JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

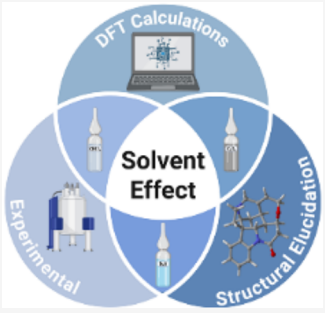
Solvent Matters: Bridging Theory and Experiment in Quantum-Mechanical NMR Structural Elucidation

Iván Cortés,[⊥] Cristina Cuadrado,[⊥] José A. Gavín, María Marta Zanardi, Antonio Hernández Daranas,* and Ariel M. Sarotti*

Cite This: https://doi.org/10.1021/acs.jcim.5c02506 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Quantum-mechanical NMR (QM-NMR) is widely used in structure elucidation. A long-sought holey grail in this field is solving structures from a simple ^{1}H NMR spectrum with AI-driven workflows. Yet, solvent effects on chemical shifts, though long recognized, remain overlooked. We show in a theory−experiment study that implicit solvation models miss solvent-induced variations and introduce a Python tool to quantify solvent sensitivity, aiding more reliable QM-NMR structural assignments.



Nuclear magnetic resonance (NMR) spectroscopy is a fundamental tool for chemists, offering unparalleled capabilities in structural analysis.[1−3] However, despite the advancements in NMR techniques and the wide array of experiments available, errors in structure elucidation persist and are frequently detected and revised.[4,5] Over the past decade, quantum-mechanical (QM) NMR calculations and machine learning approaches have emerged as transformative complements to experiment,[6−12] providing unprecedented accuracy in confirming molecular structures and revising misassigned compounds.[6−12] One of the central methodologies in this approach is the calculation of the so-called DP4 probabilities, which aid in the structural elucidation process by statistically comparing calculated chemical shifts to experimental data.[13−17] In addition to providing robust structural information, this can also reveal specific molecular geometries and interactions with neighboring molecules.[18,19]

NMR measurements are typically performed in solution, and the choice of the solvent is generally guided by sample solubility—making chloroform the default. However, alternative solvents are often required, and it is well-established that solvents can strongly modulate chemical shifts.[20−27] Many NMR practitioners have taken advantage of this effect to aid in structure elucidation, particularly when overlapping signals make assignments ambiguous. While the influence of solvents on chemical shifts is widely acknowledged, clear predictions about the nuclei involved and the magnitude of the changes remain elusive. Although computational methods offer a potential means to address this issue, a comprehensive assessment of how reliably standard DFT captures solvent-induced chemical-shift variations (SIV) is, to the best of our knowledge, still lacking.

This gap matters. Modern QM-NMR methods critically depend on the agreement between the calculated and experimental data. Moreover, in many cases, the distinction between diastereoisomers may hinge on just a few resonances. Even small solvent-induced discrepancies can therefore determine the success or failure in structural assignments. However, how much these SIV influence the robustness of structure determination is still an open question.

In this work, we address this long-standing blind spot with a comprehensive theoretical−experimental study. We reveal how the solvent impacts NMR chemical shifts, benchmark the performance of DFT methods in capturing SIV, and introduce practical tools to incorporate solvent sensitivity into computational workflows. These results establish solvent effects as a decisive factor for robust, accurate, and ultimately automatable NMR-based structure elucidation.

We began the study by selecting a set of 20 molecules with a wide range of solubilities as well as structural and conformational diversity (some representative examples are shown in Figure 1a, and the full set is in Figure S1). Their ^{1}H and ^{13}C NMR spectra were experimentally recorded in nine different deuterated solvents ($CDCl_3$, CD_2Cl_2, acetone-d_6, DMSO-d_6, CD_3CN, benzene-d_6 (C_6D_6), pyridine-d_5, CD_3OD, and D_2O) whenever possible. Hydroxyl and NH resonances were

Received: October 15, 2025
Revised: December 13, 2025
Accepted: January 6, 2026

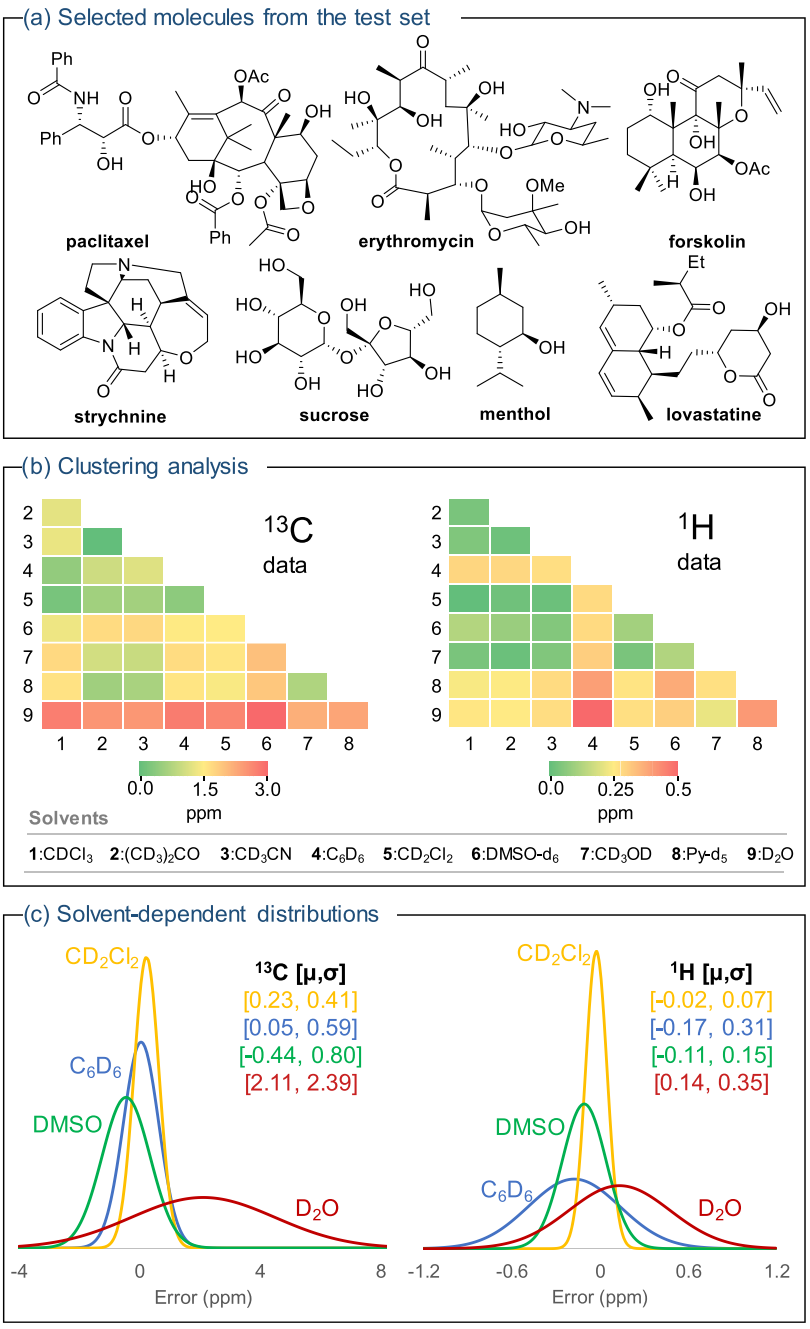


Figure 1. (a) Representative molecules from the test set. For the full list, see Figure S2. (b) AMA-$\Delta\delta$ values measured for all solvent pairs. (c) Normal distribution series obtained by correlating the $\Delta\delta$ values between $CDCl_3$ and four selected solvents: CD_2Cl_2 (yellow), benzene-d_6 (blue), DMSO-d_6 (green), and D_2O (red). The mean (μ) and standard deviation (σ) of each series are shown between the brackets.

excluded because they are generally nondiagnostic and show variable chemical shifts due to hydrogen bonding and rapid rotational/exchange freedom, complicating direct comparison with calculated values. The δ_H and δ_C were assigned by using 2D experiments. To our knowledge, this is the first study to systematically explore such a large and complex set of molecules in this manner. Consistent with previous observations, we noted a much higher variability in response to solvent changes when analyzing ^1H spectra compared to ^{13}C spectra.[21,22,28,29]

With 36 possible combinations of solvent pairs, the detailed analysis of such a large amount of information proved to be challenging. Initially, we studied the MM$\Delta\delta$ parameter, defined as $\mathrm{Max}[\delta i_{1-9}]-\mathrm{Min}[\delta i_{1-9}]$, where δi_{1-9} are the chemical shifts of nucleus i recorded in each of the nine selected solvents. This parameter delivers a global diagnosis to evaluate the sensitivity of a particular atomic environment to solvent variations. The arithmetic mean of the MM$\Delta\delta$ values for each molecule was defined as MA-MM$\Delta\delta$ (mean absolute MM$\Delta\delta$, $\Sigma_n|\mathrm{MM}\Delta\delta|/n$). Finally, the average MA-MM$\Delta\delta$ (AMA-MM$\Delta\delta$) was computed to represent the global behavior of the data set. The resulting AMA-MM$\Delta\delta$ values were 2.6 ppm for ^{13}C and 0.59 ppm for ^1H, with maximum values of MM$\Delta\delta$ (MV-MM$\Delta\delta$) reaching up to 11.4 and 1.9 ppm, respectively. One important issue is that despite the fact that we tried several approaches to observe any meaningful

https://doi.org/10.1021/acs.jcim.5c02506
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

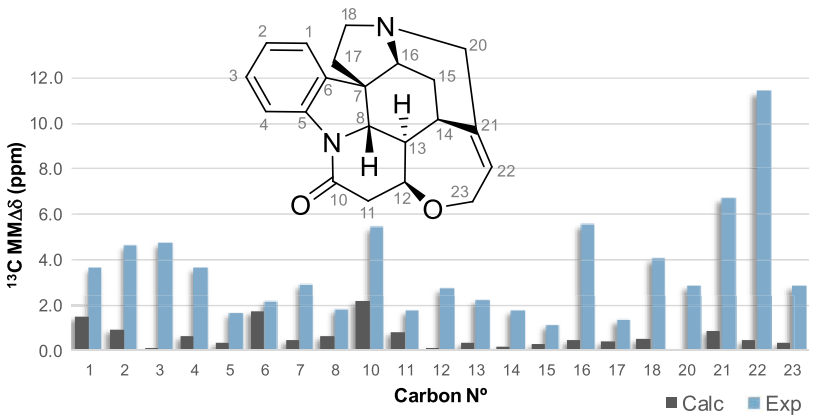


Figure 2. Experimental and calculated ^{13}C MM$\Delta\delta$ values computed for strychnine at the PCM/mPW1PW91/6-31+G**//B3LYP/6-31G* level of theory.

correlation between MM$\Delta\delta$ and the corresponding chemical environment (Figure S2), our failure underscores the inherent challenges in accurately predicting which nuclei are most susceptible to solvent-induced changes. We also analyzed correlations of $\Delta\delta$ and bulk solvent polarity descriptors (dielectric constant, Kamlet−Taft parameters, etc.),[30] but no significant trends were observed—an expected outcome, since solvent-induced changes are site-specific. In addition, we performed subset and outlier analyses (protic vs aprotic solvents), which confirmed the absence of meaningful correlations. This indicates that bulk polarity descriptors alone cannot account for the observed chemical-shift variations (see Table S23, Figures S4 and S5). We also evaluated the possible relationship between SIV and molecular conformational freedom. This analysis aimed to determine whether solvent sensitivity might be associated with changes in the preferred conformations when the chemical environment is modified. For this reason, fairly rigid structures such as strychnine as well as more flexible ones such as erythromycin were included in the experimental panel of molecules. Thus, we identified the total number of possible conformations through a systematic search using the MMFF force field with a 5 kcal/mol energy cutoff. Our results show that the MA-MM$\Delta\delta$ values exhibit no correlation with the number of generated conformations ($R^2 < 0.1$). This indicates that for the molecules studied, SIV does not depend on the complexity or breadth of the accessible conformational space (see Table S28 and Figure S9). In other words, the variation observed across solvents does not appear to originate from conformational redistribution, but rather from other aspects of solute−solvent interactions that are not directly linked to the number of available conformations. This conclusion is further supported by direct comparison of systems with markedly different conformational flexibility. For instance, despite being a rigid structure, strychnine (**1**) exhibits large solvent-dependent chemical-shift deviations (MA-MM$\Delta\delta$ of 3.5 and 0.74 ppm for ^{13}C and ^{1}H, respectively). In contrast, highly flexible molecules (such as lovastatin, **11**) show significantly smaller deviations (MA-MM$\Delta\delta$ of 2.0 and 0.37 ppm for ^{13}C and ^{1}H, respectively).

To explore SIV trends, we clustered solvents based on the MA-$\Delta\delta$ parameter (mean absolute difference in chemical shifts between two solvents). Subsequently, to assess general trends for ^{1}H and ^{13}C, we clustered solvents based on the average mean absolute difference (AMA-$\Delta\delta$, Figure 1b). For ^{1}H data, benzene-d_6, pyridine-d_5, and D_2O showed distinct behaviors while the rest grouped into two clusters: DMSO-d_6/CD_3CN (cluster 1) and $CDCl_3$/CD_2Cl_2/acetone-d_6 (cluster 2). For ^{13}C, D_2O and DMSO-d_6 displayed the largest deviations (2.1 and 1.2 ppm, respectively), whereas other solvents formed two clusters with a lower internal variability (~0.5 ppm). Notably, $CDCl_3$, CD_2Cl_2, and acetone-d_6 showed consistent clustering across both nuclei.

Given $CDCl_3$'s widespread use, we examined its $\Delta\delta$ differences with other solvents in more detail. In all cases, $\Delta\delta$ values followed normal distributions [μ,σ] (Figure 1c), with μ aligning with the values shown in Figure 1b and σ indicating variability. However, all [μ,σ] data are in Table S21 and Figure 1b highlights CH_2Cl_2, benzene-d_6, DMSO-d_6, and D_2O due to their popularity and stronger differentiation. For ^{13}C, most solvents showed low dispersion near zero, except D_2O, which shifted nearly 2 ppm (σ = 2.39 ppm). For ^{1}H, benzene-d_6 and D_2O showed the highest variability (σ = 0.31−0.35 ppm). In summary, although it is not possible to predict how each individual signal will shift, we can estimate how the signals will shift on average, which will be essential for understanding the potential impact on a structure-elucidation process (*vide infra*).

Next, to address the important question of how current DFT methods are able to predict the experimental SIV, we selected 8 representative molecules with different chemical environments, and their chemical shifts were calculated at the level of theory recommended for DP4+ calculations (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*)[16] across the nine solvents employed. Although explicit-solvent and hybrid approaches (e.g., explicit MM or QM solvent combined with CPCM) are, in principle, more physically appropriate for capturing short-range specific interactions such as hydrogen bonding,[31−36] their use in routine structure-elucidation workflows remains limited by cost and time demands. In a typical workflow, multiple candidate isomers—often dozens—must be modeled, often displaying substantial conformational flexibility that leads to hundreds or thousands of geometries requiring optimization prior to chemical-shift calculations. For this reason, despite the rapid advances in computational hardware and increasingly efficient parallelization techniques, explicit-solvent dynamic simulations, while potentially valuable for providing deeper insight into solute−solvent interactions relevant to NMR predictions, remain impractical for routine applications. Indeed, all widely used NMR-based structure-

https://doi.org/10.1021/acs.jcim.5c02506
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

elucidation methodologies (including DP4,[13−15] DP4+,[16] J-DP4,[17] CASE-3D,[37] DU8+,[38] among others) rely exclusively on implicit solvation models. To remain consistent with current practice in the field, and to better understand the performance of this commonly employed strategy, the present study focuses solely on implicit-solvent approaches,[39,40] aiming to elucidate their behavior under varying solvent conditions.

Strychnine was selected as the initial test molecule due to its rigid three-dimensional structure, which minimizes potential conformational effects while still offering substantial structural complexity. Under all calculated conditions, the results were remarkably consistent, showing only minor variations, with computed MM$\Delta\delta$ values notably smaller than those observed experimentally (MA-MM$\Delta\delta$ 0.63 vs 3.53 ppm for ^{13}C data; and 0.07 vs 0.74 ppm for ^{1}H data), indicating that the computational models failed to reproduce the experimentally observed solvent-dependent shifts. More importantly, as depicted in Figure 2, not only were the calculations unable to predict the magnitude of the changes but they also failed to disclose which nuclei would be most affected by solvent changes. To ensure that strychnine was not an outlier, seven additional structures with varying chemical environments were evaluated in the same way (Table 1; for complete data, see

Table 1. Experimental and Calculated MA-MM$\Delta\delta$ Values of Eight Representative Molecules of the Test Set at the PCM/mPW1PW91/6-31+G//B3LYP/6-31G* Level of Theory**

		calcd MA-MM$\Delta\delta$		exp. MA-MM$\Delta\delta$	
entry	structure	^{13}C	^{1}H	^{13}C	^{1}H
1	strychnine	0.6	0.07	**3.5**	**0.74**
2	digitoxigenin	0.8	0.09	**2.0**	**0.52**
3	dihydroepivulgarin	0.5	0.10	**3.2**	**0.69**
4	*O*-methylglucose	0.3	0.09	**3.1**	**0.90**
5	menthol	0.3	0.05	**2.6**	**0.31**
6	quinidine	0.6	0.05	**4.4**	**0.72**
7	reserpine	0.6	0.08	**2.2**	**0.44**
8	zoanthamine	0.9	0.09	**3.6**	**0.82**
	average	**0.6**	**0.08**	**3.1**	**0.64**

Section 1 worksheet of the supporting Excel file). Averaging all the analyzed cases, the theoretical MA-MM$\Delta\delta$ was 0.6 ppm for ^{13}C and 0.08 ppm for ^{1}H, remarkably smaller than the experimental values of 2.5 and 0.59 ppm, respectively. To generalize this observation, we investigated 11 additional levels of theory of similar computational cost, including the functionals B3LYP, M06-2X, and mPW1PW91 and the 6-31+G**, 6-311G*, 6-311G**, and cc-PVDZ basis sets. The results obtained were practically the same as for the previously tested "DP4+" level in all cases (detailed values are provided in Section 1 worksheet of the supporting Excel file).

The previous study was conducted using B3LYP/6-31G* geometries. While this level is suitable for broad DP4+ calculations, it is known to be less than optimal for other applications.[41] Understanding that small geometric variations can trigger significant changes in the calculated shielding constants,[42] other levels of theory were subsequently evaluated for both the geometry optimization and NMR calculation stages. To conduct the most thorough evaluation possible, we explored 6 functionals (B3LYP, mPW1PW91, ωB97XD, M06-2X, TPSSTPSS, and BMK), 4 basis sets (6-311+G**, 6-311++G(3df,2pd), def2-TZVPD, and pcSseg-2), 3 solvation models (PCM, CPCM, and SMD), and dispersion corrections (Grimmés GD3). The consideration of alternative solvation approaches is also relevant given the well-known limitation of PCM in accurately reproducing the chemical shifts of polyhydroxylated systems, due to its tendency to overestimate the relative stability of conformations featuring intramolecular hydrogen bonds.[43] However, the results obtained were practically the same as for the "DP4+" level in all cases (see Sections 2−4 worksheets of the supporting Excel file).

To avoid an exorbitant number of permutations, each level was used in tandem for the geometry optimization and NMR calculation. Additionally, we selected $CDCl_3$, benzene-d_6, and D_2O as the solvents for study because they showed the greatest experimental differences between them, leading to 360 combinations. To save computational resources, we initially conducted the entire study using strychnine as a model study. As expected, the quality of the results was heavily dependent on the functional used. For instance, using experimental values in $CDCl_3$, the corrected mean absolute errors (CMAE) range from 0.98 to 2.08 ppm for ^{13}C and from 0.08 to 0.15 ppm for ^{1}H. The effect of the functional is pronounced: the best results are obtained with mPW1PW91 and ωB97XD, while the worst results are obtained with M06-2X and M06-2X-GD3. There is no significant change with variations in the basis sets, dispersion, or solvation methods. However, the differences in the chemical shifts calculated for a pair of solvents remained virtually unchanged and were significantly smaller than those observed experimentally. Additionally, there is no correlation between the calculated values and the experimental data, as indicated by the low R^2 values (Figure 3).

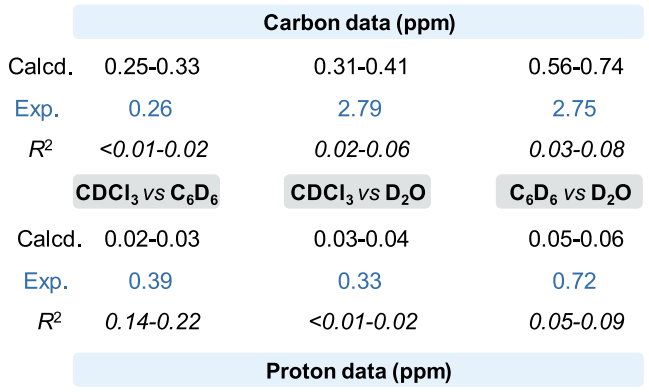


Figure 3. Average differences in the chemical shifts (MA-$\Delta\delta$) of strychnine between pairs of solvents: $CDCl_3$ vs benzene-d_6, $CDCl_3$ vs D_2O, and benzene-d_6 vs D_2O, and R^2 values from the correlation of experimental vs calculated $\Delta\delta$.

In summary, all theoretical calculations tested failed to reproduce the relative experimental changes in chemical shift due to solvent selection ($\Delta\delta$). However, significant differences were observed in terms of the absolute predictive quality. For example, the ^{13}C CMAE obtained by correlating the calculated and experimental chemical shifts in D_2O across all levels of theory are considerably higher than those obtained for $CDCl_3$ and benzene-d_6 (average 3.00 ppm vs 1.63 and 1.66 ppm, respectively). Similarly, the average ^{1}H CMAE collected in benzene-d_6 (0.22 ppm) were larger than those for D_2O (0.17 ppm) and greatly increased than those for $CDCl_3$ (0.08 ppm). This can be understood as the scaling process depends on the correction between the calculated and experimental data for

https://doi.org/10.1021/acs.jcim.5c02506
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

each solvent, which suggests that $CDCl_3$ tends to provide better linearity compared to other solvents.

This finding triggered a further study of the effect of scaling. This step is commonly used to remove systematic errors arising from the DFT calculation process, using the correction formula $\delta_{sc} = (\delta_{un} - b)/m$, where m and b are the slope and intercept, respectively, obtained from a plot of experimental vs calculated data.[6] There are two main methods for performing these correlations, namely, SC1 and SC2. In SC1, a large data set is used to get general $[m,b]$ values, which depend only on the level of theory used in the NMR calculation procedure. In SC2, which is used in DP4 and DP4+, $[m,b]$ are obtained from the direct correlation of the set of experimental and calculated data for each compound.[13,16,17] Hence, the scaling factors depend on both the level of theory as well as the unscaled and experimental values. Both approaches were tested using strychnine as a model compound, with the unscaled NMR chemical shifts computed at the PCM/mPW1PW9/6-311+G-(2d,p)//B3LYP/6-31+G(d,p) level using D_2O, $CDCl_3$, and benzene-d_6. The SC1 was performed using the $[m,b]$ values reported by Tantillo.[44]

For the ^1H and ^{13}C series, the computed unscaled MA-MM$\Delta\delta$ are markedly smaller than the experimental ones, confirming that calculations tend to yield nearly identical results. As expected, scaling narrows the gap, especially with SC2, which forces a match to each molecule's experimental data. Still, both scaling methods perform poorly, showing little correlation with experimental MM$\Delta\delta$ values ($R^2 < 0.2$). A more detailed analysis showed that certain solvents provide better agreement between calculated and experimental shifts upon scaling. As a representative example, in strychnine (**1**) while the solvent effect is underestimated in the case of H-14 (exp 1.28 ppm; SC1 0.43 ppm; SC2 0.81 ppm), it is significantly overestimated in the case of H-11b (exp 0.08 ppm; SC1 0.30 ppm; SC2 0.76 ppm). Similar trends are observed at other levels of theory, molecules, and solvents (see Figure S3).

To address this in more detail, we selected seven additional molecules from the test set and computed the NMR chemical shifts at 12 affordable levels of theory for DP4+ calculations. The unscaled chemical shifts were adjusted using the SC2 method, based on experimental values recorded in the corresponding solvents (Figure 4). For ^{13}C, $CDCl_3$, CD_2Cl_2, and acetone-d_6 performed best, while D_2O, DMSO-d_6, and CD_3OD showed poor alignment. For ^1H, most solvents behaved similarly, except pyridine-d_5, benzene-d_6, and D_2O, which performed notably worse. This may be attributed to the fact that certain solvents, such as D_2O or benzene-d_6, can generate highly specific solvation environments—through hydrogen bonding, $\pi-\pi$ interactions, or strong anisotropic effects—that are not adequately captured by implicit solvation models.[39,40] Therefore, from a purely computational standpoint, certain solvents appear less suitable for achieving reliable agreement between experimental and calculated values. In contrast, CD_3OD and DMSO-d_6 remain practical alternatives for polar compounds with a low solubility in $CDCl_3$.

It is important to remember that the central aim of this work extends beyond evaluating solvent effects on individual NMR chemical shifts. The key objective is to understand how such variations influence the reliability and robustness of probabilistic DP4 approaches in structure elucidation. In practice, these methods do not rely on the precise prediction of each chemical shift but on the collective agreement between

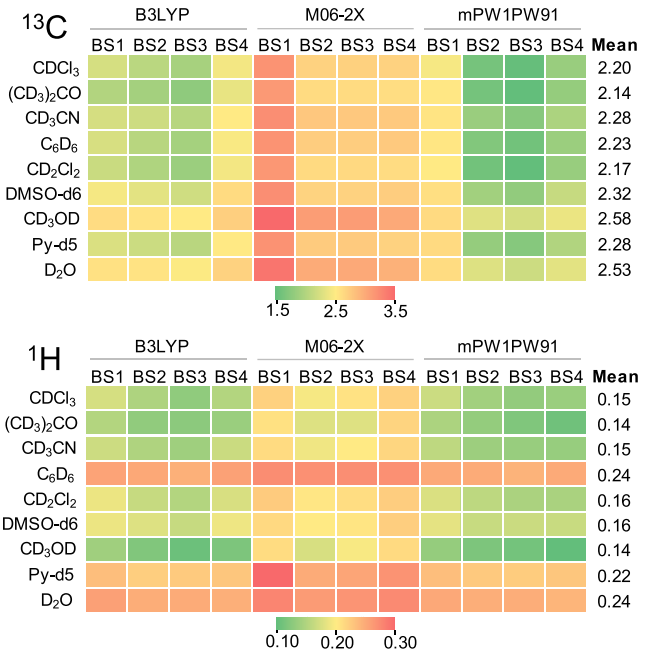


Figure 4. Averaged CMAE values (in ppm) obtained for a representative set of 8 molecules depicted in Table 1 with the nine solvents under study at 12 levels of theory. BS1: cc-PVDZ, BS2: 6-311G*, BS3: 6-311G**, and BS4: 6-31+G**.

calculated and experimental data sets. However, when solvent-induced deviations alter experimental shifts in ways not adequately captured by current DFT models, the resulting probabilities and thus the structural conclusions drawn from them may lose reliability. Recognizing and quantifying this sensitivity is therefore essential to ensure that DP4-like analyses remain dependable tools for accurate structure determination under varying solvent conditions

As DP4+ was parametrized using $CDCl_3$, we evaluated its sensitivity to solvent changes using a subset of seven compounds. Chemical shifts for all diastereomers were calculated and correlated with experimental data in each solvent using DP4+. While most assignments were consistent with those obtained in $CDCl_3$, some showed striking reversals. For instance, quinidine shifted from >99.9% in $CDCl_3$ to just 9.4% in CD_3CN, and epoxone dropped from >99.9% in $CDCl_3$ to 0.2% in D_2O, underscoring the potential significance of solvent effects (see Table S29 for the complete analysis).

Analysis of scaled (*s*DP4+) and unscaled (*u*DP4+) probabilities revealed that while some compounds exhibited strong robustness to solvent changes, others showed significant sensitivity, with at least one probability term deviating markedly. In most cases, the final assignment remained correct despite such shifts; however, in some instances—especially in polar solvents, strong imbalances led to incorrect results.

Although more reliable energies, such as those obtained at the CCSD(T)/CBS level,[45,46] could refine the Boltzmann conformational profiles of flexible systems and potentially enhance the prediction accuracy, the computational cost of such calculations remains prohibitive for routine applications. In practice, achieving an appropriate balance between predictive quality, simplicity, and low computational expense becomes essential.

In this context, anticipating the impact of solvent choice is critical, especially since implicit solvation models in DFT offer

https://doi.org/10.1021/acs.jcim.5c02506
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

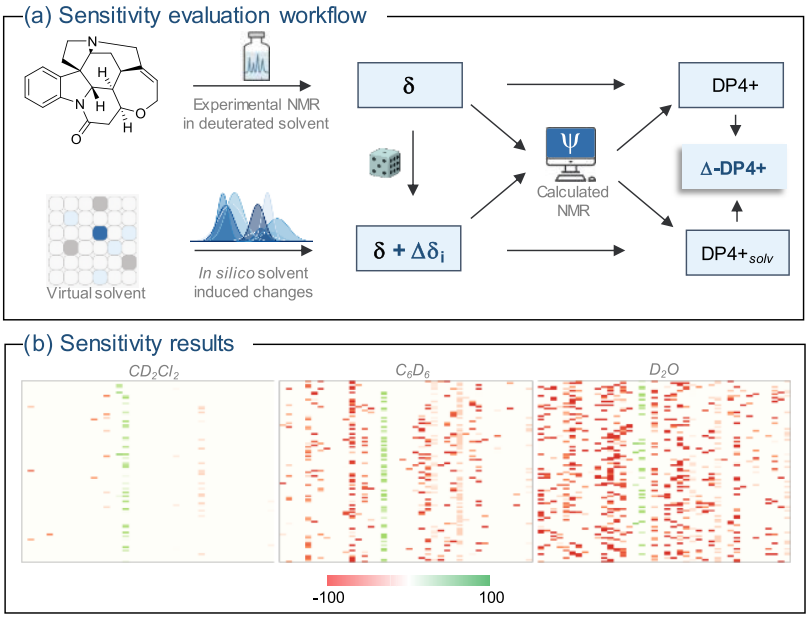


Figure 5. (a) General workflow to calculate the Δ-DP4+ values. (b) Δ-DP4+ values computed for the validation set (40 examples, horizontal axis) with 3 different solvents. To simplify the presentation, only the first 100 iterations are shown (vertical axis).

little predictive power. To address this, we adopted a stochastic approach to assess the molecular sensitivity through thousands of statistically consistent estimations. The workflow is summarized in Figure 3a, starting with a standard DP4+ calculation protocol in a given deuterated solvent. Next, solvent-induced changes in experimental chemical shifts ($\Delta\delta_i$) were simulated by adding Gaussian noise with solvent-pair mean (μ) and standard deviation (σ) derived from experimental data (Figure 1b), yielding data sets consistent with expected solvent-dependent variations. A new DP4+ probability ($DP4+_{solv}$) is recomputed with the modified experimental values. The resulting Δ-DP4+ = $DP4+_{solv}$ − DP4+ quantifies the assignment's sensitivity to solvent effects. While each set is statistically plausible, they are unlikely to replicate real experimental changes exactly. Therefore, the process was repeated 10,00,000 times to capture general trends, and the results are averaged. This approach was validated with an independent set of 40 known molecules, and the results obtained for CD_2Cl_2, benzene-d_6, and D_2O are shown in Figure 5b; full data are available in Figure S12.

The averaged Δ-DP4+ values are negative across all solvents, indicating that $CDCl_3$ tends to provide the most reliable assignments. Minimal differences are seen with CD_2Cl_2, acetone-d_6, and CD_3CN (Δ-DP4+ ≈ 2.3−4.0%), but larger deviations arise with benzene-d_6, pyridine-d_5, and especially D_2O (10.4−17.1%). Proton data are more affected (avg. Δ-^{1}H-DP4+ = 5.9%) than carbon (avg. Δ-^{13}C-DP4+ = 1.2%), suggesting a stronger solvent effect on ^{1}H-based assignments. Solvents more distinct from $CDCl_3$ yield higher Δ-DP4+ values, particularly for proton shifts—as seen with benzene-d_6 (Δ-^{1}H = 11.3% vs Δ-^{13}C = 0.6%). Unscaled data are more sensitive to solvent changes than scaled data (7.7 vs 3.2%).

Importantly, variations in the magnitude of DP4+ do not necessarily alter the direction of the assignment. In 91% of cases, the classification outcome was preserved across solvents, highlighting the robustness of DP4+ and its reliability. However, in the remaining 9% of cases, the most probable isomer differed between $CDCl_3$ and the alternative solvent. Statistically, solvent changes cannot be predicted to improve or impair the assignment reliability. The main issue is the unpredictable effect on specific signals. Nonetheless, D_2O, benzene-d_6, and pyridine-d_5 consistently show behavior that makes them less suitable for standard DP4+ protocols. This is further confirmed by an inverse DP4+ analysis, where calculated shifts were compared to experimental data from nine solvents. $CDCl_3$ consistently produced the highest probability, reinforcing its status as the preferred solvent for DP4+ (see the SI). Despite this, practical considerations such as solubility, availability, or user preference may lead to alternative solvent choices, introducing uncertainty. To address this, we developed a Python-based tool that evaluates a molecule's sensitivity to solvent changes and offers valuable insight into how the solvent choice may influence structural assignments. Released under the MIT open-source license, the program is accessible at both GitHub (https://github.com/Sarotti-Lab/DP4plus-solv) and the Python Package Index (https://pypi.org/project/DP4plus-solv/), accompanied by comprehensive tutorials to facilitate widespread adoption.

In summary, although current simulations struggle to accurately predict solvent-induced chemical-shift changes, variations relative to $CDCl_3$ have a smaller impact on DP4-based, computer-assisted structure determination than anticipated. This outcome reinforces the value of the QM-NMR methods. Successful applications using solvents such as CD_2Cl_2, acetone-d_6, CD_3CN, or CD_3OD further confirm their reliability as alternatives.[47−49] In contrast, we advise against aromatic solvents (e.g., benzene-d_6, pyridine-d_5) and D_2O as they show significant deviations from $CDCl_3$-based parameters. While these changes could in principle be better captured by a MD+DFT approach, potentially providing improved treatment of π-stacking or H-bonding interactions,[50] from a computational cost and operational simplicity perspective this does not seem feasible for highly flexible systems with many isomers. Hence, if solubility mandates solvent change, DMSO-d_6 is preferable, yielding more consistent computational NMR predictions.

https://doi.org/10.1021/acs.jcim.5c02506
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Our results recast solvent effects from a nuisance to a design parameter. We intend this study to serve as a cornerstone for solvent-aware simulations, enabling general strategies to model and correct solvent-induced chemical-shift changes. Embedding solvent-sensitivity metrics into DFT protocols and ML pipelines should yield reliable predictions and bring us closer to the field's long-sought goal: fully automated structure elucidation from nothing more than a table of ^{1}H and ^{13}C chemical shifts.

■ ASSOCIATED CONTENT

Data Availability Statement

The data underlying this study are available in the published article and its Supporting Information.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.5c02506.

All experimental data associated with this study (PDF)

All computational data associated with this study (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Antonio Hernández Daranas − *Instituto de Productos Naturales y Agrobiología. Consejo Superior de Investigaciones Cientificas (IPNA-CSIC), San Cristobal de La Laguna 38206, Spain;* orcid.org/0000-0001-8376-7941; Email: sarotti@iquir-conicet.gov.ar

Ariel M. Sarotti − *Instituto de Química Rosario (IQUIR, CONICET-UNR) - Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario S2002LRK, República Argentina;* orcid.org/0000-0002-8151-0306; Email: adaranas@ipna.csic.es

Authors

Iván Cortés − *Instituto de Química Rosario (IQUIR, CONICET-UNR) - Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario S2002LRK, República Argentina*

Cristina Cuadrado − *Instituto de Productos Naturales y Agrobiología. Consejo Superior de Investigaciones Cientificas (IPNA-CSIC), San Cristobal de La Laguna 38206, Spain;* orcid.org/0000-0001-9282-1575

José A. Gavín − *Department of Organic Chemistry, Universidad de La Laguna, San Cristobal de La Laguna 38206, Spain*

María Marta Zanardi − *Instituto de Investigaciones en Ingeniería Ambiental, Química y Biotecnología Aplicada (INGEBIO), Facultad de Química e Ingeniería del Rosario, Pontificia Universidad Católica Argentina, Rosario S2002QEO, Argentina*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.5c02506

Author Contributions

$^{\perp}$I.C. and C.C. contributed equally to this work. The manuscript was written through contributions of all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

I.C. thanks CONICET for a postdoctoral fellowship, and C.C. thanks ACIISI and FSE (Programa Operativo Integrado de Canarias 2014-2020, Eje 3, Tema Prioritario 74−85%) for a predoctoral fellowship. This research was funded by the UNR (BIO 500 and 567), ANPCyT (PICT-2019-4052 and PICT-2021-CAT-II-00033), CONICET (PIP 11220200102205CO and PIBAA 28720210101005CO), MICINN (PID 2019-109476RB-C21), and ACIISI-FEDER (ProID2021010118).

■ DEDICATION

Dedicated to the memory of Dr. Raquel María Dorta Rodríguez.

■ REFERENCES

(1) Williams, A.; Martin, G.; Rovnyak, D. Modern NMR Approaches to the Structure Elucidation of Natural Products. In *Volume 2: Data Acquisition and Applications to Compound Classes*; Royal Society of Chemistry, 2016.

(2) Nath, N.; Schmidt, M.; Gil, R. R.; Williamson, R. T.; Martin, G. E.; Navarro-Vázquez, A.; Griesinger, C.; Liu, Y. Determination of Relative Configuration from Residual Chemical Shift Anisotropy. *J. Am. Chem. Soc.* **2016**, *138* (30), 9548−9556.

(3) Pauli, G. F.; Niemitz, M.; Bisson, J.; Lodewyk, M. W.; Soldi, C.; Shaw, J. T.; Tantillo, D. J.; Saya, J. M.; Vos, K.; Kleinnijenhuis, R. A.; Hiemstra, H.; Chen, S. N.; McAlpine, J. B.; Lankin, D. C.; Friesen, J. B. Toward Structural Correctness: Aquatolide and the Importance of 1D Proton NMR FID Archiving. *J. Org. Chem.* **2016**, *81* (3), 878−889.

(4) Nicolaou, K. C.; Snyder, S. A. Chasing Molecules That Were Never There: Misassigned Natural Products and the Role of Chemical Synthesis in Modern Structure Elucidation. *Angew. Chem., Int. Ed.* **2005**, *44* (7), 1012−1044.

(5) Chhetri, B. K.; Lavoie, S.; Sweeney-Jones, A. M.; Kubanek, J. Recent Trends in the Structural Revision of Natural Products. *Nat. Prod. Rep.* **2018**, *35* (6), 514−531.

(6) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational Prediction of 1H and 13C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **2012**, *112* (3), 1839−1862.

(7) Marcarino, M. O.; Zanardi, M. M.; Cicetti, S.; Sarotti, A. M. NMR Calculations with Quantum Methods: Development of New Tools for Structural Elucidation and Beyond. *Acc. Chem. Res.* **2020**, *53* (9), 1922−1932.

(8) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. Determination of Relative Configuration in Organic Compounds by NMR Spectroscopy and Computational Methods. *Chem. Rev.* **2007**, *107* (9), 3744−3779.

(9) Gerrard, W.; Bratholm, L. A.; Packer, M. J.; Mulholland, A. J.; Glowacki, D. R.; Butts, C. P. Impression-Prediction of NMR Parameters for 3-Dimensional Chemical Structures Using Machine Learning with near Quantum Chemical Accuracy. *Chem. Sci.* **2020**, *11* (2), 508−515.

(10) Guan, Y.; Shree Sowndarya, S. V.; Gallegos, L. C.; St John, P. C.; Paton, R. S. Real-Time Prediction of 1H and 13C Chemical Shifts with DFT Accuracy Using a 3D Graph Neural Network. *Chem. Sci.* **2021**, *12* (36), 12012−12026.

(11) Cortes, I.; Cuadrado, C.; Hernandez Daranas, A.; Sarotti, A. M. Machine Learning in Computational NMR-aided structural elucidation. *Front. Nat. Prod.* **2023**, *2*, No. 1122426.

(12) Kleine Büning, J. B.; Grimme, S. Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT. *J. Chem. Theory Comput.* **2023**, *19* (12), 3601−3615.

(13) Smith, S. G.; Goodman, J. M. Assigning Stereochemistry to Single Diastereoisomers by GIAO NMR Calculation: The DP4 Probability. *J. Am. Chem. Soc.* **2010**, *132* (37), 12946−12959.

(14) Ermanis, K.; Parkes, K. E. B. B.; Agback, T.; Goodman, J. M. Doubling the Power of DP4 for Computational Structure Elucidation. *Org. Biomol. Chem.* **2017**, *15* (42), 8998−9007.
(15) Ermanis, K.; Parkes, K. E. B.; Agback, T.; Goodman, J. M. The Optimal DFT Approach in DP4 NMR Structure Analysis-Pushing the Limits of Relative Configuration Elucidation. *Org. Biomol. Chem.* **2019**, *17* (24), 5886−5890.
(16) Grimblat, N.; Zanardi, M. M.; Sarotti, A. M. Beyond DP4: An Improved Probability for the Stereochemical Assignment of Isomeric Compounds Using Quantum Chemical Calculations of NMR Shifts. *J. Org. Chem.* **2015**, *80* (24), 12526−12534.
(17) Grimblat, N.; Gavín, J. A.; Hernández Daranas, A.; Sarotti, A. M. Combining the Power of J Coupling and DP4 Analysis on Stereochemical Assignments: The J-DP4Methods. *Org. Lett.* **2019**, *21* (11), 4003−4007.
(18) Li, S. W.; Cuadrado, C.; Yao, L. G.; Daranas, A. H.; Guo, Y. W. Quantum Mechanical-NMR-Aided Configuration and Conformation of Two Unreported Macrocycles Isolated from the Soft Coral Lobophytum Sp.: Energy Calculations versus Coupling Constants. *Org. Lett.* **2020**, *22* (11), 4093−4096.
(19) Cuadrado, C.; Cen-Pacheco, F.; Daranas, A. H. Computationally Assisted Analysis of NMR Chemical Shifts as a Tool in Conformational Analysis. *Org. Lett.* **2024**, *26* (31), 6529−6534.
(20) Dračínský, M.; Bour, P. Computational Analysis of Solvent Effects in NMR Spectroscopy. *J. Chem. Theory Comput.* **2010**, *6* (1), 288−299.
(21) Ronayne, J.; Williams, D. H. Solvent Effects in Proton Magnetic Resonance Spectroscopy. *Annu. Rep. NMR Spectrosc.* **1969**, *2*, 83−124, DOI: 10.1016/S0066-4103(08)60320-8.
(22) Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I. NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **2010**, *29* (9), 2176−2179.
(23) Schaefer, T.; Schneider, W. G. On the Nature of Solvent Effects in the Proton Resonance Spectra of Unsaturated Ring Compounds. I. Substituted Benzenes. *J. Chem. Phys.* **1960**, *32* (4), 1218−1222.
(24) Paudler, W. W.; Humphrey, S. A. Solvent Effects in Nuclear Magnetic Resonance Spectra of Some Pyrazines, Pyrimidines and Their N-oxides. *Org. Magn. Reson.* **1971**, *3* (2), 217−220.
(25) Matsuo, T. Studies of the Solvent Effect on the Chemical Shifts in Nmr Spectroscopy. II. Solutions of Succinic Anhydride, Maleic Anhydride, and the N-Substituted Imides. *Can. J. Chem.* **1967**, *45* (16), 1829−1835.
(26) Buckingham, A. D.; Schaefer, T.; Schneider, W. G. Solvent Effects in Nuclear Magnetic Resonance Spectra. *J. Chem. Phys.* **1960**, *32* (4), 1227−1233.
(27) Babij, N. R.; McCusker, E. O.; Whiteker, G. T.; Canturk, B.; Choy, N.; Creemer, L. C.; De Amicis, C. V.; Hewlett, N. M.; Johnson, P. L.; Knobelsdorf, J. A.; et al. NMR Chemical Shifts of Trace Impurities: Industrially Preferred Solvents Used in Process and Green Chemistry. *Org. Process Res. Dev.* **2016**, *20* (3), 661−667.
(28) Domínguez, H.; Crespín, G. D.; Santiago-Beńitez, A. J.; Gaǘin, J. A.; Norte, M.; Fernández, J. J.; Daranas, A. H. Stereochemistry of Complex Marine Natural Products by Quantum Mechanical Calculations of NMR Chemical Shifts: Solvent and Conformational Effects on Okadaic Acid. *Mar. Drugs* **2014**, *12* (1), 176−192.
(29) Pierens, G. K.; Venkatachalam, T. K.; Reutens, D. C. NMR and DFT Investigations of Structure of Colchicine in Various Solvents Including Density Functional Theory Calculations. *Sci. Rep.* **2017**, *7* (1), No. 5605.
(30) Katritzky, A. R.; Fara, D. C.; Yang, H.; Tämm, K.; Tamm, T.; Karelson, M. Quantitative Measures of Solvent Polarity. *Chem. Rev.* **2004**, *104* (1), 175−198.
(31) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55* (22), 2471.
(32) Iftimie, R.; Minary, P.; Tuckerman, M. E. Ab Initio Molecular Dynamics: Concepts, Recent Developments, and Future Trends. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6654−6659.
(33) Bagno, A.; Rastrelli, F.; Saielli, G.; Chimiche, S.; Marzolo, V.; Padova, S. Prediction of the 1 H and 13 C NMR Spectra of R- D -Glucose in Water by DFT Methods and MD Simulations. *J. Org. Chem.* **2007**, *72*, 7373−7381.
(34) Palivec, V.; Pohl, R.; Kaminsky, J.; Martinez-Seara, H. Efficiently Computing NMR 1H and 13C Chemical Shifts of Saccharides in Aqueous Environment. *J. Chem. Theory Comput.* **2022**, *18* (7), 4373−4386.
(35) Toukach, F. V.; Ananikov, V. P. 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.
(36) Watson, A.; Simmermaker, C.; Aung, E.; Do, S.; Hackbusch, S.; Franz, A. H. NMR Analysis and Molecular Dynamics Conformation of α-1, 6-Linear and α-1, 3-Branched Isomaltose Oligomers as Mimetics of α-1, 6-Linked Dextran. *Carbohydr. Res.* **2021**, *503*, No. 108296.
(37) Navarro-Vázquez, A.; Gil, R. R.; Blinov, K. Computer-Assisted 3D Structure Elucidation (CASE-3D) of Natural Products Combining Isotropic and Anisotropic NMR Parameters. *J. Nat. Prod.* **2018**, *81* (1), 203−210.
(38) Kutateladze, A. G.; Reddy, D. S. High-Throughput in Silico Structure Validation and Revision of Halogenated Natural Products Is Enabled by Parametric Corrections to DFT-Computed 13C NMR Chemical Shifts and Spin-Spin Coupling Constants. *J. Org. Chem.* **2017**, *82* (7), 3368−3381.
(39) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378−6396.
(40) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105* (8), 2999−3094.
(41) Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry**. *Angew. Chem., Int. Ed.* **2022**, *61* (42), No. e202205735.
(42) Cuadrado, C.; Daranas, A. H.; Sarotti, A. M. May the Force (Field) Be with You: On the Importance of Conformational Searches in the Prediction of NMR Chemical Shifts. *Mar. Drugs* **2022**, *20* (11), No. 699.
(43) Zanardi, M. M.; Marcarino, M. O.; Sarotti, A. M. Redefining the Impact of Boltzmann Analysis in the Stereochemical Assignment of Polar and Flexible Molecules by NMR Calculations. *Org. Lett.* **2020**, *22* (1), 52−56.
(44) Merrill, A. T.; Tantillo, D. J. Solvent Optimization and Conformational Flexibility Effects on 1H and 13C NMR Scaling Factors. *Magn. Reson. Chem.* **2020**, *58* (6), 576−583.
(45) Marianski, M.; Supady, A.; Ingram, T.; Schneider, M.; Baldauf, C. Assessing the Accuracy of Across-the-Scale Methods for Predicting Carbohydrate Conformational Energies for the Examples of Glucose and α-Maltose. *J. Chem. Theory Comput.* **2016**, *12* (12), 6157−6168.
(46) Csonka, G. I.; Kaminsky, J. Accurate Conformational Energy Differences of Carbohydrates: A Complete Basis Set Extrapolation. *J. Chem. Theory Comput.* **2011**, *7* (4), 988−997.
(47) Paterson, I.; Dalby, S. M.; Roberts, J. C.; Naylor, G. J.; Guzmán, E. A.; Isbrucker, R.; Pitts, T. P.; Linley, P.; Divlianska, D.; Reed, J. K.; Wright, A. E. Leiodermatolide, a Potent Antimitotic Macrolide from the Marine Sponge Leiodermatium Sp. *Angew. Chem.* **2011**, *123* (14), 3277−3281.
(48) Domínguez, H. J.; Napolitano, J. G.; Fernández-Sánchez, M. T.; Cabrera-García, D.; Novelli, A.; Norte, M.; Fernández, J. J.; Daranas, A. H. Belizentrin, a Highly Bioactive Macrocycle from the Dinoflagellate Prorocentrum Belizeanum. *Org. Lett.* **2014**, *16* (17), 4546−4549.

https://doi.org/10.1021/acs.jcim.5c02506
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(49) Anderl, F.; Größl, S.; Wirtz, C.; Fürstner, A. Total Synthesis of Belizentrin Methyl Ester: Report on a Likely Conquest. *Angew. Chem., Int. Ed.* **2018**, *57* (33), 10712−10717.
(50) Saielli, G. Computational NMR Spectroscopy of Ionic Liquids. In *NMR Spectroscopic Parameters: Theories and Models, Computational Codes and Calculations*; Aucar, G. A., Ed.; Royal Society of Chemistry, 2025; Vol. *38*, pp 395−429 DOI: 10.1039/9781837678020-00395.

https://doi.org/10.1021/acs.jcim.5c02506
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX