Published in final edited form as:

Anal Chem. 2018 September 18; 90(18): 10646–10649. doi:10.1021/acs.analchem.8b02660.

Applications of Parametrized NMR Spin Systems of Small Molecules

Hesam Dashti[†], **Jonathan R. Wedell**[‡], **William M. Westler**[†], **Marco Tonelli**[†], **David Aceti**[†], **Gaya K. Amarasinghe**[§], **John L. Markley**[*,†,‡], and **Hamid R. Eghbalnia**[*,†,‡]

[†] National Magnetic Resonance Facility at Madison, University of Wisconsin, Madison, Wisconsin 53706, United States

[‡] BioMagResBank, Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, United States

[§] Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, United States

Abstract

We have developed technology for producing accurate spectral fingerprints of small molecules through modeling of NMR spin system matrices to encapsulate their chemical shifts and scalar couplings. We describe here how libraries of these spin systems utilizing unique and reproducible atom numbering can be used to improve NMR-based ligand screening and metabolomics studies. We introduce new Web services that facilitate the analysis of NMR spectra of mixtures of small molecules to yield their identification and quantification. The library of parametrized compounds has been expanded to cover simulations of ^{1}H NMR spectra at a variety of magnetic fields of more than 1100 compounds, included are many common metabolites and a library of drug-like molecular fragments used in ligand screening. The compound library and related Web services are freely available from http://gissmo.nmrfam.wisc.edu/.

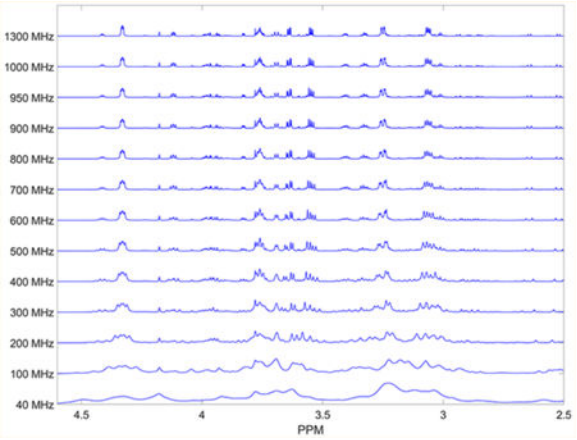
SUMMARY

The current release of the GISSMO library contains optimized spin system matrices corresponding to more than 1100 compounds. These include common metabolites and molecular fragments routinely utilized in ligand screening investigations. Recently implemented Web services provide tools for identifying compounds from ^{1}H NMR spectra collected at a wide variety of magnetic field strengths and for verifying analyses of compounds and their concentrations in ^{1}H NMR spectra of mixtures.

Abstract

[*]**Corresponding Authors** Phone: 608-262-3759. jmarkley@wisc.edu. [*]Phone: 608-261-1167. heghbaln@wisc.edu.

The authors declare no competing financial interest.



NMR spin system matrices of chemical compounds enable parametrization of their spectral features under different conditions, including magnetic field strength. We have developed an approach called Guided Ideographic Spin System Model Optimization (GISSMO)[1] that facilitates the derivation of NMR spin system matrices by optimizing the fit between an experimental one-dimensional ^{1}H NMR spectrum and the theoretical spin system matrix. GISSMO offers a graphical user interface (GUI) containing several optimization modules that semiautomate the process. This GUI supports the splitting of spin system matrices so that portions can be optimized prior to merging. It also allows for the consideration of couplings between ^{1}H nuclei and other types of nuclei (e.g., ^{31}P), and the GUI supports the use of auxiliary two-dimensional spectra in refining chemical shifts and couplings. GISSMO utilizes the Atom Label Assignment Tool using InChI String[2] (ALATIS) software package to generate unique and reproducible atom nomenclature within the library of parametrized compounds. The ALATIS naming convention has been adopted by the BMRB[3] small molecule NMR archive and by the NMReData initiative,[4] and ALATIS is being utilized to federate a variety of small molecule databases.[2]

CURRENT STATUS OF THE GISSMO SPIN SYSTEM MATRIX LIBRARY

The initial release of the GISSMO GUI reported spin system matrices for about 400 compounds from the BMRB archive. In the current release, the number of entries from BMRB has increased to 511. In addition, the library now contains optimized spin system matrices for 664 molecular fragments from the Maybridge Ro3 fragment library (https://www.maybridge.com/), which are routinely utilized in NMR-based ligand screening for drug development. For every GISSMO entry in the library, we utilized ALATIS to identify its corresponding HMDB[5-8] entry and applied an in-house text-processing module to extract the associated biolocations of the compounds. The result, which is available on the GISSMO Web site, indicates that 338 out of 511 (66%) of the compounds have been observed in at least one cellular or tissue location. This library of optimized spin system matrices is publicly available from the GISSMO's Web site. Figure 1 displays a histogram showing compounds as a function of the number of NMR spins analyzed and a histogram with a validation of the optimized spin system matrices in terms of the normalized $RMSD_{100}$[1] between the experimental and fitted 1D-^{1}H NMR spectra.

SPECTRAL SIMULATIONS AT A VARIETY OF NMR FIELD STRENGTHS

Because the ^{1}H NMR spectra of many compounds are not strictly first order, a major challenge of databases containing reference NMR data is how to deal with their dependence on field strength. One approach is to collect reference spectra at a single field strength and require that this field strength be used for collecting experimental spectra. Alternatively, some reference databases contain data collected at more than one field strength. A more general solution to this challenge is to derive parametric representations of ^{1}H NMR spectra in terms of spin system matrices, which can be utilized to generate spectra at any desired magnetic field strength. We have taken advantage of this feature of spin system matrices to produce spectra of all compounds in the GISSMO database at ^{1}H resonance frequencies of 40, 100, 200, 300, 400, 500, 600, 700, 750, 800, 900, 950, 1000, and 1300 MHz. These spectra, which cover the range of magnetic fields used in NMR spectroscopy and MRI, can be accessed and downloaded from the GISSMO Web site. To display the spectra, we utilize the open source graphing library Plotly (https://plot.ly/), which provides interactive zooming and pan features for visual investigation of the spectra. The downloadable spectra are formatted in two columns (ppm and amplitude) as a comma separated file (CSV). These files can be loaded into NMR software programs such as Mestrelab Mnova (http://mestrelab.com) and NMRFx,[9] which are accessible through the NMRbox project[10] (https://www.nmrbox.org/). In addition, these files can be easily loaded using any scripting and programming languages.

ADVANTAGES OF SPECTRA SIMULATED FROM SPIN SYSTEM MATRICES AS REFERENCES IN SPECTRAL PROFILING

Spectral peak pattern matching is a common approach for profiling of small molecules. This approach involves peak picking the NMR spectra and subsequently using the resulting chemical shifts to search for matching peak patterns in a small molecule reference database. This process of identification relies strongly on the spectral peak lists, especially those archived in the reference databases. Because these databases utilize peak picking programs on their archived experimental spectra, the accuracy and reliability of the reference spectral peaks depend on a variety of factors including the correct identification of the reference compound, the presence of impurities, spectral artifacts from water signal suppression or other sources, and the signal-to-noise ratio of the experimental spectra. By contrast, the reference spectra generated from parametrized spin system matrices are noise-free, contain no impurity peaks, and are free from spectral artifacts. In addition, they serve to validate the identity of the reference compound.

PEAK LISTS AND SEARCHING

Generating peak lists from these highly refined spectra can be readily achieved by standard peak picking approaches. We utilized the peak picking modules of the Mnova program, under both the “Standard” and “Global Spectral Deconvolution (GSD)” options, to generate peak lists from the entire library of compounds parametrized by GISSMO at all aforementioned magnetic field strengths. We used these to generate interactive lists of chemical shifts and peak amplitudes (standard or GSD) for each compound at a selectable

magnetic field strength, which are available on the GISSMO Web site. Clicking on a chemical shift brings up a region of the ^{1}H NMR spectrum with the corresponding peak identified.

A “Peak Search” module is now available on the GISSMO Web site. The search is linked to a PostgreSQL (https://www.postgresql.org/) database containing all curated spectral peak lists. The user can query the database by specifying one or more peak positions in ppm (standard or GSD) with a specified tolerance at a selected magnetic field strength. The result is a list of compounds associated with the queried peaks within the specified matching tolerance. The compounds returned from the query are sorted based on the minimum differences between the queried peaks and those archived in the database. Users can investigate the resulting compounds by browsing the corresponding GISSMO Web pages, download the GISSMO entries, or download the output in CSV format. Additional details regarding the results are provided on the Web site.

SIMULATED SPECTRA OF COMPOUND MIXTURES

^{1}H NMR is widely used to identify and quantify metabolites in biological fluids or tissue extracts. We are developing tools based on spin system matrix parametrizations for use in NMR- based metabolomics. As a first step, we have created a module for simulating ^{1}H NMR spectra of compound mixtures. This module utilizes the archived spin system matrices in the GISSMO library that have been manually optimized against the reference spectra. As its input, the module accepts a list of these compounds and their corresponding concentrations. Users can upload a CSV file or provide them through an interactive webpage on GISSMO’s Web site. The spectrum of the simulated mixture can be downloaded as a two-column CSV file. A user-controlled slider enables adjustments of the components in the simulated spectrum to achieve the best match to an experimental spectrum containing the same components. The mixture simulation module uses the Plotly library to display the spectra. As an example, Figure 2 shows a simulated NMR spectrum generated from a list of metabolites and their concentrations reported in a published metabolomics study.[11]

The mixture module accepts an optional experimental spectrum in CSV format and displays an overlay of the uploaded experimental and simulated mixtures on the Web site. Comparisons of simulated spectra with experimental spectra of mixtures can be used to validate prior analyses of compounds present and their relative concentrations. If experimental spectra of mixtures have been collected at multiple fields, the simulation can provide an additional layer of validation. Note that experimental spectra of biological samples may reflect effects of macromolecules or contaminants in samples or molecular interactions between compounds. These effects are not represented by the optimized spin systems of reference spectra. We are currently working to expand this first step of processing mixture spectra by developing ways of optimizing spin system matrices against experimental spectra of biological samples.

ACKNOWLEDGMENTS

This study made use of the National Magnetic Resonance Facility at Madison, WI, which is supported by National Institutes of Health (NIH) Grant P41GM103399. Acquisition of experimental data were supported by NIH Grants

GM109046, AI120943, AI109664, and AI123926. H.D., H.R.E., and J.R.W. are supported in part by the National Center for Biomolecular NMR Data Processing and Analysis, which is supported by NIH Grant P41GM111135.

REFERENCES

(1). Dashti H; Westler WM; Tonelli M; Wedell JR; Markley JL; Eghbalnia HR Anal. Chem. 2017, 89, 12201–12208. [PubMed: 29058410]

(2). Dashti H; Westler WM; Markley JL; Eghbalnia HR Sci Data 2017, 4, 170073. [PubMed: 28534867]

(3). Ulrich EL; Akutsu H; Doreleijers JF; Harano Y; Ioannidis YE; Lin J; Livny M; Mading S; Maziuk D; Miller Z; Nakatani E; Schulte CF; Tolmie DE; Kent Wenger R; Yao H; Markley JL Nucleic Acids Res. 2008, 36, D402–D408. [PubMed: 17984079]

(4). Pupier M; Nuzillard JM; Wist J; Schlörer NE; Kuhn S; Erdelyi M; Steinbeck C; Williams AJ; Butts C; Claridge TDW; Mikhova B; Robien W; Dashti H; Eghbalnia HR; Fares C; Adam C; Kessler P; Moriaud F; Elyashberg M; Argyropoulos D; et al. Magn. Reson. Chem. 2018, 56, 703–715. [PubMed: 29656574]

(5). Wishart DS; Jewison T; Guo AC; Wilson M; Knox C; Liu Y; Djoumbou Y; Mandal R; Aziat F; Dong E; Bouatra S; Sinelnikov I; Arndt D; Xia J; Liu P; Yallou F; Bjorndahl T; Perez-Pineiro R; Eisner R; Allen F; et al. Nucleic Acids Res. 2012, 41, D801–807. [PubMed: 23161693]

(6). Wishart DS; Feunang YD; Marcu A; Guo AC; Liang K; vázquez-Fresno R; Sajed T; Johnson D; Li C; Karu N; Sayeeda Z; Lo E; Assempour N; Berjanskii M; Singhal S; Arndt D; Liang Y; Badran H; Grant J; Serra-Cayuela A; et al. Nucleic Acids Res. 2018, 46, D608–D617. [PubMed: 29140435]

(7). Wishart DS; Knox C; Guo AC; Eisner R; Young N; Gautam B; Hau DD; Psychogios N; Dong E; Bouatra S; Mandal R; Sinelnikov I; Xia J; Jia L; Cruz JA; Lim E; Sobsey CA; Shrivastava S; Huang P; Liu P; et al. Nucleic Acids Res. 2009, 37, D603–D610. [PubMed: 18953024]

(8). Wishart DS; Tzur D; Knox C; Eisner R; Guo AC; Young N; Cheng D; Jewell K; Arndt D; Sawhney S; Fung C; Nikolai L; Lewis M; Coutouly MA; Forsythe I; Tang P; Shrivastava S; Jeroncic K; Stothard P; Amegbey G; et al. Nucleic Acids Res. 2007, 35, D521–526. [PubMed: 17202168]

(9). Norris M; Fetler B; Marchant J; Johnson BA J. Biomol NMR 2016, 65, 205–216. [PubMed: 27457481]

(10). Maciejewski MW; Schuyler AD; Gryk MR; Moraru I; Romero PR; Ulrich ER; Eghbalnia HR; Livny M; Delaglio F; Hoch J Biophys. J. 2017, 112, 1529. [PubMed: 28445744]

(11). Whigham LD; Butz DE; Dashti H; Tonelli M; Johnson LK; Cook ME; Porter WP; Eghbalnia HR; Markley JL; Lindheim SR; Schoeller DA; Abbott DH; Assadi-Porter FM Curr. Metabolomics 2014, 1, 269–278.

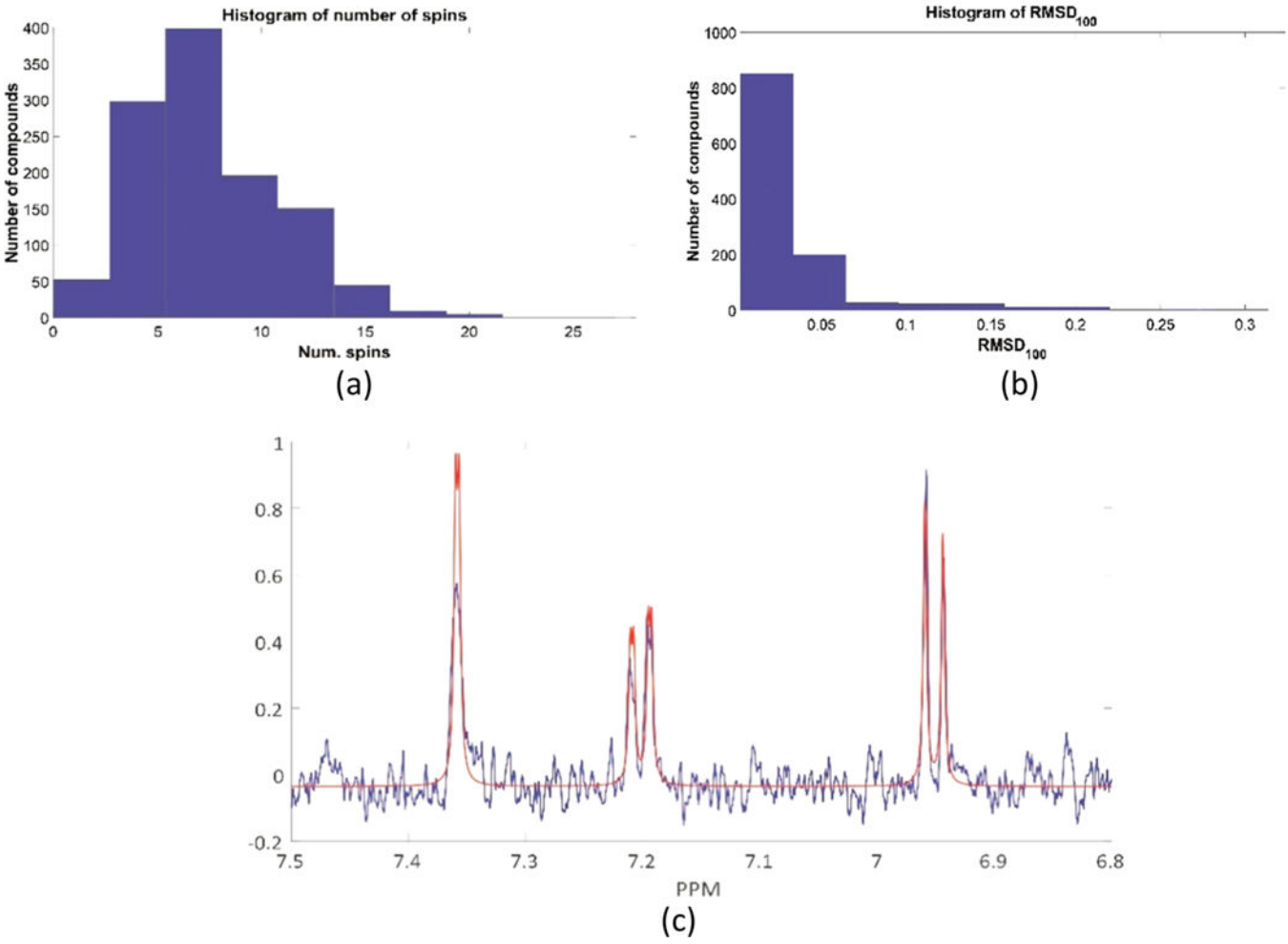


Figure 1.

(a) Histogram of compounds in the GISSMO library as a function of their number of parametrized spins. There is no limitation on the number of spins that the GISSMO tools can handle, although compounds with higher numbers of spins require more time to optimize. (b) Histogram of normalized $RMSD_{100}$ of the optimized spin system matrices. The majority of compounds (93%) have RMSD less than 0.1. Higher RMSDs typically result from optimization of spin system matrices fitted to experimental spectra with low signal-to-noise ratio (SNR). (c) Overlay of a section of the experimental (blue) and simulated (red) spectra of the compound (4-chloro-2-(trifluoromethoxy)aniline) that exhibited the largest normalized $RMSD_{100}$ (0.304). The poor fit resulted from a combination of poor signal-to-noise and partial relaxation effects.

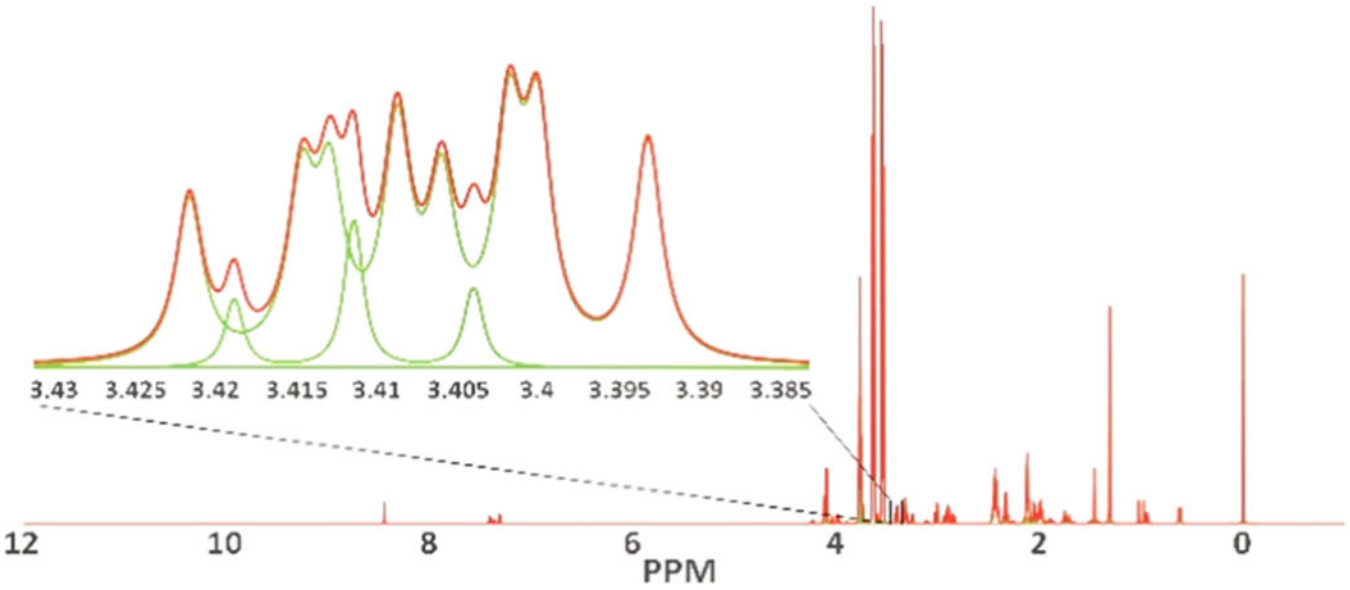


Figure 2.

Sample output of the mixture module. The red line indicates the mixture spectrum and the green lines show GISSMO's spectral representation of the compounds. The zoomed region shows a portion of taurine and proline spectra (in green) that resulted in the mixture spectra shown in red.