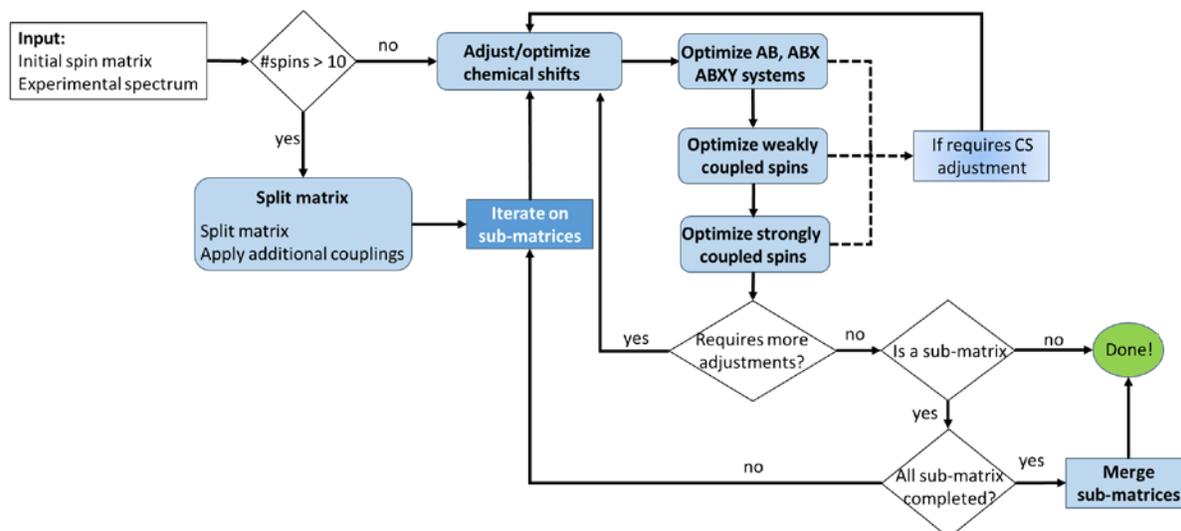


Figure S2. Flowchart for optimizing initial spin system matrices. The dashed lines show a verification step that considers adjusting chemical shift values after every step of optimizing the coupling constants. In addition to visual inspection, the normalized RMSD, reported on the GUI, can be used to determine whether further adjustment is needed or the optimization is completed.

III.1. 'File' for creating, loading, and saving a project file

GISSMO delivers a flexible model for accepting input files in a number of formats. The starting point in the process is a *project file*. In order to create a *project file*, users are required to provide an initial spin system matrix that contains spin names, their corresponding chemical shifts in ppm, and spin-spin coupling constants in Hz. Initial spin system matrices can be provided in any of the following five formats: (I) a CSV file that contains the spin system matrix; (II) a previously generated XML file (by GISSMO); (III) the output from the NMRdb² software; (IV) the output from the Gaussian software package (<http://www.gaussian.com/>); (V) interactive creation of a new matrix—the user will be asked for the number of spins, and the program generates a template for the spin system matrix accordingly. The values of this template can be adjusted later by using the GUI. The experimental spectrum to be used for the optimization of the spin system matrix can be provided in one of the following formats: processed Bruker spectrum, JCAMP-DX³ format (<http://www.jcamp-dx.org/>), NMRPipe⁴ processed spectra, or a CSV file containing the chemical shifts and amplitudes of a spectrum in two columns. The CSV file can be generated by using one of the many NMR spectral processing software packages including NmrGlue⁵ and the MestreNova (<http://mestrelab.com/>). In addition to these essential requirements (spin system matrix and experimental spectrum), the GUI accepts optional information about the compound; the corresponding standard InChI⁶ (<http://www.inchi-trust.org/>) string and a two-dimensional structure file of the compound. This optional information is used to show the structure of the molecule with unique and reproducible atom labels generated by ALATIS.⁷ The user can rename the spin names with these atom labels so that they conform to this reproducible standard.

We have developed an XML data-format representation of spin system matrices for use in storing the optimized spin system matrices. These can be downloaded from our website. Spin system matrices for metabolites can be loaded via the GISSMO GUI and used to simulate spectra at any magnetic field strength (example shown in Figure 6 of the manuscript). The GUI, along with the current set of data, is installed and ready to use as a virtual machine (VM), which can be downloaded from (<http://gissmo.nmrfam.wisc.edu/>). In addition, the software package is available through the NMRBox⁸ project (<https://nmrbox.org/>).

After a spin system matrix is created, the GUI has the capability of loading it for optimization purposes. The workspace and intermediate (or final) data can be saved for further analyses (or deposition to the database) at the user's convenience. Another useful feature of the GUI is the ability to load a database of stored results in the form of a hierarchy of folders. A database can be a folder containing several sub-folders corresponding to different projects. Users can navigate through a database and load/process different compounds.

III.2. 'Tools' for processing a spin system matrix

Rename atoms

The first option under the "Tools" tab of the GUI is "rename atom", which replaces the atom labels with the unique, reproducible ones generated from the ALATIS⁹ web-service. The ALATIS web-service provides a structure file that can be used to generate a 2D representation of the compound of interest (with the reproducible atom labels) through the "Structure to figure" option on the GISSMO web server. These 2D plots can be loaded when creating a *project file* in the GISSMO software package. The "Show 2D figure" option on the main window of the GUI (Figure S1, Box 3) can be used to show the unique atom labels.

Splitting spin system matrix option

This option under the 'Tools' tab facilitates the splitting a spin system matrix, the processing individual sub-matrices, and the merging the submatrices and their corresponding simulated spectra to simulate the experimental spectrum.

Additional coupling constant

As discussed in the manuscript for the example of acetyl-L-carnitine, when spin system matrices contain separate blocks of coupled spins, it is a straightforward process to divide the spin system matrix into smaller matrices that are merged together after being processed individually. However, when spin system matrices do not contain separated blocks of coupled spins, it may be necessary to add additional couplings between the sub-matrices. To accommodate the processing of these types of spin system matrices, we have implemented an option in the GUI for the addition of additional coupling constants in a matrix. An example of this process is shown in Figure S3. In this example, we consider the spin system matrix of L-isoleucine and split the spin system matrix into two sub-matrices: (1) five hydrogen atoms of methyl C1 and methylene C3, (2) five hydrogen atoms of C4, C5 and methyl C2. Vicinal three-bond coupling is expected between H18 and H16 and H17 of methylene C3. However, because the spin system matrix is divided such that the atoms are considered in two different sub-matrices, the couplings between them cannot be considered during the simulation process.

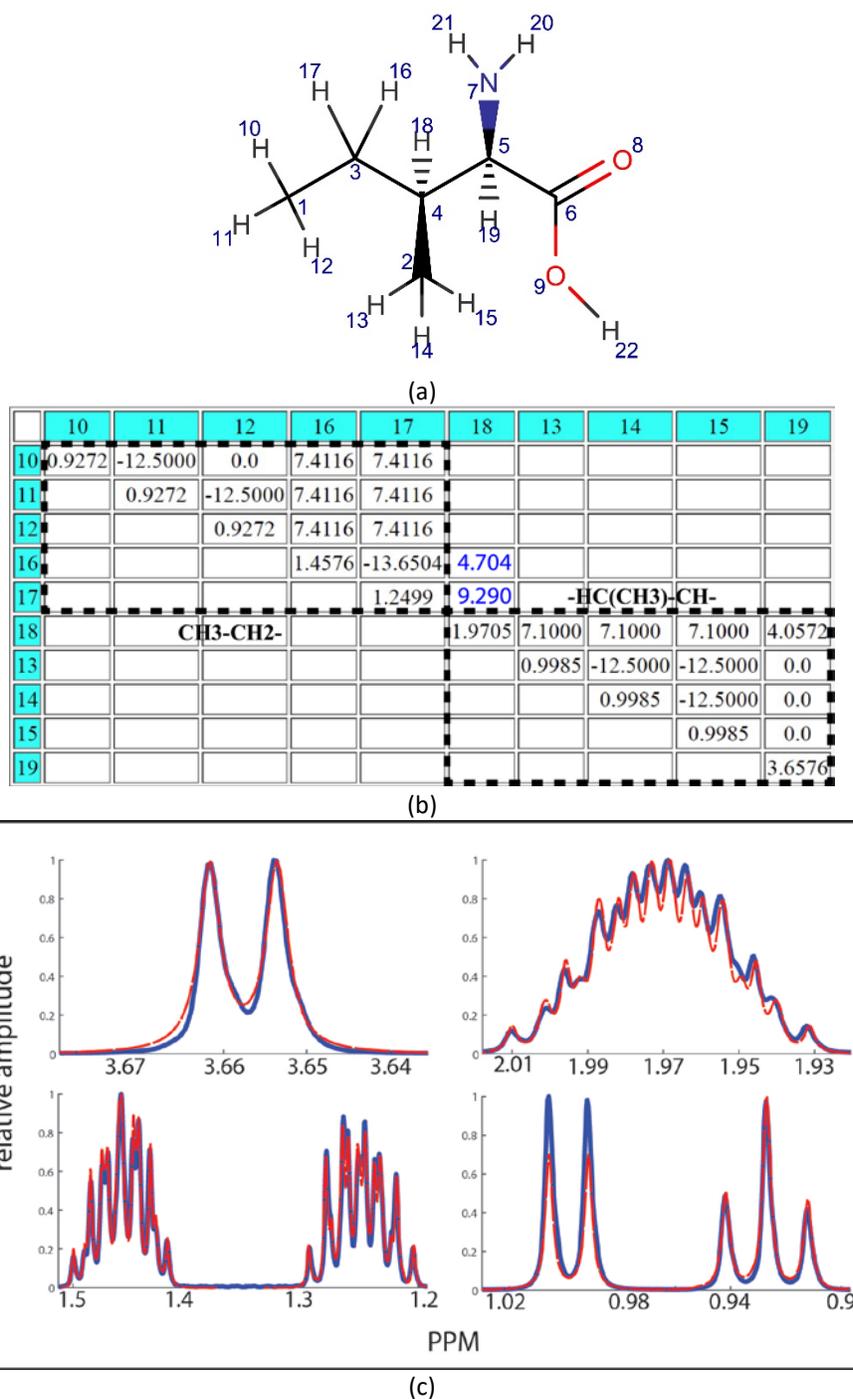


Figure S3. L-isoleucine spin system matrix and its additional couplings. (a) Unique and reproducible atom labels of BMRB entry bmse000041 from ALATIS. (b) The optimized spin system matrix of the molecule. The dashed lines show the sub-matrices considered. From (a) it is clear that H18 in the lower-right sub-matrix is coupled to both H17 and H16 upper-left sub-matrix. Therefore, we applied additional couplings to these atoms and optimized these couplings, which resulted in $J_{H16, H18} = 4.704$ Hz and $J_{H17, H18} = 9.290$ Hz. These additional couplings are shown in blue. (c) Superposition of the simulated spectrum after merging the submatrices (red) and the experimental spectrum from the BMRB entry (blue).

The additional coupling option is used to accommodate these vicinal couplings ($J_{H16, H18}$ and $J_{H17, H18}$), where two additional couplings are considered for H18 and one additional coupling for each of the H16 and H17 atoms. The outcome of this process is shown in Figure S3.

In addition to its use in the splitting process, the additional coupling constants option of the GUI can be used to incorporate couplings from other nuclei that affect the spin simulation. One example is the effect of the phosphate in D-(-)-3-phosphoglyceric acid on hydrogen atoms H12 and H13 (Figure S4). Because ^{31}P is not considered in the hydrogen spin system matrix, the additional coupling option is used to handle the effect of this spin $\frac{1}{2}$ nucleus on the simulated 1D- ^1H spectrum. For this molecule, we considered two additional couplings (one for H12 and one for H13), and the optimization processes resulted in assigning $^3J_{\text{PH}}$ couplings of 5.862 Hz and 5.887 Hz to H12 and H13, respectively.

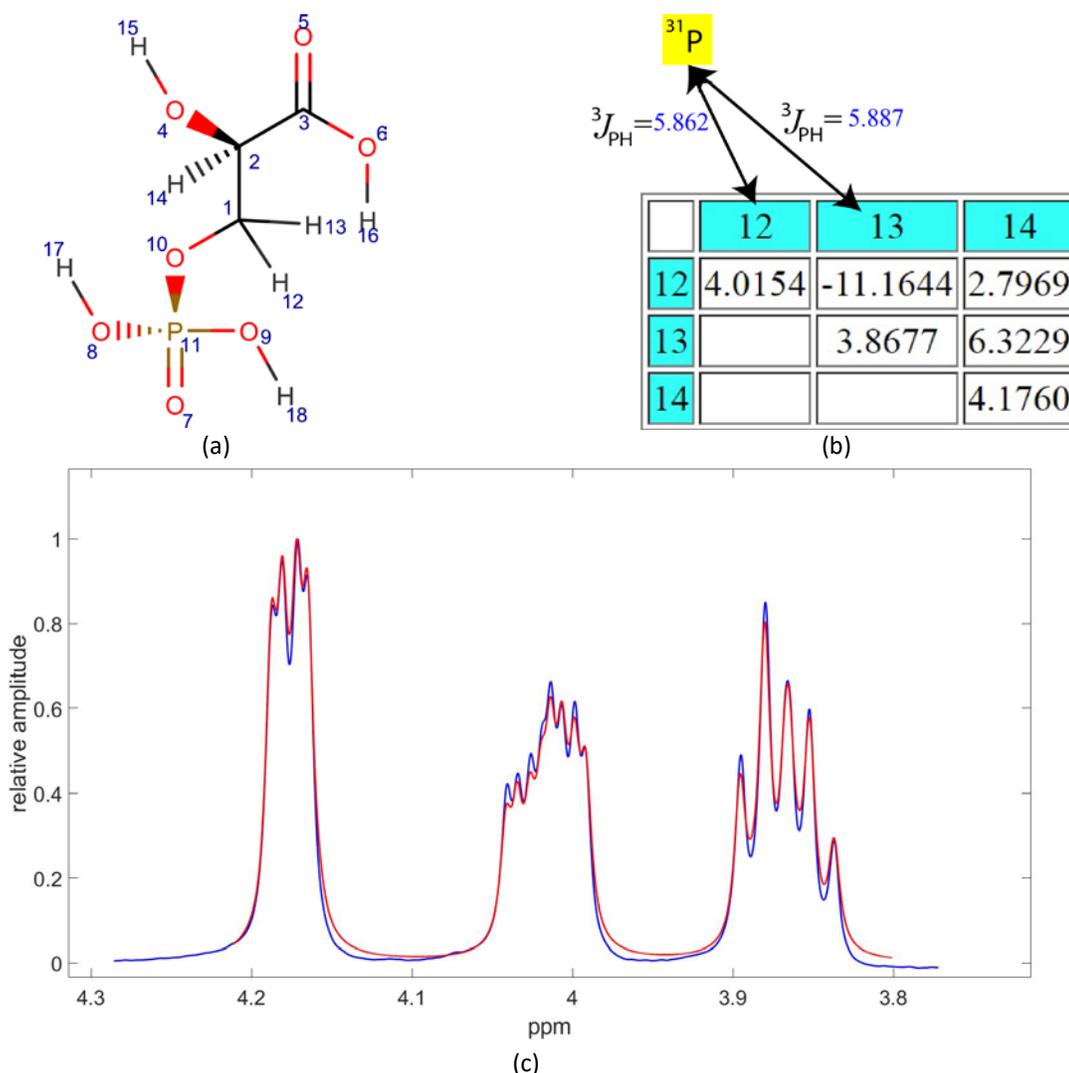


Figure S4. Example of the effect of an additional spin (^{31}P) on the spin system matrix and its simulation. (a) 2D representation of the molecule. (b) Optimized spin system matrix. The additional couplings between P11 and H12 and H13 are indicated by arrows. (c) Superposition of the simulated spectrum derived from the spin system matrix (red) on the experimental spectrum (blue) of the compound from BMRB entry bmse000007.

Optimization

Several options are offered under the “Tools” tab of the GUI to optimize the spin system matrix and its corresponding additional couplings against the experimental spectrum. All of the optimization processes are based on a core simplex optimization as described in the manuscript.

Optimize line shape.

This process optimizes line shape parameters: the line width, Lorentzian and Gaussian coefficients of the Voight function. The line width is optimized to a value between 0 and 5 Hz, and the coefficients in the Voight function are optimized such that their sum always equals 1. These predefined constraints can be adjusted as discussed below, and alternatively, the line shape parameters can be adjusted manually in the main window of the GUI without any predefined constraints.

Optimize spin system matrix.

Optimization of the spin system matrix can be performed in two different ways, as described below.

(I) Optimization on selected cells. Simplex optimization can be performed on a selected set of cells: these can contain both diagonal cells (chemical shift) and off-diagonal cells (spin-spin couplings). The objective of this optimization process is to minimize the differences between the simulated spectrum and the experimental spectrum around the chemical shifts of the corresponding selected spins. In this process, a range of ± 0.1 ppm is used for optimization on the diagonal cells of the spin system matrix. For example, if a coupling constant between spin 1 (3.69 ppm) and spin 4 (5.18 ppm) is selected to be optimized, the optimization process aims for minimizing the RMSD₁₀₀ differences between the simulated and experimental spectra within the ranges of [3.59, 3.79] ppm and [5.08, 5.28] ppm. The “Optimize” button on the main GUI (Figure S1, Box 9) provides easy access to this optimization option from the main window.

(II) Optimization over groups of cells. It is clear that for A₂ or A₃ spin systems (i.e. methylene and methyl groups), the corresponding hydrogen atoms could be assigned to a single chemical shift and would have the same coupling constants to the other spins in a spin system matrix. For these cases, instead of optimizing each of the spins individually, the GUI contains a feature that allows the user to group the cells and optimize them as a group. The same range (± 0.1 ppm) of the frequency domain around the grouped cells will be used during the optimization process.

Optimize chemical shifts (grained).

This option allows users to define a range on the frequency domain to optimize the chemical shift of a spin. This option is usually used for fine-tuning the diagonal cells of a spin system matrix.

Guided optimization.

This option is used to optimize the chemical shifts and coupling constants of AB, ABX and ABXY systems. Whereas the optimization processes considers a wide range of possible strong coupling values, weak coupling optimizations are based on the Bothner-By equation.¹⁰ During this optimization process, users are asked to identify strongly coupled spins (AB) and one (X) or two

(XY) spins that are weakly coupled to AB. Then the GUI considers different possible 3D structural angles between the AB and X(Y) spins and uses the Bothner-By equation to calculate the coupling constants between the spins. The user can choose possible angles between 0° to 360° at intervals of 30° or 60°, and at each angle chosen, simplex optimization is performed on regions (± 0.1 ppm) around the chemical shifts of the identified A, B, X, and Y spins. The angle that results in the minimum RMSD₁₀₀ difference between the simulated and experimental spectra in these regions is used, and the coupling constants in the spin system matrix are adjusted according to this exploratory optimization process.

Optimize additional couplings.

In addition to the optimization options on the spin system matrix, users can also optimize additional couplings considered in the simulation process.

Auxiliary spectrum

Users can employ this option to load a two-dimensional spectrum (in CSV, Bruker, or NMRPipe processed file formats) and select a ¹H trace to be used as the experimental spectrum during the optimization processes. In addition, the auxiliary spectrum facilitates drawing a box around a peak on the 2D spectrum to construct a ¹H trace from the selected 2D peak. This option helps to optimize specific spins against a portion of a ¹H trace.

In practice, there are cases in which it is convenient to simulate and optimize some spins of a spin system matrix against a ¹H trace from the corresponding ¹³C-HSQC spectrum. When the chemical shifts of uncoupled spins overlap and cannot be resolved in the 1D-¹H spectrum, the 2D ¹³C-HSQC spectrum may provide sufficient resolution to distinguish the overlapping spins. Figure S5 shows how a 2D spectrum was used in calculating the spin system matrix of sucrose (BMRB ID bmse000119). The simulated spectrum and the spin system matrix of this compound are available on our website (<http://gissmo.nmrfam.wisc.edu/>). In this example, the perfect HSQC pulse program¹¹ was used to avoid phase and amplitude modulations due to proton-proton scalar coupling.

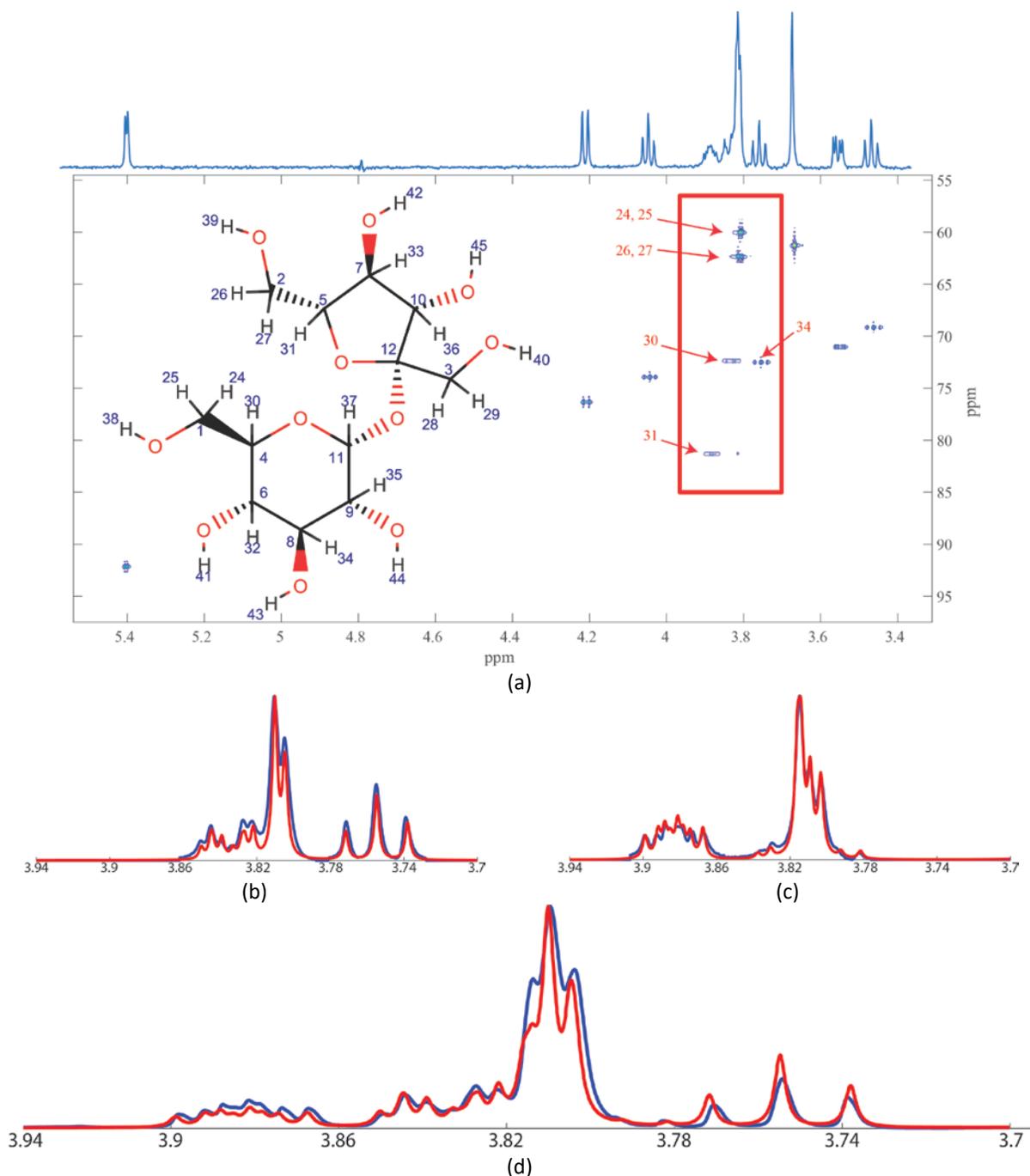


Figure S5. Application of auxiliary 2D spectrum. (a) The unique and reproducible atom names for sucrose generated by ALATIS are shown on the 2D representation of the structure of sucrose. These 14 spins can be divided into two sub-matrices: sub-matrix 1 containing spins 24, 25, 30, 32, 34, 35, 37, and sub-matrix 2 containing spins 36, 33, 31, 26, 27, 28, 29. The red box on the 2D ^{13}C -HSQC spectrum indicates spins whose proton chemical shifts overlap on the 1D spectrum. By using the traces from the HSQC spectrum (blue lines), it was possible to separate the overlapping spins and optimize them individually (red lines) as shown for the domain of [3.7, 3.94] ppm in (b) for spins 24, 25, 30, and 34 from the first sub-matrix, and in (c) for spins 26, 27, and 31. (d) After merging the sub-matrices the simulated spectrum reached RMSD_{100} of 0.05361 against the original 1D- ^1H spectrum from BMRB. In these plots, the optimized and experimental data are normalized based on their maximum amplitude.

Auxiliary tools

Get spectral information.

This option provides utilities for calculating the average distance between selected peaks from the experimental spectra in both ppm and Hz. This option can be used to estimate the chemical shifts and coupling constants of spins and to manually update the spin system matrix on the main GUI window. Figure S6 shows an example of this option of the GUI.

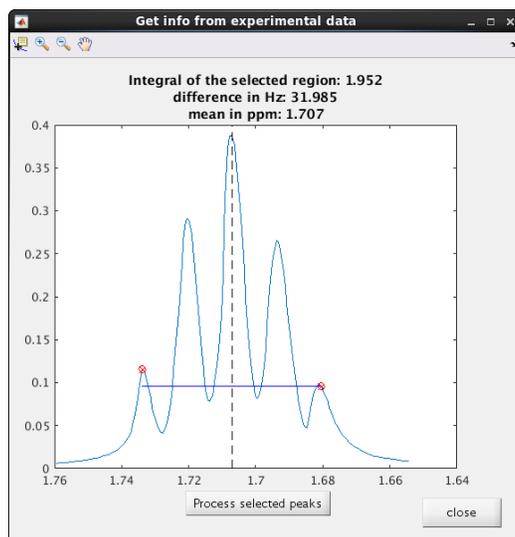


Figure S6. Report on experimental spectra. The red points are selected manually; and after processing, the selected peaks, the vertical dashed line shows the mean of the selected points. The information on the top of the figure can be used to determine the number of spins giving rise to the peaks (the integral of the region between the points), their corresponding scalar coupling (their differences in Hz), and their relative chemical shifts (based on the mean in ppm).

Drawing Tools.

Two options are available for drawing the structure of a compound of interest: (a) “Draw InChI string graph” accepts an InChI string and draws a graph representing the structure captured in the standard InChI string. (b) “Plot a molecule” accepts a 2D or 3D structure file and opens it in the user’s native web-browser by using the Jmol program (<http://jmol.sourceforge.net/>).

Notes

This option can be used to add notes to projects. These notes are stored in the XML *project file*, and the Note option on the GUI provides an interface to open/edit these notes.

Parameters

All of the predefined constant parameters in the GUI can be accessed and adjusted through this option. These parameters include the optimization parameters (the range for exploring strong and weak coupling constants, the domain and interval used in Bothner-By equation, or the line shape optimization range) and the frequency range for calculating RMSD₁₀₀ between experimental and simulated spectra.

III.3. 'Help' tab

This tab provides information about the running version of the GISSMO software program and also provides a hyperlink to the GISSMO manual on GISSMO's web server.

IV. Output formats

A project file can be written as an XML file, which can be loaded using the GUI, or as an HTML file, which can be opened by web-browsers. Although the XML file is machine readable, the HTML file reports the spin system matrix and the corresponding parameters (field strength, Voigt coefficients, etc.) in a human readable format. In addition to these ways of exporting the workspace, the GUI has options to save the simulated spectra in CSV format (chemical shifts and amplitudes in two consecutive columns) or as an image in different formats (eps, tiff, jpeg, etc.). The superposition of the simulated and experimental spectra can be exported with adjustable ppm and amplitude offsets in different image formats.

V. Frequency of metabolites having strongly coupled spins

The BMRB archives maintain 1D-¹H reference NMR spectra for several hundred metabolites. More than 50% of the ¹H NMR spectra were acquired using spectrometers operating at 500 MHz (Table S2). We considered the base magnetic field for the purpose of identifying compounds with strong couplings as that resonating ¹H at 500 MHz.

Table S2. Number of 1D-¹H spectra of small molecules available from the BMRB archive collected at different field strengths. Only tabulated are spectra collected at compound concentration = 100 mM, temperature = 298 K, and pH = 7.4.

Field strength yielding 1H frequency (MHz)	Number of 1D- ¹ H spectra
400	237
500	718
600	84

The ratio of the differences in the chemical shifts of two spins (in a spin system) divided by the coupling constant between the spins is used as an indicator of strong coupling ($\Delta\delta/J \leq 5$) - $\Delta\delta$ and J are in units of Hz. By using GISSMO to analyze over 440 metabolites commonly identified in biological fluids, we found that close to 150 of them (approximately one-third) contain at least one pair of strongly coupled spins. As expected, the $\Delta\delta/J$ ratio is field dependent and decreases at higher magnetic fields (Figure S7). However, the results make it clear that a significant number of metabolites have strongly coupled spins at the magnetic fields used in NMR metabolomics experiments, and, thus, field-dependent changes to peak patterns are to be expected.

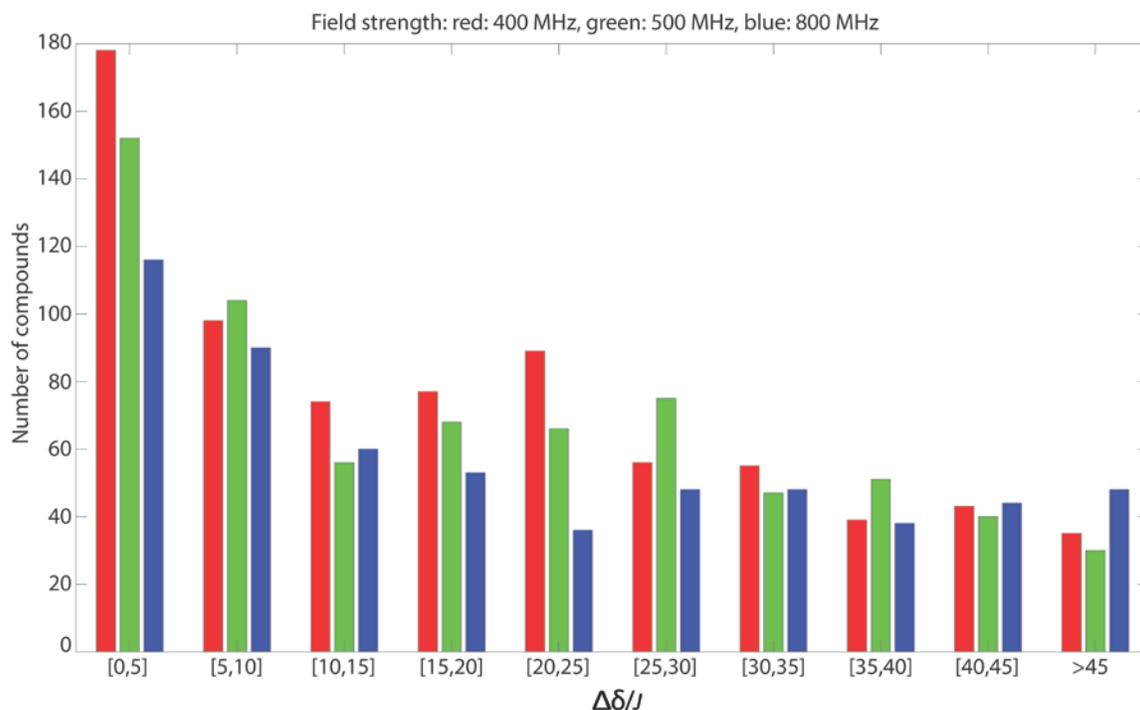


Figure S7. Histogram of the ratio ($\Delta\delta/J$). In these plots, the x-axis shows the domain of ratios with a bin-size of 5. The y-axis indicates the number of metabolites that have at least one pair of spins with the ratio indicated on the x-axis. The ratios were calculated for ^1H data acquired at field strengths corresponding to 400 MHz (red bars), 500 MHz (green bars), and 800 MHz (blue bars).

Supplemental References

- (1) Pochapsky, T. C.; Pochapsky, S. *NMR for physical and biological scientists*, 1 ed.; Taylor & Francis, 2006, p 394.
- (2) Binev, Y.; Marques, M. M. B.; Aires-de-Sousa, J. *Journal of chemical information and modeling* **2007**, *47*, 2089-2097.
- (3) Davies, A. N.; Lampen, P. *Applied Spectroscopy* **1993**, *47*, 1093-1099.
- (4) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J Biomol NMR* **1995**, *6*, 277-293.
- (5) Helmus, J. J.; Jaroniec, C. P. *Journal of biomolecular NMR* **2013**, *55*, 355-367.
- (6) Heller, S. R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. *Journal of Cheminformatics* **2015**, *7*, 23.
- (7) Dashti, H.; Westler, W. M.; Markley, J. L.; Eghbalnia, H. R. *Scientific data* **2017**, *4*, 170073.
- (8) Maciejewski, M. W.; Schuyler, A. D.; Gryk, M. R.; Moraru, I.; Romero, P. R.; Ulrich, E. R.; Eghbalnia, H. R.; Livny, M.; Delaglio, F.; Hoch, J. *Biophysical Journal* **2017**, *112*.
- (9) Dashti, H.; Westler, W. M.; Markley, J. L.; Eghbalnia, H. R. *Scientific Data* **2017**, *4*, 170073.
- (10) Bothner-By, A. A. In *Advances in Magnetic and Optical Resonance*, John, S. W., Ed.; Academic Press, 1965, pp 195-316.
- (11) Castañar, L.; Sistaré, E.; Virgili, A.; Williamson, R. T.; Parella, T. *Magnetic Resonance in Chemistry* **2015**, *53*, 115-119.