Title

The impact of conformational sampling on first-principles calculations of vicinal COCH J-couplings in carbohydrates

Journal
Glycobiology, 33(1)

ISSN
0959-6658

Authors
Reeves, Hannah L
Wang, Lee-Ping

Publication Date
2023-01-08

DOI
10.1093/glycob/cwac073

Peer reviewed

Glycobiology, 2023, **33**, 1, 38–46
https://doi.org/10.1093/glycob/cwac073
Advance access publication date 2 November 2022
Original Article

The impact of conformational sampling on first-principles calculations of vicinal COCH J-couplings in carbohydrates

Hannah L Reeves, Lee-Ping Wang*

Department of Chemistry, University of California at Davis, 1 Shields Ave, Davis, CA 95616, USA
*Corresponding author: Department of Chemistry, University of California at Davis, 1 Shields Ave, Davis, CA 95616, USA.
Email: leeping@ucdavis.edu

Dihedral angles in organic molecules and biomolecules are vital structural parameters that can be indirectly probed by nuclear magnetic resonance (NMR) measurements of vicinal J-couplings. The empirical relations that map the measured couplings to dihedral angles are typically determined by fitting using static structural models, but this neglects the effects of thermal fluctuations at the finite temperature conditions under which NMR measurements are often taken. In this study, we calculate ensemble-averaged J-couplings for several structurally rigid carbohydrate derivatives using first-principles molecular dynamics simulations to sample the thermally accessible conformations around the minimum energy structure. Our results show that including thermal fluctuation effects significantly shifts the predicted couplings relative to single-point calculations at the energy minima, leading to improved agreement with experiments. This provides evidence that accounting for conformational sampling in first-principles calculations can improve the accuracy of NMR-based structure determination for structurally complex carbohydrates.

Key words: vicinal coupling; nuclear magnetic resonance; dynamics; density functional theory; carbohydrate.

Introduction

Nuclear magnetic resonance (NMR) is commonly used to study the structural features of carbohydrates and other biomolecules, for example by measuring vicinal J-couplings, which are strongly correlated with dihedral angles (Karplus 1963; Tvaroška et al. 1989; San Fabián et al. 2019). The J-coupling is the result of an interaction between nuclear magnetic dipole moments (Karplus 1959) and can be determined experimentally (Tvaroska and Taravel 1995) by measuring line spacings from triple resonance spectra such as HNCA (Wang and Bax 1995) or coupled Heteronuclear Single Quantum Coherence (HSQC) experiments (Zimmer et al. 1996; Li et al. 2022). Because of its sensitivity to molecular conformation (Cano et al. 1986), the J-coupling is often applied to determine the dihedral angle of glycosidic linkages (Tvaroska and Taravel 1995).

The empirical Karplus relation provides an approximate mapping between dihedral angles and J-couplings, and has the typical functional form $J = A\cos^2\phi + Bcos\phi + C$ (Karplus 1963). Additional terms and empirical parameters may be added to provide more flexibility when fitting the equation to experimental data (Hackbusch et al. 2017). Because the coefficients depend on the chemical environment of the dihedral angle being measured (Karplus 1963), they are usually fitted by first measuring several dihedral angles in chemically similar environments via X-ray crystallography experiments, then fitting the curves to J-couplings measured from solution-phase NMR experiments (Case et al. 2000).

Karplus relations are only reliable when the dihedral angle being studied is chemically similar to the dihedral angles used to fit the parameters; for example, a Karplus equation fitted for proteins may not be applicable to carbohydrates as the atoms and functional groups surrounding the bond of interest differ (Tafazzoli and Ghiasi 2007). However, an error of at least 1 Hz can be expected even when the appropriate Karplus equation is used (Cano et al. 1987; Richards et al. 2013). In several carbohydrate studies, the Karplus estimations were even less accurate for conformations where the J-coupling was at a local maximum (Engelsen et al. 1995; Cloran et al. 1999; Richards et al. 2013). A source of the error could be the differences between physical environments in NMR and X-ray crystallography experiments. In typical NMR experiments the molecules are in solution (Toukach and Ananikov 2013), whereas in X-ray crystallography experiments the molecules are in a periodic crystal (Coxon 2009). These differences in conditions may lead to systematic inaccuracies when the Karplus equation is used to infer dihedral angles from couplings or vice versa (Hoch et al. 1985).

J-couplings can also be estimated with quantum chemistry (QM) calculations (Taha et al. 2010). An advantage of QM calculations is that the J-coupling is calculated from a model structure where the dihedral angle is specified exactly; the angle can then be scanned to explore conformations that may not be available from crystal structures (Cloran et al. 1999). A significant limitation of QM calculations is that they are approximate, with different methods having different accuracy and cost (San-Fabián et al. 1993). Methods at the ab initio level such as the coupled-cluster iterative triples (CC3) model can calculate J-couplings and with a high degree of accuracy, but are currently too computationally demanding to be performed on larger molecules, including many carbohydrates (Krivdin 2021). Density functional theory (DFT) methods provide a good compromise between accuracy and computational cost for medium-sized systems such as oligosaccharides. These methods account for electron correlation effects implicitly through the electron density instead of explicitly through the wavefunction and are thus much less computationally demanding, but their results have a high dependency on the choice of functional and basis set (Krivdin 2021). Furthermore, QM calculations of J-couplings

Received: July 14, 2022. **Revised:** October 14, 2022. **Accepted:** October 25, 2022

are typically performed on single energy-minimized structures, whereas J-couplings measured from NMR are derived from molecules in solution with multiple conformational states (Engelsen et al. 1995). The variance in conformations away from the energy-minimized structure has the potential to introduce systematic errors into the predicted coupling values.

Previous DFT studies demonstrate the impact of ensemble-averaging on the J-coupling in NMR. In a 2007 study by Tafazzoli et al., DFT calculations using the B3LYP functional were performed to calculate vicinal COCH J-couplings in aldohexopyranoside derivatives using energy-minimized structures (Tafazzoli and Ghiasi 2007). When the calculated J-couplings were compared with empirical J-couplings, the DFT-calculated J-couplings were accurate within ~1 Hz. A 2010 study by Taha et al, used DFT calculations at the B3LYP level to calculate vicinal J-couplings in arabofuranosides (Taha et al. 2010); this study attempted to replicate the conformational sampling in solution by calculating J-couplings for several known conformations of the molecules under study. The prevalence of each conformation was evaluated through a 200-ns molecular dynamics (MD) simulation to provide a weighted-average J-coupling for each dihedral angle (Taha et al. 2010). The weighted-average J-couplings were more accurate than the empirical Karplus relations in every angle observed, and in 8 of the 17 angles examined, the DFT-calculated J-coupling was within 0.4 Hz of the NMR-derived J-coupling (Taha et al. 2010). However, in 7 of the 17 angles examined, the DFT-calculated J-couplings had an error of more than 1.5 Hz (Taha et al. 2010). These inaccuracies could be due to the MD simulations themselves, which are based on empirical force fields and are approximate in their estimates of conformational probabilities.

The purpose of the current study is to assess whether the accuracy of theoretical J-couplings could be improved by sampling from the thermodynamic ensemble while minimizing the errors that could come from inaccurate sampling of multiple conformations. We selected carbohydrate-like compounds with relatively rigid COCH dihedral angles (Fig. 1) by virtue of the central bond being located in a ring or macrocycle. We gathered experimentally measured J-coupling values from the literature, many of which were measured using a selective 2-D J resolved experiment (Mulloy et al. 1988; Tvaroška et al. 1989). For each molecule we carried out ab initio molecular dynamics (AIMD) simulations using DFT and implicit solvent models. From the AIMD simulation trajectories, structures were sampled at regular intervals to be used as input to QM J-coupling calculations, and the resulting values were averaged to obtain the ensemble-averaged coupling. To quantify the effect of ensemble-averaging, the level of agreement of these ensemble-averaged values was then compared with control values computed by energy minimization of the crystal structure followed by a single-point calculation of the J-coupling.

It was found in this study that conformational sampling introduces a significant shift in the ensemble-averaged J-coupling compared with the value calculated from a single energy-minimized structure. This is most apparent at dihedral angle values close to 0 degrees, where past studies have shown that J-coupling calculations from energy-minimized structures had greater deviations from experiment. Accounting for conformational sampling reduced the root-mean-squared error (RMSE) between the calculated and experimentally measured J-coupling from 0.77 to 0.52 Hz. Our results indicate that including the effects of conformational sampling can significantly improve the accuracy of predicting J-couplings measured via NMR.

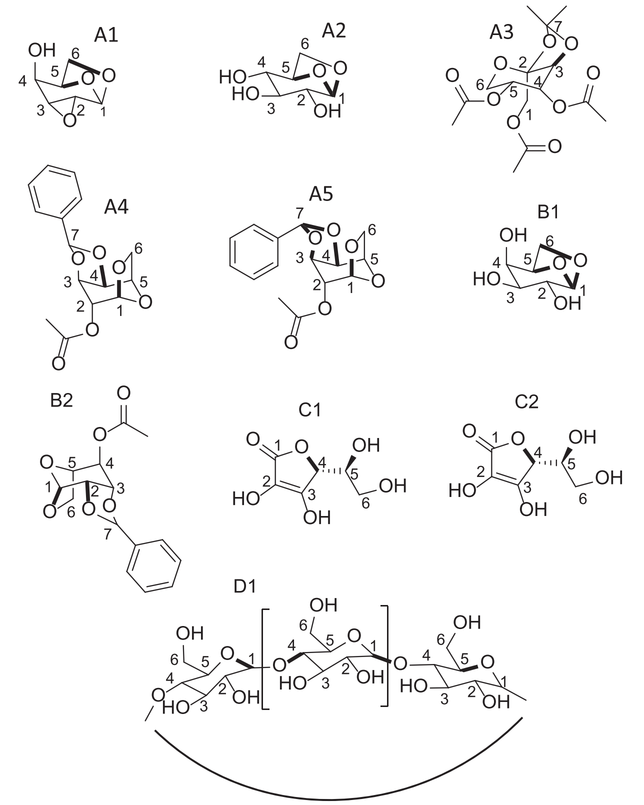

Fig. 1. Structures of molecules investigated in this study. Each molecule has 1 or more rings resulting in some degree of rigidity. Bonds in bold represent central bond of dihedral angles examined. The cyclodextrin molecule (D1) has α-glycosidic linkages.

Materials and methods

Selection of experimental data

The molecules selected for this study were carbohydrate-like molecules with known experimental J-coupling values for rigid COCH dihedral angles (Berking and Seeman 1971; Park et al. 1971; Koll and Kopf 1976; Bock and Pedersen 1977; Ogawa et al. 1977; Hamer et al. 1978; Paukstelis et al. 1982; Cano et al. 1986; Tvaroška et al. 1989; Tvaroska and Taravel 1995). Many of these molecules were taken from (Tvaroška et al. 1989), in which the authors also reported a Karplus relation for the COCH coupling. All of these dihedral angles are part of rings or macrocycles, and are not expected to populate multiple distinct free energy minima (Minch 1994), and therefore the effects of ensemble-averaging may be simulated on ab initio MD-accessible time scales (<1 ns; Catoire et al. 1997). The molecules and couplings used and their sources are included in Fig. 1 and Table 1. The initial structures of the molecules used in the MD simulations were derived from X-ray crystal structures downloaded from the Cambridge Structural Database.

Energy minimization and ab initio MD calculations

The calculations in this study can roughly be divided into those that produce a sequence of structures (energy minimization and AIMD) and those that compute the single-point J-couplings for a given structure. The calculations in the

Table 1. Molecules and angles.

Label	References	CSD ID	Angle number	Vicinal atoms	Crystal structure dihedral angle (degrees)	Experi-mental J (Hz)	Energy-minimized dihedral angle (degrees)	Energy-minimized J (Hz)	MD-average dihedral angle (degrees)	MD-average J (Hz)
A1	Tvaroška et al. 1989	AHGULP10	1	H6(endo)-C1	109.00	1.4	111.88	1.58	113.15	1.79
			2	H6(exo)-C1	−128.53	2.2	−127.89	2.21	−127.40	2.28
			3	H5-C1	153.85	5.6	157.86	5.99	157.00	5.96
A2	Tvaroška et al. 1989	AHGLPY01	4	H1-C6	140.90	5.1	141.73	5.81	139.63	5.22
			5	H1-C5	−160.87	5.1	−161.70	5.30	−160.86	5.12
			6	H5-C1	164.03	5.9	163.53	5.82	164.44	6.03
A3	Tvaroška et al. 1989	OIPFRP	7	H3-C7	85.69	0	82.20	−0.31	84.26	−0.20
A4	Tvaroška et al. 1989	FOJLUL	8	H4-C7	118.67	2.3	117.31	2.82	111.27	2.05
			9	H1-C6	135.74	5.1	140.30	5.50	139.07	5.40
A5	Tvaroška et al. 1989	FOJMAS	10	H3-C7	−81.56	0	−78.52	0.00	−88.47	0.07
			11	H4-C7	130.85	3.2	120.94	2.98	124.60	3.59
			12	H1-C6	137.02	5.2	142.52	6.98	140.05	5.68
			13	H7-C3	−144.35	4.1	−153.66	6.42	−141.39	4.89
			14	H7-C4	127.26	5.2	136.41	5.03	126.18	3.75
B1	Tvaroska and Taravel 1995	AHGALP	15	H6(endo)-C1	135.59	3.7	124.30	3.48	128.80	4.29
			16	H5-C1	−173.36	6.1	163.20	6.13	163.95	6.15
			17	H1-C5	−147.54	5.0	−161.89	5.30	−160.35	5.07
			18	H1-C6	120.13	5.2	142.15	6.07	138.19	5.03
B2	Tvaroska and Taravel 1995	FOJMEW	19	H2-C7	−116.61	2.8	−120.45	3.28	−116.14	2.89
			20	H1-C6	145.13	6.0	144.53	6.64	142.36	6.35
C1	Paukstelis et al. 1982	COFKOA	21	H4-C1	−112.03	1.87	−121.50	2.36	−119.10	2.01
C2	Paukstelis et al. 1982	LASCAC15	22	H4-C1	−116.67	1.87	−122.37	2.46	−120.65	2.31
D1	Tvaroska and Taravel 1995	GOQZUH	23	H4′-C1	0.61	4.8	7.02	6.26	9.95	5.69
			24	H5-C1	−66.2	N/A	−62.44	1.92	−61.17	2.01
			25	H1-C4′	5.26	4.8	−4.16	6.11	−8.71	5.86
			26	H1-C5	178.47	6.85	178.70	7.21	178.44	7.04

Note: Labels refer to molecules in Fig. 1. CSD ID refers to the X-ray crystal structure found in the Cambridge Structural Database. The 4 columns on the right are data generated by this study.

former category used the B3LYP density functional approximation with a D3(BJ) dispersion correction (Becke 1993; Grimme et al. 2011). The def-TZVP basis set was used for all atoms (Schäfer et al. 1994). The SWIG-PCM implicit solvent model (Lange and Herbert 2010; Liu et al. 2015) was used with the dielectric constant of 78.4 corresponding to liquid water at room temperature (Wagner and Pruß 2002). Seven explicit water molecules were included in the structure of α-cyclodextrin in order to solvate the hydrogen bond donor and acceptors in the ring interior, and no other simulations included explicit waters. These calculations were run using the TeraChem quantum chemistry software (Seritan et al. 2020, 2021). Sample input files for the TeraChem calculations are available in the Supporting Information Tables S1 and S2.

In order to investigate the effects of thermal fluctuations on the computed J-couplings, the thermodynamic ensemble in the neighborhood of the experimental structure was sampled using ab initio MD simulations. These simulations use the quantum-chemical potential energy surface instead of a classical force field, which gives increased accuracy at significantly increased computational cost (Islam and Roy 2012). This was done in order to eliminate the possible errors in results that could come from using a force field, and the molecules should be sufficiently rigid that long trajectories are not needed to cross over barriers and sample from multiple potential energy basins (Tvaroška et al. 1989). The MD simulations were carried out using a time step of 1.0 fs, at constant temperature (300 K) using Langevin dynamics and a time constant of 300 steps. A total of 100.0 ps of MD simulation was carried out for each molecule. Structures were saved at 1.0-ps intervals after a 10.0-ps equilibration period; for each AIMD trajectory, J-couplings were calculated as averages over 90 structures.

Calculating the J-coupling

Single-point calculations of the J-couplings were performed with Q-Chem 4.4 (Shao et al. 2015) using the B3LYP density functional approximation and SWIG-PCM implicit solvent model, which was also used in the TeraChem calculations. These J-coupling calculations used specialized basis sets needed to describe the properties of the region of the wave function close to the nucleus (Jensen 2006). Our procedure follows (Helgaker et al. 2016), where it was demonstrated that couplings calculated with B3LYP/pcJ-1 produced highly accurate results with an RMSE of 0.18 Hz vs. experiment for the strychnine molecule (Helgaker et al. 2016).

There are 4 physical mechanisms that contribute to the calculated J-coupling: Fermi contact (FC), Paramagnetic spin orbital (PSO), Diamagnetic spin orbital (DSO), and spin dipole (SD; Ramsey 1953). The FC mechanism is the result of the change in interaction energy between nuclear moments and electron spins when the nuclear spin changes in response to radiation (Minch 1994). The spin orbital mechanisms arise from changing nuclear magnetic moments inducing orbital electronic fields that influence the nuclear magnetic moment of the coupled proton. The DSO mechanism arises from the effect from electrons in the ground state, whereas the PSO mechanism takes into account all excited states. Finally, the SD mechanism takes into account the effect of the change in the nuclear magnetic moment on its coupled nuclear magnetic moment through changes in electric magnetic moments (Minch 1994). Of these mechanisms, FC is by far the dominant component (San-Fabián et al. 1993), and is cheaper to compute than the remaining 3 mechanisms that make up the J-coupling (Bally and Rablen 2011). Previous studies have demonstrated that the non-FC components of the J-coupling can be calculated with a smaller basis set with the effect of significantly reducing computational costs while sacrificing little accuracy (Fukui et al. 1995;Bally and Rablen 2011 ; Helgaker et al. 2016). In some studies, reasonably-accurate J-couplings were predicted even when the non-FC components were completely omitted from the calculations (Bally and Rablen 2011; Helgaker et al. 2016). For this reason a hybrid calculation was used in this study, where the FC was calculated using the pcJ-1 basis set while the remaining components of the coupling were calculated using the smaller pcJ-0 basis set following the procedure developed by (Helgaker et al. 2016). An additional set of calculations was performed on the energy-minimized structures using the larger pcJ-1 basis set for the entire J-coupling to test the validity of this approach. Sample Q-Chem input files are available in Supporting Information, Tables S3 and S4. For the energy-minimized structures with implicit solvent, the RMSE for the hybrid basis approach vs. pcJ-1 for all components is 0.131 Hz, which is small compared with the standard deviation of ∼0.5 Hz that we calculated from multiple experimental measurements of the same J-coupling in the literature, such as angles 3, 10, 12, 13, 25, and 26 (Berking and Seeman 1971; Park et al. 1971; Koll and Kopf 1976; Hamer et al. 1978; Cano et al. 1986; Mulloy et al. 1988; Tvaroška et al. 1989; Tvaroska and Taravel 1995). Therefore, we deemed the hybrid basis approach to be appropriate for including the non-FC contributions to the J-coupling.

Statistics and analytical methods

To determine whether ensemble-averaging J-couplings from multiple structures derived from AIMD improved agreement with experiment, the RMSE for the ensemble-averaged J-couplings was compared with the RMSE for the energy-minimized ones. A 1-tailed *t*-test was performed to assess the statistical significance of any observed improvement in agreement with the experimental J-couplings when accounting for ensemble-averaging. For a *t*-test to provide the most reliable results, the 2 groups being compared typically need to meet 3 conditions: The groups must be of equal size, have equal variance, and be normally distributed (Havlicek and Peterson 1974). If the groups are not normally-distributed, a reliable *t*-statistic may still be used as long as the first 2 conditions are met and the distributions are the same shape (Havlicek and Peterson 1974). Since both the ensemble-averaged and energy-minimized groups were of equal size, had roughly equal variance, and had left-skewed distributions, the conditions for the *t*-test were met.

A closer examination of the distribution of data points in the ensemble-averaged and energy-minimized groups revealed data points far from the range where most data points were clustered. Tukey's rule was applied to identify potential outliers, where an outlier was defined as any data point >1.5 times the value of the interquartile range greater than the value of the upper quartile (Hoaglin et al. 1986). Four data points were identified as potential outliers, corresponding to angles 13, 14, 23, and 25. The RMSEs of the J-couplings are calculated twice by including and excluding the potential outliers from the data set.

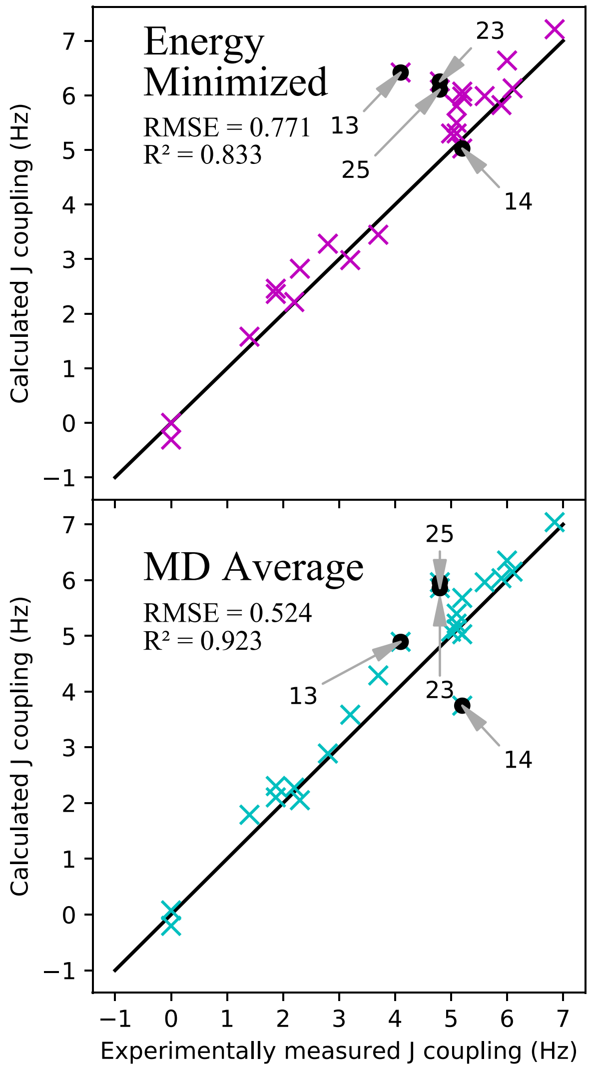

Fig. 2. Calculated vs experimental J-couplings. The level of disagreement between theory and experiment is indicated by the vertical distance of data points from the diagonal line. Potential outliers are labeled in black with arrows indicating angle numbers.

Results and discussion

Accuracy of calculated J-couplings—energy-minimized vs ensemble-averaged

A total of 25 dihedral angles in this study had experimental J-couplings available. For each angle, 2 J-coupling values were calculated: A value taken from the energy-minimized structure and an ensemble-averaged value from the MD snapshots. A comparison of the calculated J-coupling and the experimental J-coupling in each group can be seen in Fig. 2.

Overall, the RMSE in the MD-averaged group was 0.52 Hz, which was lower than the RMSE of 0.77 Hz in the energy-minimized group. A single-tailed *t*-test with 24 degrees of freedom was performed to determine whether this difference was statistically significant. The *t*-statistic was −1.69 with a corresponding *P*-value of 0.052. This indicates that there is a 94.8% confidence level that the difference in accuracy between the MD-averaged J-couplings and the energy-minimized J-couplings are statistically significant. A second

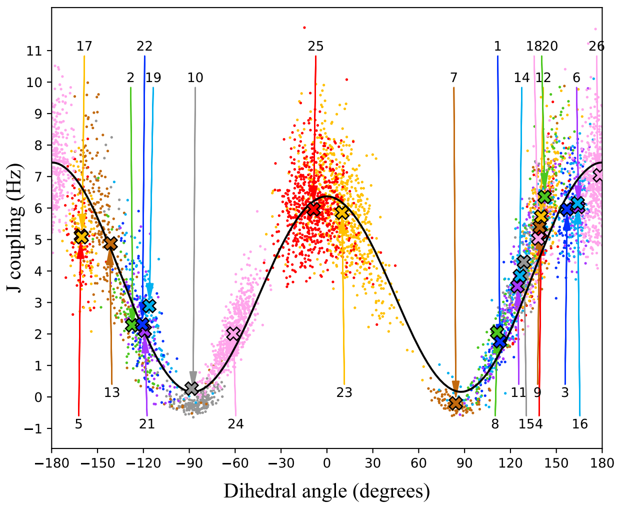

Fig. 3. Correlation between computed J-coupling and dihedral angles. Points represent calculated J-couplings and dihedral angles for every structure from the 90 MD snapshots for each angle. Points belonging to each angle are grouped by color coding with an arrow pointing to the corresponding angle number from Table 1. The "X" symbol represents the average J-coupling and dihedral angle value in degrees for each angle. Curve with solid line represents 3-parameter Karplus curve fitted to the computational data.

set of statistical calculations was performed with the 4 potential outliers omitted to determine the effects of the outliers on the results. Without outliers, the RMSE for the MD-averaged J-couplings was 0.28 Hz and the RMSE for the energy-minimized J-couplings was 0.45 Hz. This difference was again statistically significant with a *P*-value of 0.011 indicating a 98.9% confidence level.

A possible origin of systematic error in J-coupling calculations

The J-couplings calculated from the energy-minimized structures tended to be higher than the experimental J-couplings, especially in angles where the experimental *J*-coupling was in the range of 4–6 Hz. The ensemble-averaged J-couplings derived from the MD samples were still higher than the experimental J-couplings, but to a lesser extent. From these results, it is possible that the experimental J-couplings are lower than the energy-minimized computed ones due to ensemble-averaging. A plot of calculated J-couplings vs dihedral angles for all conformations examined was created to identify explanations for this trend in Fig. 3.

In the scatter plot of the *J*-coupling vs dihedral angle (Fig. 3), 2 distinct local minima and local maxima can be observed as expected from the Karplus relationship. The areas near −90 and +90 degrees correspond to lower computed *J*-couplings, and in these regions the coupling does not change by a large amount away from the local minima. For instance, in the 60-degree areas ranging from −120 to −60 degrees and 60–120 degrees, the J-coupling only ranges from approximately −0.5 to 0.5 Hz. Therefore, in angles where the J-coupling is low, the average J-coupling stays close to any single *J*-coupling observed in a randomly-selected conformation. This can explain why errors in the energy-minimized group were smaller for angles with smaller J-couplings. In contrast, the local maxima in the *J*-couplings have larger fluctuations away from the average. In the 60-degree area ranging from −30 to 30 degrees, the calculated single-point J-couplings range widely, from 3.5 to 10 Hz. Although the average *J*-coupling for angles in this region are close to 5 Hz, it would be reasonable for a randomly-selected conformation to be off by ∼2 Hz. Moreover, due to the downward curvature in the Karplus relation in the neighborhood of the local maxima, thermal fluctuations in the dihedral angle have the effect of systematically lowering the calculated coupling, compared with a single-point calculation near the maximum. This may explain why errors in the energy-minimized group were higher for angles with larger J-couplings and why accounting for conformational averaging significantly reduced this error.

Effect of omitting non-Fermi-contact contributions

The FC makes up the largest contribution toward the total *J*-coupling and represents only a small fraction of the total computational cost of calculating the *J*-coupling using ab initio calculations. If the contribution of the non-FC coupling components is small enough, it may be possible to considerably reduce computational costs by omitting the non-FC components from the *J*-coupling calculation without significantly reducing accuracy.

To test this idea 2 sets of ensemble-averaged *J*-couplings were compared with the experimental values: One set where all components were calculated, and one where only the FC contribution was calculated. The RMSE between the experimental *J*-coupling and the total J-coupling was 0.52 Hz, and the R^2 value was 0.92. In contrast, when the FC-only *J*-coupling was compared with experiment, the RMSE increased only slightly to 0.57 Hz, whereas the R^2 value was reduced to 0.91. Even after neglecting the non-FC contributions, the ensemble-averaged values were still more accurate than the energy-minimized results, which had an RMSE with experiment of 0.77 Hz and an R^2 value of 0.83. The loss in accuracy when calculating only the FC portion of the J-coupling was relatively minor. In exchange for the slightly reduced accuracy, the FC-only calculations in α-cyclodextrin took 33% less CPU time to complete than the hybrid pcJ-1 and pcJ-0 approach. In the smaller molecules labeled A through C, the FC-only calculations only took 50% as much CPU time to complete than the hybrid approach. It may therefore be justifiable to omit the non-FC coupling components from the J-coupling calculations to save computational costs without sacrificing much accuracy, especially for smaller molecules where greater savings can be expected.

Deriving a new Karplus equation for predicting J-couplings from MD simulations

A 3-parameter Karplus equation was fitted to the dihedral angles and J-couplings derived from the conformations extracted from the MD simulations. There were a few regions in Fig. 3 that had greater data point density than others. Weights were applied to each data point to prevent a fitting bias in the denser regions of the plot, with the value of each weight determined by counting the number of points within 5 degrees of each point and defining the weight for each point by the reciprocal of that count. The resulting equation can be seen superimposed over the MD-derived data in Fig. 3

depicted by a black line. The resulting equation for the 3-parameter curve is:

$$J = 6.735\cos^2\phi - 0.538\cos\phi + 0.173$$

The fit of the new Karplus equation was evaluated by calculating the RMSE of the curve vs the data points. The resulting RMSE for the newly fitted curve was 1.177 Hz. RMSEs were also calculated for existing empirically derived Karplus equations for the COCH angle. These equations were developed by Tvaroška et al. (1989), Mulloy et al. (1988), and Anderson and Ijeh (1994). Supporting Information Fig. S1 (see online supplementary material for a color version of this figure) shows these experimental Karplus curves as well as the newly fitted Karplus curve over the dihedral angle and J-coupling data from the MD simulations. The (Tvaroška et al. 1989) equation had the lowest RMSE of the empirical equations at 1.257 Hz, which was ~6% higher than the RMSE for the newly fitted equation. This demonstrates that the newly-developed Karplus equation is suitable for predicting J-couplings from MD-derived data.

There are a few possible reasons for the success of the new Karplus equation. One explanation is that more data were available to complete the curve fitting. In the fittings of the pre-existing empirical Karplus relations, for example, crystal structures with angles in the −60 to 60 degree region were rare (Tvaroška et al. 1989), resulting in limitations in predicting J-couplings in this region. Another factor is the aforementioned use of experimental crystal structures from which the dihedral angles were obtained for fitting the NMR J-couplings, rather than taking an ensemble average of structures, which is more representative of experimental conditions.

Effects attributed to incomplete sampling or approximate potential energies

The MD-averaged J-couplings tended to estimate the experimental couplings more closely than the energy-minimized results. However, in angle 14 there was a significant difference between the MD-average J-coupling and the experimental J-coupling, and the energy-minimized result was actually more accurate for this angle by ~1.3 Hz. Errors in the single-point J-coupling calculations for the angle 14 conformations are not suspected, because the data points for angle 14 in Fig. 3 are consistent with the surrounding data for other angles and follow the expected sinusoidal pattern; if the single-point J-coupling calculations were performed incorrectly, the Karplus-like relationship would likely not be preserved.

The MD trajectory of molecule A5 containing angle 14 was observed to identify potential sources of error. Approximately 50 ps into the simulation, a ring flip was observed in the 5-membered ring that angle 14 was part of. The 5-membered ring almost immediately reverted to its original state after the ring flip. During this brief conformational change, the value of the dihedral angle for angle 14 changed from ~144 to ~79 degrees, and the corresponding calculated couplings decreased from 6.32 to −0.11 Hz. The simulation length was extended to 300 ps to determine if the flipped ring conformation became stable and to capture any alternative conformational data that may not have been captured in the original 100-ps simulation time and could have contributed to a higher ensemble-averaged J-coupling in an experimental setting. The brief ring flip was observed approximately every 50 ps; however the flipped conformation did not stabilize throughout the extended simulation and the MD-averaged J-coupling for angle 14 remained approximately the same for the 100 and 300 ps trajectories. Since the MD-averaged J-coupling did not improve even when the simulation length was extended 3-fold, no strong evidence was found to indicate that the observed error was the result of a lack of conformational sampling.

Additional angles in molecule A5 were investigated to determine whether the brief fluctuations in the conformation of the 5-membered ring were responsible for the increased error seen in angle 14. Supporting information Fig. S2 (see online supplementary material for a color version of this figure) shows the values of the dihedral angles for angles 12, 13, and 14. The 2 angles that were part of the 5-membered ring, 13 and 14, have brief and significant changes in dihedral angles that accompany a ring flip observed in the trajectory. Angle 12, which was not part of the 5-membered rings, did not display such large fluctuations. Compared with the energy-minimized J-couplings, the MD-averaged J-couplings for both 13 and 14 were both lower by ~1.5 Hz. In angle 13, where the energy-minimized J-coupling overestimated the experimental value by ~2.3 Hz, the lowered MD-average J-coupling resulted in a better fit with the experiment. However in angle 14 the energy-minimized J-coupling was already lower than the experimental J-coupling, which made the lowered MD-average J-coupling a poorer fit. Therefore, we think that the observed ring flipping does not worsen (or improve) overall agreement with experiment.

A surprising result apparent from Fig. 3 is the significant fluctuation of the J-couplings away from the ensemble average at a given dihedral angle, with a significant number of structures deviating by 2 Hz or more from the fitted Karplus curve. Because any fluctuations in the J-couplings would have to originate from the other degrees of freedom in the chemical environment of the coupled nuclei, we looked for correlations between the coupling and local degrees of freedom (i.e. C–O, O–C, C–H bond stretching; C–O–C, O–C–H angle bending) for structures selected from a narrow window of the dihedral angle (10 degrees) (Supporting information, Fig. S3, see online supplementary material for a color version of this figure). When angles 20 and 25 were examined within a 10 degree window of ~155 degrees and −15 degrees, respectively, we found a significant negative correlation between the J-coupling and the central O–C bond stretch. We created a modified model to incorporate the central O–C bond length into the estimation of the J-coupling in the form: $J = \left(A\cos^2\phi + B\cos\phi + C\right)\left(1 + b\left(r - r_0\right)\right)$. The fitted equation was found to be:

$$J = (8.870\cos^2\phi - 0.751\cos\phi + 0.319)\,(1 - 2.557\,(r - 1.330))$$

A visual representation of this relationship can be seen in Fig. 4. Incorporating the central O–C bond length into the model lowered the RMSE with our data by ~0.15 Hz compared with the Karplus relation without the bond length term. Although the lowering of the RMSE is significant, we expect it could be lowered further by incorporating more features of the

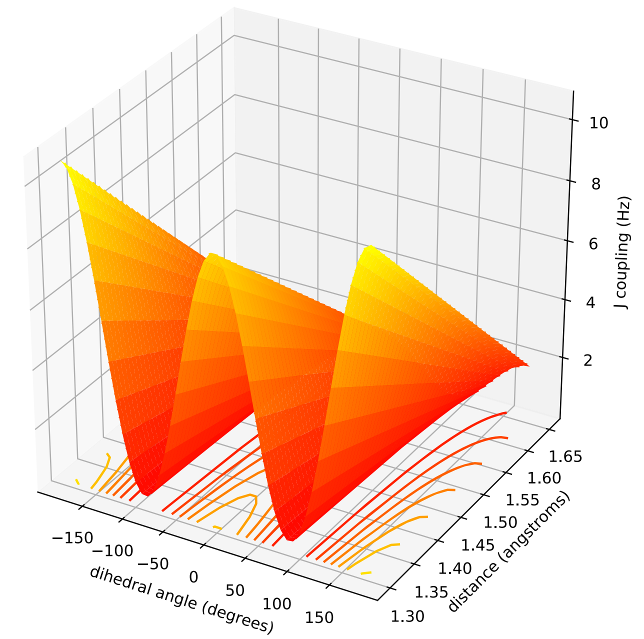

Fig. 4. J-coupling vs dihedral angle and central bond length. Modified Karplus curve with distance parameters is demonstrated by 3-D contour. Level curves are shown on XY plane.

chemical environment, possibly using neural network models following longstanding methods for predicting chemical shifts (Han et al. 2011; Yang et al. 2021). These are promising areas for future study.

Conclusion

The present study indicates that ensemble-averaging using AIMD simulations has a significant effect on computed *J*-couplings and could lead to improved agreement with experiment. The Karplus relation developed in this study is expected to be more reliable for estimating *J*-couplings from simulations, because the data used to fit the relation came from single-point calculations on different individual structures, rather than a single ensemble-averaged value.

Accounting for conformational sampling with QM methods may still prove difficult in more flexible molecules; since less-rigid molecules have greater degrees of freedom in movement, it may not always be possible to access the molecule's full range of conformations within ab initio MD-accessible timeframes without using additional approximations. For more challenging cases, umbrella sampling approaches may provide accurate estimates of relative free energies between rotamers, and hybrid QM/MM simulations that use polarizable solvent models could improve the description of solvent effects beyond the implicit solvent employed here. Overall, the ability to compute J-couplings more accurately will provide greater physical insight into the structures of glycoproteins and carbohydrate-protein interactions as well as their biological functions.

Supplementary data

Supplementary material is available at *Glycobiology* Journal online.

Supporting Information for: The Impact of Conformational Sampling on First-Principles Calculations of Vicinal J-Couplings in Carbohydrates: Figs. S1–S3 and Tables S1–S4.

Acknowledgments

We would like to acknowledge Xi Chen, Daron Freedberg, Lisa Oh, and Ajit Varki for helpful discussions.

Data availability statement

The simulation data that support the results of this study are available from the corresponding author, L.-P. W., upon reasonable request.

Funding

This work was funded by The United States National Institutes of Health (NIH) grant number R014AI130684.

Conflict of interest statement

None declared.

References

Anderson JE, Ijeh AI. Eclipsed ground-state conformations for methoxycyclohexanes with adjacent methyl-group substitution. An NMR criterion and molecular mechanics calculations. *J Chem Soc Perkin Trans*. 1994:**2**(9):1965–1967. https://doi.org/10.1039/P29940001965.

Bally T, Rablen PR. Quantum-chemical simulation of 1H NMR spectra. 2. Comparison of DFT-based procedures for computing proton–proton coupling constants in organic molecules. *J Org Chem*. 2011:**76**(12):4818–4830. https://doi.org/10.1021/jo200513q.

Becke AD. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys*. 1993:**98**(7):5648–5652. https://doi.org/10.1063/1.464913.

Berking B, Seeman NC. The crystal structure of 1,6:2,3-dianhydro-β-d-gulopyranose. *Acta Crystallogr B*. 1971:**27**(9):1752–1760. https://doi.org/10.1107/S0567740871004746.

Bock K, Pedersen C. Reaction of carbohydrates with hydrogen bromide – preparation of some 6-deoxy-D-mannofuranoses. *Acta Chem Scand Ser B-Org Chem Biochem*. 1977:**31**(3):248–250. https://doi.org/10.3891/acta.chem.scand.31b-0248.

Cano FH, Foces-Foces C, Jimenez-Barbero J, Bernabe M, Martin-Lomas M. Application of vicinal carbon-proton coupling constants to the conformational analysis of benzylidene-type acetals of 1,6-anhydro-β-d-hexopyranoses. *Carbohydr Res*. 1986:**155**:1–10. https://doi.org/10.1016/S0008-6215(00)90128-9.

Cano FH, Foces-Foces C, Jimenez-Barbero J, Alemany A, Bernabe M, Martin-Lomas M. Experimental evidence of deviations from a Karplus-like relationship of vicinal carbon-proton coupling constants in some conformationally rigid carbohydrate derivatives. *J Org Chem*. 1987:**52**(15):3367–3372. https://doi.org/10.1021/jo00391a036.

Case DA, Scheurer C, Brüschweiler R. Static and dynamic effects on vicinal scalar J couplings in proteins and peptides: a MD/DFT analysis. *J Am Chem Soc*. 2000:**122**(42):10390–10397. https://doi.org/10.1021/ja001798p.

Catoire L, Braccini I, Bouchemal-Chibani N, Jullien L, du Penhoat CH, Perez S. NMR analysis of carbohydrates with model-free spectral densities: the dispersion range revisited. *Glycoconj J*. 1997:**14**(8):935–943. https://doi.org/10.1023/A:1018518928122.

Cloran F, Carmichael I, Serianni AS. Density functional calculations on disaccharide mimics: studies of molecular geometries and trans-O-glycosidic 3JCOCH and 3JCOCC spin-couplings. *J Am Chem Soc*. 1999:**121**(42):9843–9851. https://doi.org/10.1021/ja984384t.

Coxon B. 2009. Chapter 3 developments in the Karplus equation as they relate to the NMR coupling constants of carbohydrates. In: *Advances in carbohydrate chemistry and biochemistry*. Vol. **62**. Baltimore, MD: Academic Press. p. 17–82. [accessed 9 October 2021]. https://www.sciencedirect.com/science/article/pii/S0065231809000031.

Engelsen SB, Pérez S, Braccini I, Du Penhoat CH. Internal motions of carbohydrates as probed by comparative molecular modeling and nuclear magnetic resonance of ethyl β-lactoside. *J Comput Chem*. 1995:**16**(9):1096–1119. https://doi.org/10.1002/jcc.540160905.

Fukui H, Inomata H, Baba T, Miura K, Matsuda H. Calculation of nuclear spin–spin couplings. VIII. Vicinal proton–proton coupling constants in ethane. *J Chem Phys*. 1995:**103**(15):6597–6600. https://doi.org/10.1063/1.470388.

Grimme S, Ehrlich S, Goerigk L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem*. 2011:**32**(7):1456–1465. https://doi.org/10.1002/jcc.21759.

Hackbusch S, Watson A, Franz AH. Development of a Karplus equation for 3JCOCH in ester-functionalized glucopyranoses and methylglucuronate. *Arkivoc*. 2017:5:268–292. https://doi.org/10.24820/ark.5550190.p010.113.

Hamer GK, Balza F, Cyr N, Perlin AS. A conformational study of methyl β-cellobioside-d8 by 13C nuclear magnetic resonance spectroscopy: dihedral angle dependence of 3JC–H in 13C—O—C—1H arrays. *Can J Chem*. 1978:**56**(24):3109–3116. https://doi.org/10.1139/v78-508.

Han B, Liu Y, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR*. 2011:**50**(1):43–57. https://doi.org/10.1007/s10858-011-9478-4.

Havlicek LL, Peterson NL. Robustness of the T test: a guide for researchers on effect of violations of assumptions. *Psychol Rep*. 1974:**34**(3_suppl):1095–1114. https://doi.org/10.2466/pr0.1974.34.3c.1095.

Helgaker T, Jaszuński M, Świder P. Calculation of NMR spin–spin coupling constants in strychnine. *J Org Chem*. 2016:**81**(22):11496–11500. https://doi.org/10.1021/acs.joc.6b02157.

Hoaglin DC, Iglewicz B, Tukey JW. Performance of some resistant rules for outlier labeling. *J Am Stat Assoc*. 1986:**81**(396):991–999. https://doi.org/10.1080/01621459.1986.10478363.

Hoch JC, Dobson CM, Karplus M. Vicinal coupling constants and protein dynamics. *Biochemistry*. 1985:**24**(15):3831–3841. https://doi.org/10.1021/bi00336a003.

Islam SM, Roy P-N. Performance of the SCC-DFTB model for description of five-membered ring carbohydrate conformations: comparison to force fields, high-level electronic structure methods, and experiment. *J Chem Theory Comput*. 2012:**8**(7):2412–2423. https://doi.org/10.1021/ct200789w.

Jensen F. The basis set convergence of spin−spin coupling constants calculated by density functional methods. *J Chem Theory Comput - J CHEM THEORY COMPUT 2*. 2006:**2**:1360–1369. https://doi.org/10.1021/ct600166u.

Karplus M. Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys*. 1959:**30**(1):11–15. https://doi.org/10.1063/1.1729860.

Karplus M. Vicinal proton coupling in nuclear magnetic resonance. *J Am Chem Soc*. 1963:**85**(18):2870–2871. https://doi.org/10.1021/ja00901a059.

Koll P, Kopf J. Derivatives of fructose in unusual conformation - crystal and molecular-structure of 1,4,5-tri-omicron-acetyl-2,3 -omicron-isopropylidene-beta-D-fructopyranos E. *Chem Berichte-Recl*. 1976:**109**(10):3346–3357. https://doi.org/10.1002/cber.19761091010.

Krivdin LB. Computational NMR of carbohydrates: theoretical background, applications, and perspectives. *Molecules*. 2021:**26**(9):2450. https://doi.org/10.3390/molecules26092450.

Lange AW, Herbert JM. A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: the switching/Gaussian approach. *J Chem Phys*. 2010:**133**(24):244111. https://doi.org/10.1063/1.3511297.

Li W, Battistel MD, Reeves H, Oh L, Yu H, Chen X, Wang L-P, Freedberg DI. A combined NMR, MD and DFT conformational analysis of 9-O-acetyl sialic acid-containing GM3 ganglioside glycan and its 9-N-acetyl mimic. *Glycobiology*. 2022:**30**(10):787–801. https://doi.org/10.1093/glycob/cwaa040. [accessed 2020 Sep 18]. https://academic.oup.com/glycob/advance-article/doi/10.1093/glycob/cwaa040/5826789.

Liu F, Luehr N, Kulik HJ, Martínez TJ. Quantum chemistry for solvated molecules on graphical processing units using polarizable continuum models. *J Chem Theory Comput*. 2015:**11**(7):3131–3144. https://doi.org/10.1021/acs.jctc.5b00370.

Minch MJ. Orientational dependence of vicinal proton-proton NMR coupling constants: the Karplus relationship. *Concepts Magn Reson*. 1994:**6**(1):41–56. https://doi.org/10.1002/cmr.1820060104.

Mulloy B, Frenkiel TA, Davies DB. Long-range carbon—proton coupling constants: application to conformational studies of oligosaccharides. *Carbohydr Res*. 1988:**184**:39–46. https://doi.org/10.1016/0008-6215(88)80004-1.

Ogawa T, Uzawa J, Matsui M. A 13C-n.m.r. study on the conformation of L-ascorbic acid in deuterium oxide. *Carbohydr Res*. 1977:**59**(2):C32–C35. https://doi.org/10.1016/S0008-6215(00)83210-3.

Park YJ, Kim HS, Jeffrey GA. The crystal structure of 1,6-anhydro-β-d-glucopyranose. *Acta Crystallogr B*. 1971:**27**(1):220–227. https://doi.org/10.1107/S0567740871001936.

Paukstelis JV, Mueller DD, Seib PA, Lillard DW. NMR spectroscopy of ascorbic acid and its derivatives. In: *Ascorbic acid: chemistry, metabolism, and uses*. Vol. **200**. Washington, D.C.: American Chemical Society. (Advances in Chemistry); 1982. pp. 125–151. [accessed 18 September 2020] https://doi.org/10.1021/ba-1982-0200.ch006

Ramsey NF. Electron coupled interactions between nuclear spins in molecules. *Phys Rev*. 1953:**91**(2):303–307. https://doi.org/10.1103/PhysRev.91.303.

Richards MR, Bai Y, Lowary TL. Comparison between DFT- and NMR-based conformational analysis of methyl galactofuranosides. *Carbohydr Res*. 2013:**374**:103–114. https://doi.org/10.1016/j.carres.2013.03.030.

San Fabián J, Omar S, García de la Vega JM. Computational protocol to evaluate side-chain vicinal spin–spin coupling constants and karplus equation in amino acids: alanine dipeptide model. *J Chem Theory Comput*. 2019:**15**(7):4252–4263. https://doi.org/10.1021/acs.jctc.9b00131.

San-Fabián J, Guilleme J, Díez E, Lazzeretti P, Malagoli M, Zanasi R. Vicinal proton—proton coupling constants. Basis set dependence in SCF ab initio calculations. *Chem Phys Lett*. 1993:**206**(1):253–259. https://doi.org/10.1016/0009-2614(93)85549-4.

Schäfer A, Huber C, Ahlrichs R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J Chem Phys*. 1994:**100**(8):5829–5835. https://doi.org/10.1063/1.467146.

Seritan S, Bannwarth C, Fales BS, Hohenstein EG, Kokkila-Schumacher SIL, Luehr N, Snyder JW, Song C, Titov AV, Ufimtsev IS et al. TeraChem: accelerating electronic structure and *ab initio* molecular dynamics with graphical processing units. *J Chem Phys*. 2020:**152**(22):224110. https://doi.org/10.1063/5.0007615.

Seritan S, Bannwarth C, Fales BS, Hohenstein EG, Isborn CM, Kokkila-Schumacher SIL, Li X, Liu F, Luehr N, Snyder JW et al. TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIREs Comput Mol Sci*. 2021:**11**(2):e1494. https://doi.org/10.1002/wcms.1494. [accessed 8 July 2022]. https://onlinelibrary.wiley.com/doi/10.1002/wcms.1494.

Shao Y, Gan Z, Epifanovsky E, Gilbert ATB, Wormit M, Kussmann J, Lange AW, Behn A, Deng J, Feng X et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol Phys*. 2015:**113**(2):184–215. https://doi.org/10.1080/00268976.2014.952696.

Tafazzoli M, Ghiasi M. Conformational study of anomeric center in some carbohydrate derivatives. *J Mol Struct THEOCHEM*. 2007:**814**(1):127–130. https://doi.org/10.1016/j.theochem.2007.03.006.

Taha HA, Castillo N, Sears DN, Wasylishen RE, Lowary TL, Roy P-N. Conformational analysis of arabinofuranosides: prediction of 3JH,H using MD simulations with DFT-derived spin−spin coupling profiles. *J Chem Theory Comput*. 2010:**6**(1):212–222. https://doi.org/10.1021/ct900477x.

Toukach FV, Ananikov VP. Recent advances in computational predictions of NMR parameters for the structure elucidation of carbohydrates: methods and limitations. *Chem Soc Rev*. 2013:**42**(21):8376–8415. https://doi.org/10.1039/C3CS60073D.

Tvaroska I, Taravel FR. 1995. Carbon-proton coupling constants in the conformational analysis of sugar molecules. In: Horton D, editor. *Advances in carbohydrate chemistry and biochemistry*. Vol. **51**. Bratislava, Slovak Republic: Academic Press. p. 15–61. [accessed 18 September 2020]. http://www.sciencedirect.com/science/article/pii/S0065231808601912.

Tvaroška I, Hricovíni M, Petráková E. An attempt to derive a new Karplus-type equation of vicinal proton-carbon coupling constants for C-O-C-H segments of bonded atoms. *Carbohydr Res*. 1989:**189**:359–362. https://doi.org/10.1016/0008-6215(89)84112-6.

Wagner W, Pruß A. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *J Phys Chem Ref Data*. 2002:**31**(2):387–535. https://doi.org/10.1063/1.1461829.

Wang AC, Bax A. Reparametrization of the Karplus relation for 3J(H.alpha.-N) and 3J(HN-C') in peptides from uniformly 13C/15N-enriched human ubiquitin. *J Am Chem Soc*. 1995:**117**(6):1810–1813. https://doi.org/10.1021/ja00111a021.

Yang Z, Chakraborty M, White AD. Predicting chemical shifts with graph neural networks. *Chem Sci*. 2021:**12**(32):10802–10809. https://doi.org/10.1039/D1SC01895G.

Zimmer DP, Marino JP, Griesinger C. Determination of homo- and heteronuclear coupling constants in uniformly 13C,15N-labeled DNA oligonucleotides. *Magn Reson Chem*. 1996:**34**(13):S177–S186. https://doi.org/10.1002/(SICI)1097-458X(199612)34:13<S177::AID-OMR76>3.0.CO;2-S.