

MA'AT: A Web-Based Application to Determine Rotamer Population Distributions in Solution from Nuclear Magnetic Resonance Spin-Coupling Constants

Reagan J. Meredith, Luke Sernau, and Anthony S. Serianni*



Cite This: *J. Chem. Inf. Model.* 2022, 62, 3135–3141



Read Online

ACCESS |



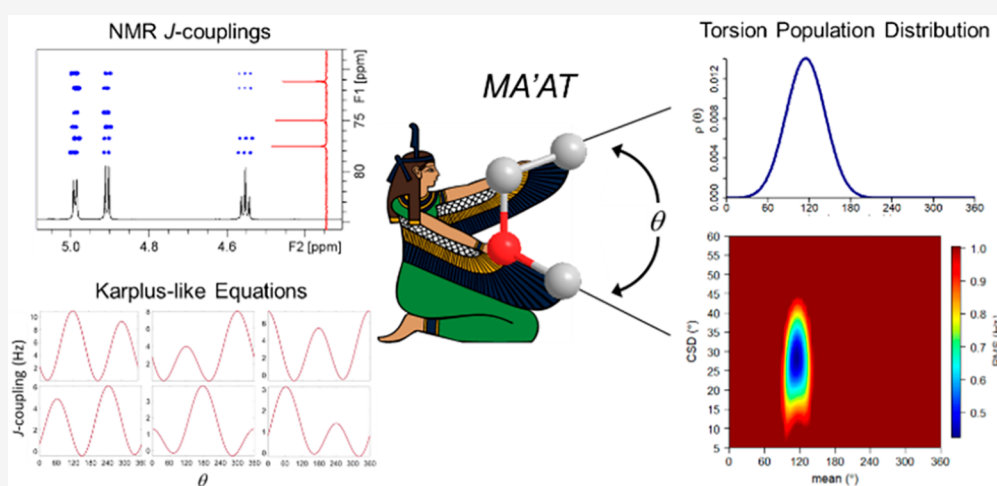
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: A hybrid experimental–computational method to determine conformational equilibria of molecules in solution has been developed based on the use of redundant nuclear magnetic resonance (NMR) spin–spin coupling constants (spin-couplings; J -couplings), density functional theory (DFT) calculations, and circular statistics. The mathematics that underpins the method, known as MA'AT analysis, is presented, and key components of a computer program that applies this algorithm are discussed. The method was tested using single-state and multi-state models to identify the factors required to obtain reliable results, to establish the limitations of the method, and to highlight techniques to evaluate the uniqueness of solution.

INTRODUCTION

Molecular structures are determined routinely in crystalline solids^{1,2} and more recently in the gas phase.^{3,4} However, determinations of molecular structures in solution remain challenging. While computational methods have improved significantly, experiment-based structure determinations of molecules in solution have not advanced appreciably over the past few decades, especially for more dynamic structures. Many molecules are structurally flexible in solution, which complicates the interpretation of experimental parameters. The structure of a molecule in solution may differ from that observed in a crystalline state or when complexed to a receptor. To identify how biomolecular structure dictates function *in vivo*, quantitative relationships between covalent structure and solution conformation and dynamics are needed. This need can only be met by an experimental technique that provides information on all biomolecular conformations in solution over time.

Nuclear magnetic resonance (NMR) spectroscopy is an experimental method that provides information to construct detailed structural models of molecules in solution. The work described here involves the measurement and analysis of multiple, redundant NMR spin–spin coupling constants, which are highly abundant in many molecules and average linearly in the presence of conformational exchange (i.e., the observed J -value is determined by the weighted average of the individual J -values for each conformer in solution in the rapid-exchange limit). Ensembles of J -couplings that display distinctly different dependencies on the same conformational element (i.e., a molecular torsion angle) are extremely useful in structural

Received: September 23, 2021

Published: June 22, 2022



determinations of conformationally flexible molecules such as saccharides and intrinsically disordered proteins.⁵

Spin-coupling constants can be used to characterize conformational equilibria in solution when quantitative relationships between their magnitudes and signs with one or more molecular parameters are known. Karplus reported the first such relationship in which three-bond (vicinal) $^3J_{\text{HCCH}}$ spin-coupling constants were found through computational studies to obey a $\cos^2(\theta)$ function, where θ corresponds to the H–C–C–H torsion angle.⁶ Karplus-like relationships have subsequently been developed for geminal and vicinal homonuclear and heteronuclear J -couplings involving heavy atoms as coupled spins.^{7,8} Prior work has shown that many of these relationships are not generalizable but depend on the identities of internal and external substituents along the coupling pathway and on their relative orientations.⁹ Thus, J -couplings have limited utility unless quantitative relationships between their magnitudes/signs and molecular structure are either available or can be readily obtained. Density functional theory (DFT) has been used to obtain relationships between individual J -couplings and one or more molecular properties.^{5,7,10,11}

Current DFT calculations, molecular dynamics (MD) simulations, and other computational methods^{12,13} are able to treat molecules that exist as conformational populations in solution. Conversely, determining conformational populations (equilibria) in solution from experimental data is not straightforward. This situation has led to a heavy reliance on MD simulations and other theoretical methods to interpret experimental data^{14,15} despite the fact that experimental validations of MD simulation results are largely lacking. MD force fields are currently evaluated using available structural data to model conformational behavior in solution, and these data are commonly provided by X-ray crystallography, especially for saccharides. Consequently, the models obtained by MD simulations can be biased toward structures observed in the crystalline state and may not faithfully recapitulate those that are found in solution. This report describes a new experimental method, MA'AT analysis, to determine conformational populations in solution that promises to circumvent this problem and enable more rigorous validation of the results obtained from MD and other computational methods.

MATHEMATICS UNDERLYING MA'AT ANALYSIS

The dependencies of NMR J -couplings on molecular torsion angles are well understood,^{6,16} making it possible to obtain functions of $J(\theta)$ that describe the dependencies of molecular torsion angles on J -couplings. Experimental J -couplings are time-averaged values over the entire population in solution. Therefore, an experimental J -coupling (J_{exp}) can be described mathematically by eq 1, where $J(\theta)$ equals the J -coupling at torsion angle θ , and $p(\theta)$ is the population density.

$$J_{\text{exp}} = \int_0^{2\pi} J(\theta)p(\theta)d\theta \quad (1)$$

Parametrized equations can be obtained experimentally or from DFT^{5,7,10,11} or other calculations for an ensemble of J -couplings that are sensitive to a specific torsion angle, θ . The challenge becomes how to use this information to determine $p(\theta)$. Several computational methods have been developed to solve this problem. The simplest method assumes that $p(\theta)$ is nonzero for a few discrete values of θ and zero for all remaining values. While the simplicity of this method is appealing, there

are several drawbacks. Assuming that θ only adopts a few discrete values introduces bias in the treatment. Furthermore, this model of θ has no physical basis since it is highly likely that the molecule experiences some degree of libration about the optimal value or values of θ and/or that θ values in addition to the optimal value and values contribute to the experimental J -couplings.

Continuous models have been proposed to overcome these problems. These models structure $p(\theta)$ as a sum of basis functions to better approximate the true distribution. An example of this approach is the CUPID method (ContinUous ProbabIlity Distribution of rotamers) in which a Fourier series is used to represent $p(\theta)$.¹⁷ Fourier series are well understood functions, and with techniques like fast Fourier transformation (FFT), computations are possible even on limited hardware. However, this approach suffers from several drawbacks. Trigonometric functions are unattractive when describing a probability distribution because they can produce negative populations. More importantly, because there are often only a few J -couplings in an ensemble, the Fourier series must be truncated after a few terms. This truncation produces a function that derives much of its shape from the trigonometric basis instead of from the data.

A logical choice to represent $p(\theta)$ would be a sum of Gaussian distributions. Gaussian-like distributions are commonly observed when investigating torsional behavior from crystal databases or MD simulations and provide a straightforward population distribution for torsion angle modeling.¹⁸ However, Gaussian distributions are not periodic; that is, they are defined over the entire real line, whereas an angular probability distribution should be defined over a finite interval (e.g., from 0 to 2π). Wrapped normal distributions are analogous to Gaussian distributions wrapped around a circle to eliminate the periodicity problem but are computationally demanding to integrate when used in eq 1. Furthermore, evaluating the integral in eq 1 numerically for many sample points as part of a nonlinear optimization is computationally demanding, especially as the dimension of the sample space increases (increasing the modality introduces three additional dimensions).

The MA'AT (named after the Egyptian goddess, MA'AT) method to determine rotamer population distributions evaluates eq 1 analytically, thereby eliminating the need for computationally costly numeric integration which allows a large number of population distributions to be sampled efficiently. This is achieved by representing the $J(\theta)$ term as a modified Karplus relationship¹⁹ (eq 2) in the form of a trigonometric polynomial, where k , c , and s are constants for a specific J -coupling, and N is the degree of the polynomial (usually 2 or 3).

$$J(\theta) = k + \sum_{n=1}^N c_n \cos(n\theta) + \sum_{n=1}^N s_n \sin(n\theta) \quad (2)$$

Initially $p(\theta)$ is modeled as a sum of wrapped normal distributions (eq 3), where P is the number of peaks, w the weighting parameter per peak, μ the mean position, and σ the circular standard deviation (CSD).

$$p(\theta) = \int_0^{2\pi} \sum_{p=0}^P \sum_{j=-\infty}^{\infty} \frac{w_p}{\sigma_p \sqrt{2\pi}} e^{-\frac{(\theta+2\pi j-\mu_p)^2}{2\sigma_p^2}} d\theta \quad (3)$$

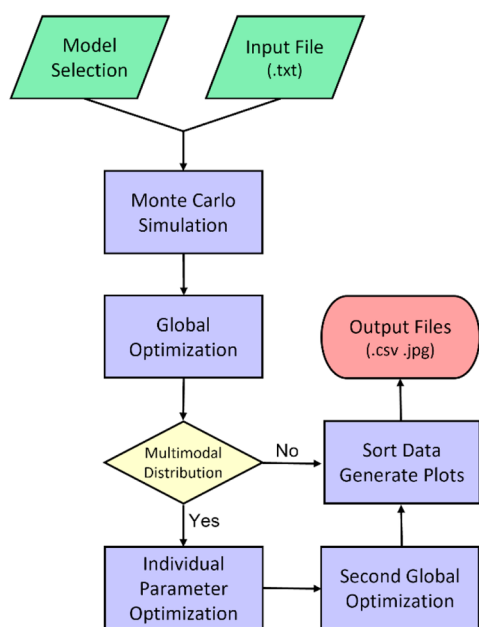
Since $p(\theta)$ is a probability distribution, the sum of the weighting parameters always equals 1. Substituting the functions in eqs 2 and 3 into eq 1 and performing analytical integration produces the MA'AT eq (eq 4) (see Supporting Information for derivation). An ensemble of J -couplings gives rise to a series of MA'AT equations, making it possible to solve for optimal values of w_p , μ_p , and σ_p .

$$J = \sum_{p=0}^P w_p \sum_{n=1}^N e^{-\frac{(\sigma_p n)^2}{2}} (c_n \cos(\mu_p n) + s_n \sin(\mu_p n)) + k \quad (4)$$

IMPLEMENTATION OF MA'AT ANALYSIS

MA'AT analysis has been encoded into a Shiny application in R. The application generates population distributions from user-supplied experimental data and parametrized Karplus-like equations. The workflow is illustrated in Scheme 1. The ability

Scheme 1. MA'AT Application Flowchart



to generate unique distribution models depends on multiple factors that must be evaluated before running an analysis. The primary consideration is the number of experimental J -couplings to be used in the analysis. Since the analysis utilizes only user-supplied data, increasing the information provided to the application increases the precision of the results. Importantly, all J -couplings included in the ensemble must have a significant dependence on a single torsion angle.

The application requires the user to upload the experimental data and parametrized J -coupling equation in a text file and then select a population model. MA'AT supports unimodal, bimodal, and trimodal wrapped normal or von Mises^{20,21} distribution models. Advanced options are also available that allow the user to restrict some or all of the distribution model parameters to a range of values. A Monte Carlo simulation is performed initially to generate population models that are used to calculate J -couplings from the user-supplied equations. A root mean squared deviation (RMSD) is then calculated for each model. The models with the lowest RMSDs are then optimized using the Nelder–Mead derivative-free algorithm.^{22,23} No further optimizations are performed when a

unimodal distribution is selected. For multimodal modeling, individual parameters (mean, CSD, and peak populations) are separately optimized followed by a second global optimization. The peak population parameters are restricted in the optimizations to ensure negative values are not produced. The data are then sorted from lowest-to-highest RMSD and tables and graphs (see below) are generated. If a von Mises distribution is being modeled, the final global optimization step is performed by evaluating the integral produced from eq 1 instead of using the MA'AT equation, which increases the run time significantly. Standard errors of each parameter are estimated by taking the square roots of the diagonal elements from an inverted Hessian matrix.²⁴

The application produces several data tables and graphs containing information on the predicted distribution model(s), the uniqueness of solution, and a comparison of experimental and calculated J -values. The outputs are separated into individual tabs, with the first tab showing the parameters for the top fits (Figure S10A, Supporting Information). The second tab contains a data table of the experimental and calculated J -couplings for each model (Figure S10B, Supporting Information). The third tab displays a plot of the top distributions (Figure S11, Supporting Information). The fourth and final tab contains two plots that report on the uniqueness of the solution (Figure S12, Supporting Information). The first plot is a histogram of the RMSDs obtained from the Monte Carlo simulation. If only a few models produce small RMSDs, then the solution is likely to be unique. However, if numerous models produce low RMSDs, then more experimental constraints are needed to generate a unique solution. The second plot differs for unimodal and multimodal analyses. For unimodal analyses, the second plot is a surface plot of the two model parameters colored by the associated RMSD (Figure S5, Supporting Information). The uniqueness of solution is determined by visual inspection of the parameter space plots and through inspection of the standard errors of each model parameter. For multimodal analyses, visualizing the uniqueness of solution is more difficult due to the increased number of variables. This problem is overcome by generating numerous fits to determine whether the overall fits converge. Histograms of mean positions, CSDs, and peak populations are used to evaluate the uniqueness of the individual parameters of the models (Figure S6 and S7, Supporting Information).

The Shiny application is available at <https://rmeredit.shinyapps.io/maat24/>. For more details on the user interface, refer to the MA'AT user manual, which can be downloaded from the application.

RESULTS

Method Testing. The ability of MA'AT analysis to reproduce population models from an ensemble of J -couplings was tested by back-calculating J -couplings from a known distribution using an ensemble of Karplus-like equations. The back-calculated J -couplings were then used as input to the MA'AT application to determine whether the original distributions could be reproduced. Initial distributions to test the application were generated in R. Random distributions were generated for each modality being tested. Distribution variables, μ , CSDs, and peak percentages (multimodal only) were produced with a random number generator. CSD values were limited to a range of 10° to 40°. For multimodal modeling, mean positions were separated by a minimum of 20°, and w_p had a minimum of 0.15 to ensure modality of the

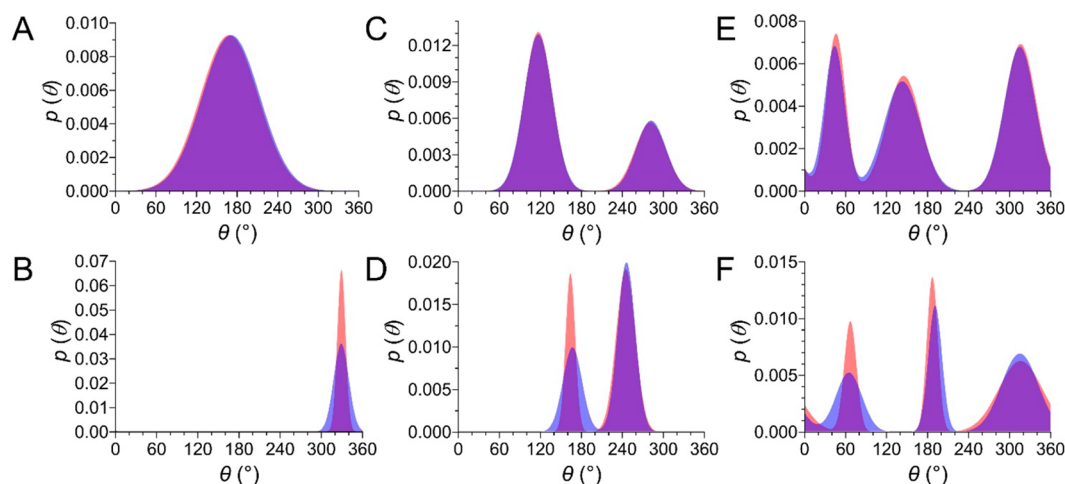


Figure 1. Comparison of six input and output distributions. Blue areas are the randomly generated input distributions. Red areas are the MA'AT output distributions. Purple areas are the regions where the distributions agree. Unimodal: 98.5% (A) and 74.5% (B). Bimodal: 98.8% (C) and 86.5% (D). Trimodal: 96.8% (E) and 84.7% (F).

distributions was retained. Since the associated J -couplings showed no significant differences when wrapped normal and von Mises distributions were generated, only wrapped normal distributions were considered for the input distributions.

Simulated J -couplings were computed in R by multiplying an ensemble of equations by an input distribution. J -couplings were then calculated by taking the Riemann sum of the resulting line ($n = 36,000$). A second set of J -couplings was generated to mimic uncertainties in the experimental J -couplings. A random number generator added noise of ± 0.8 Hz to the simulated values, and the latter were rounded to a single decimal.

One hundred random distributions were generated for each modality model (uni, bi, or tri). For the unimodal models, two ensembles composed of three (ensemble 1) or two (ensemble 2) parametrized equations each were tested (Table S1, Supporting Information). For multimodal models, a total of 10 ensembles were tested using 6–28 equations (Tables S2–S5, Supporting Information). One hundred distributions with the lowest RMSDs were output by the application. Output distributions were compared to input distributions to determine whether the application accurately reproduced the population models. Accuracy was evaluated using a percent similarity score calculated from the difference between the two different distributions (Figure 1). Percent similarity is reported as the average for an ensemble. However, this test is not completely reliable because relatively small changes in the model parameters can lead to large decreases in percent similarity. Consequently, two additional metrics were applied to assess similarity: DTS test for common distributions and a common means test (see Supporting Information for more details on DTS test).^{25,26} Common means testing was performed by calculating an RMSD between the mean values of the input and output distributions. If the RMSD value was below 18° (5% of 360°), then the models share common means. While this metric only considers the application's ability to reproduce mean positions, it is useful because of the large change in percent similarity resulting from small changes in CSDs. For example, Figure 1B shows a percent similarity score of only 74.5%, yet the models differ in their mean positions by only 0.3° and in their CSDs by only 5° .

Unimodal Models. Unimodal distributions were reproduced with an average percent similarity of greater than 97% for ensemble 1 and greater than 80% for ensemble 2. While the former consistently produced a single distribution, the latter produced two to three distributions that matched the data equally in many of the tests. When the output distributions were filtered to the distribution that was most similar to the input, the average percent similarity for both ensembles rose to $\sim 99\%$. When random noise was added, the average percent similarity was reduced to $\sim 70\%$ and $\sim 60\%$ for the two ensembles. After filtering the output distributions, the percent similarity rose to $\sim 80\%$ for both ensembles. Tables S1 and S2 contain the unimodal testing statistics, and violin plots of the percent similarities are displayed in Figure S2 in the Supporting Information.

MA'AT reproduced unimodal distributions with high accuracy even from as little as two observables, although the data needed to be filtered since the reduced number of observables can result in multiple distributions with comparable RMSDs. When random noise was added, the accuracy was reduced significantly even after filtering the data. The amount of random noise correlated strongly with the method's ability to reproduce the data.

Multimodal Models. Multimodal analysis typically generates numerous distributions instead of a single model. The models that were produced matched the input data equally and typically contained only small differences in mean positions and CSDs. To allow useful comparisons, the models with the lowest RMSDs were averaged if their parameters did not differ significantly. Occasionally, models with significant differences were produced, especially when the number of simulated experimental constraints was low. In these cases, inspection of the distribution plots is recommended.

Bimodal models were reproduced with $\sim 95\%$ similarity, and trimodal models were reproduced with $\sim 90\%$ similarity (Table 1). When the output distributions were filtered to the model that was most similar to the input, only the trimodal ensembles with six J -couplings showed a significant improvement in their accuracy. When random noise was included in these tests, these similarities dropped to $\sim 70\%$ and $\sim 55\%$, respectively. After filtering the output distributions, the percent similarity rose to $\sim 75\%$ for bimodal models and $\sim 65\%$ for trimodal

Table 1. Multimodal Testing Statistics

No. of J s in fit	Bimodal			Trimodal		
	% Similarity	DTS test ^a	Common means ^b	% Similarity	DTS test ^a	Common means ^b
29	96.0	5	1	91.4	2	6
20	95.8	7	2	91.1	0	5
15	95.2	6	2	90.8	2	5
10	95.2	5	2	90.4	0	6
7	94.7	8	0	89.9	2	7
6	93.9	4	3	80.0	10	29
20	95.7	7	1	90.9	2	7
15	95.2	9	2	91.0	0	5
10	95.0	9	1	90.7	0	3
6	91.9	10	6	78.5	7	39

^aNumber of fits in which the DTS test rejected the null hypothesis.

^bNumber of fits in which the mean positions RMSD differed significantly from the input.

models. The reduced similarities are caused by larger variations in the CSDs in the output models. Small changes in CSDs when mean positions are conserved result in significant reductions of percent similarity. Tables S3–S5 contain the multimodal testing statistics with the inclusion of random noise and the best output model statistics. Violin plots of the percent similarities are displayed in Figures S3 and S4 in the Supporting Information.

Limitations. The MA'AT method is limited primarily by the accuracy of the input J -couplings. As random noise increases, accuracy decreases. Therefore, it is imperative that only high-quality experimental J -coupling data are used. Applying a random noise of ± 0.8 Hz reduced the method's ability to reproduce the input distributions for all modalities, as expected. When the data were carefully analyzed as a function of random noise, the method gave high reproducibility when the average noise was ± 0.3 – 0.4 Hz or less. Random noise of ± 0.4 – 0.6 Hz lowered the reproducibility to $\sim 50\%$. Since a random number generator was used to generate noise, it would be expected that the number of fails would be $\sim 25\%$ which is similar to the values for testing except in the case of trimodal modeling. Trimodal modeling was able to reproduce the input distributions $\sim 70\%$ of the time when more than 10 J -couplings were included; however, these analyses required close inspection of the data since numerous solutions were produced.

The number of J -couplings used in the analysis also affects the method's ability to accurately reproduce the input distributions to varying extents. The number of J -couplings depends on the sensitivity of the J -coupling to the torsion angle being examined. A reduced number of J -couplings does not reduce accuracy as long as they all have a strong dependence on the torsion angle. However, the uniqueness of solution for these fits is likely to be reduced, resulting in numerous distributions that reproduce the data equally. The ensembles tested with fewer J -couplings are comprised of those values that are most sensitive to the torsion angle being examined. If different ensembles of 10 J -couplings were used, then reproducibility is reduced (data not shown). Adding additional J -couplings with reduced sensitivity is not expected to improve the accuracy but may reduce the number of solutions produced in any given analysis.

Multimodal modeling suffers from more limitations than unimodal modeling due to the increased complexity of the model. Determining the uniqueness of solution for unimodal analyses requires inspections of parameter landscapes, but this is not possible for multimodal analyses. Instead, the uniqueness of solution can be determined by inspecting the models that are the closest match to the data, a histogram of the RMSDs, and histograms of the different model parameters (Figures S6 and S7, Supporting Information). Given the nature of the Karplus-like equations and the population models, it is likely that multiple models that have conserved mean values but vary in CSDs and/or peak populations will fit the data equally.

CONCLUSIONS

MA'AT analysis offers significant advantages to model molecular torsion angle populations in solution compared to previous experimental methods. First, MA'AT analysis provides continuous distributions and thus more physically realistic conformational models.^{27–29} Continuous distributions eliminate bias due to partial sampling of the conformational space inherent to discrete distributions. Furthermore, circular distributions avoid models with negative populations and/or other unphysical results.¹⁷ No assumptions are made about the CSDs, thereby providing insights into torsional dynamics. MA'AT analysis has been used to study a wide range of conformational properties of saccharides, including *O*-glycosidic linkage conformations, side-chain conformations, and the pseudorotation of five-membered rings.^{9,30–34} More recent applications include the modeling of backbone conformation in oligopeptides. In principle, the method should be applicable to any molecule provided that the molecule contains sufficient numbers of redundant J -couplings to interrogate the torsion angle(s) of interest.

The MA'AT web application described in this report is user friendly with a simple GUI that only requires a set of J -coupling equations and experimental J -couplings as inputs. MA'AT analysis generates continuous multimodal distributions with speed and precision from minimal user-supplied information. Unimodal modeling produces accurate and precise population models from only a few experimental measurements. Multimodal modeling has more limitations due to the complexity of the models, but MA'AT can still generate models with accurate mean values and CSDs in many cases. The method is limited by the accuracy of the experimental J -couplings, the number of available J -couplings, the sensitivity of the J -couplings to the molecular torsion angle of interest, and the ability to generate quantitative parametrized equations for the J -couplings.

DATA SOFTWARE AND AVAILABILITY

Users can access the MA'AT application free of charge at <https://rmeredit.shinyapps.io/maat24/>. R scripts can be downloaded from the application for offline use (recommended for multimodal analysis). Testing data are contained in a zip file available as Supporting Information. The contents of the zip file are divided into three main folders (one for each modality tested). Each of these three folders contains either 2 (for 1_state) or 10 (for 2_state and 3_state) subfolders with 200 text files needed for testing.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c01166>.

Additional information on application testing, scoring metrics, testing results, methods to determine the uniqueness of solution, user interface, and the MA'AT equation proof (PDF)

Input files used for application testing (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Anthony S. Serianni – Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556-5670, United States; orcid.org/0000-0001-6114-1446; Email: aseriann@nd.edu

Authors

Reagan J. Meredith – Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556-5670, United States; orcid.org/0000-0001-7330-3190

Luke Sernau – Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556-5670, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.1c01166>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

R.J.M., L.S., and A.S. acknowledge support from the National Science Foundation (CHE 1707660 and CHE 2002625) and from Omicron Biochemicals, Inc., South Bend, Indiana.

■ REFERENCES

- (1) Harris, K. D. M.; Tremayne, M.; Kariuki, B. M. Contemporary Advances in the Use of Powder X-Ray Diffraction for Structure Determination. *Angew. Chem.* **2001**, *40*, 1626–1651.
- (2) Bunaciu, A. A.; Udristioiu, E. G.; Aboul-Enein, H. Y. X-Ray Diffraction: Instrumentation and Applications. *Crit. Rev. Anal. Chem.* **2015**, *45*, 289–299.
- (3) Harmony, M. D.; Laurie, V. W.; Kuczkowski, R. L.; Schwendeman, R. H.; Ramsay, D. A.; Lovas, F. J.; Lafferty, W. J.; Maki, A. G. Molecular Structures of Gas-Phase Polyatomic Molecules Determined by Spectroscopic Methods. *J. Phys. Chem. Ref. Data* **1979**, *8*, 619–721.
- (4) Calabrese, C.; Écija, P.; Compañón, I.; Vallejo-López, M.; Cimas, Á.; Parra, M.; Basterretxea, F. J.; Santos, J. I.; Jiménez-Barbero, J.; Lesarri, A.; Corzana, F.; Cocinero, E. J. Conformational Behavior of D-Lyxose in Gas and Solution Phases by Rotational and NMR Spectroscopies. *J. Phys. Chem. Lett.* **2019**, *10*, 3339–3345.
- (5) Klepach, T.; Zhao, H.; Hu, X.; Zhang, W.; Stenutz, R.; Hadad, M. J.; Carmichael, I.; Serianni, A. S. Informing Saccharide Structural NMR Studies with Density Functional Theory Calculations. In *Glycoinformatics: Methods in Molecular Biology*; Lütteke, T., Frank, M., Eds.; Springer: New York, 2015; pp 289–331.
- (6) Karplus, M. Contact Electron-Spin Coupling Nuclear Magnetic Moments. *J. Chem. Phys.* **1959**, *30*, 11–15.
- (7) Thibaudeau, C.; Stenutz, R.; Hertz, B.; Klepach, T.; Zhao, S.; Wu, Q.; Carmichael, I.; Serianni, A. S. Correlated C–C and C–O Bond Conformations in Saccharide Hydroxymethyl Groups: Parametrization and Application of Redundant ^1H – ^1H , ^{13}C – ^1H , and ^{13}C – ^{13}C NMR J-Couplings. *J. Am. Chem. Soc.* **2004**, *126*, 15668–15685.
- (8) Zhao, H.; Pan, Q.; Zhang, W.; Carmichael, I.; Serianni, A. S. DFT and NMR Studies of $^2J_{\text{COH}}$, $^3J_{\text{HCOH}}$, and $^3J_{\text{CCOH}}$ Spin-couplings in Saccharides: C–O Torsional Bias and H-bonding in Aqueous Solution. *J. Org. Chem.* **2007**, *72*, 7071–7082.
- (9) Zhang, W.; Turney, T.; Meredith, R.; Pan, Q.; Sernau, L.; Wang, X.; Hu, X.; Woods, R. J.; Carmichael, I.; Serianni, A. S. Conformational Populations of β -(1→4) O-Glycosidic Linkages Using Redundant NMR J-Couplings and Circular Statistics. *J. Phys. Chem. B* **2017**, *121*, 3042–3058.
- (10) Parr, R. G.; Yang, W. Density-Functional Theory of the Electronic Structure of Molecules. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701–728.
- (11) Krivdin, L. B. Computational NMR of Carbohydrates: Theoretical Background, Applications, and Perspectives. *Molecules* **2021**, *26*, 2450.
- (12) Helgaker, T.; Jaszunski, M.; Ruud, K. Ab Initio Methods for the Calculation of NMR Shielding and Indirect Spin-Spin Coupling Constants. *Chem. Rev.* **1999**, *99*, 293–352.
- (13) Fang, J.; Hu, L.; Dong, J.; Li, H.; Wang, H.; Zhao, H.; Zhang, Y.; Liu, M. Predicting Scalar Coupling Constants by Graph Angle-Attention Neural Network. *Sci. Rep.* **2021**, *11*, 18686.
- (14) Woods, R. J. Computational Carbohydrate Chemistry: What Theoretical Methods Can Tell Us. *Glycoconjugate J.* **1998**, *15*, 209–216.
- (15) Autschbach, J.; Le Guennic, B. Analyzing and Interpreting NMR Spin-Spin Coupling Constants Using Molecular Orbital Calculations. *J. Chem. Educ.* **2007**, *84*, 156–171.
- (16) Karplus, M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871.
- (17) Dzakula, Z.; Westler, W. M.; Edison, A. S.; Markley, J. L. The “CUPID” Method of Calculating the Continuous Probability Distribution of Rotamers from NMR Data. *J. Am. Chem. Soc.* **1992**, *114*, 6195–6199.
- (18) Irbäck, A.; Mohanty, S. PROFASI: A Monte Carlo Simulation Package for Protein Folding and Aggregation. *J. Comput. Chem.* **2006**, *27*, 1548–1555.
- (19) Pachler, K. G. R. Extended Hückel Theory MO Calculations of Proton-proton Coupling Constants—II: The Effect of Substituents on Vicinal Couplings in Monosubstituted Ethanes. *Tetrahedron* **1971**, *27*, 187–199.
- (20) Maksimov, V. M. Necessary and Sufficient Statistics for the Family of Shifts of Probability Distributions on Continuous BiCompact Groups. *Theory Probab. Appl.* **1967**, *12*, 267–314.
- (21) Gatto, R.; Jammalamadaka, S. R. The Generalized von Mises Distribution. *Stat. Methodol.* **2007**, *4*, 341–353.
- (22) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Computer J.* **1965**, *7*, 308.
- (23) Varadhan, R.; Borchers, H. W. *dfoptim: Derivative-Free Optimization*, ABB Corporate Research R Package, 2016, Vol. 7 <https://CRAN.R-project.org/package=dfoptim> (accessed June 2022).
- (24) Yuen, K. *Bayesian Methods for Structural Dynamics and Civil Engineering*; Wiley, 2010.
- (25) Dowd, C. twosamples: Fast Permutation Based Two Sample Tests, R package version 1.1.1, 2020. <https://CRAN.R-project.org/package=twosamples> (accessed June 2022).
- (26) Dowd, C. A New ECDF Two-Sample Test Statistic. *arXiv Preprint*, arXiv:2007.01360, 2020. DOI: [10.48550/arXiv.2007.01360](https://doi.org/10.48550/arXiv.2007.01360)
- (27) De Leeuw, F. A. A. M.; Altona, C.; Kessler, H.; Bermel, W.; Friedrich, A.; Krack, G.; Hull, W. E. Peptide Conformations. 20. Conformational Analysis of Proline Rings from Proton Spin-spin Coupling Constants and Force-field Calculations: Application to Three Cyclic Tripeptides. *J. Am. Chem. Soc.* **1983**, *105*, 2237–2246.
- (28) Haasnoot, C. A. G.; De Leeuw, F. A. A. M.; De Leeuw, H. P. M.; Altona, C. Relationship Between Proton-proton NMR Coupling Constants and Substituent Electronegativities. III. Conformational Analysis of Proline Rings in Solution Using a Generalized Karplus Equation. *Biopolymers* **1981**, *20*, 1211–1245.

(29) De Leeuw, F. A. A. M.; Altona, C. Computer-Assisted Pseudorotation Analysis of Five-Membered Rings by Means of Proton Spin-Spin Coupling Constants: Program PSEUROT. *J. Comput. Chem.* **1983**, *4*, 428–437.

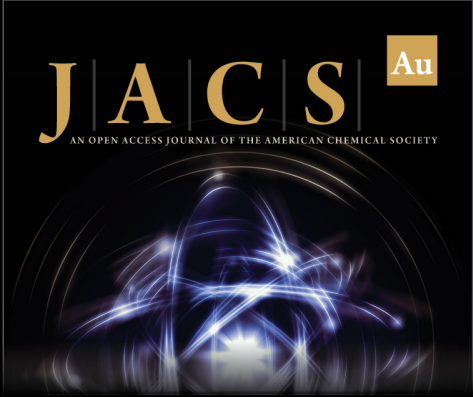
(30) Zhang, W.; Meredith, R.; Pan, Q.; Wang, X.; Woods, R. J.; Carmichael, I.; Serianni, A. S. Use of Circular Statistics To Model α Man-(1 $\rightarrow$ 2)- α Man and α Man-(1 $\rightarrow$ 3)- α / β Man O-Glycosidic Linkage Conformation in ^{13}C -Labeled Disaccharides and High-Mannose Oligosaccharides. *Biochemistry* **2019**, *58*, 546–560.

(31) Zhang, W.; Meredith, R.; Yoon, M.-Y.; Wang, X.; Woods, R. J.; Carmichael, I.; Serianni, A. S. Synthesis and O-Glycosidic Linkage Conformational Analysis of ^{13}C -Labeled Oligosaccharide Fragments of an Antifreeze Glycolipid. *J. Org. Chem.* **2019**, *84*, 1706–1724.

(32) Meredith, R. J.; Woods, R. J.; Carmichael, I.; Serianni, A. S. Reconciling MA'AT and Molecular Dynamics Models of Linkage Conformation in Oligosaccharides. *Phys. Chem. Chem. Phys.* **2020**, *22*, 14454–14457.

(33) Turney, T.; Pan, Q.; Sernau, L.; Carmichael, I.; Zhang, W.; Wang, X.; Woods, R. J.; Serianni, A. S. O-Acetyl Side-Chains in Monosaccharides: Redundant NMR Spin-Couplings and Statistical Models for Acetate Ester Conformational Analysis. *J. Phys. Chem. B* **2017**, *121*, 66–77.

(34) Meredith, R. J.; McGurn, M.; Euell, C.; Rutkowski, P.; Cook, E.; Carmichael, I.; Serianni, A. S. MA'AT Analysis of Aldofuranosyl Rings: Unbiased Modeling of Conformational Equilibria and Dynamics in Solution. *Biochemistry* **2022**, *61*, 239–251.



JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

 Editor-in-Chief
Prof. Christopher W. Jones
Georgia Institute of Technology, USA

Open for Submissions 

pubs.acs.org/jacsau

 **ACS Publications**
Most Trusted. Most Cited. Most Read.