JOC The Journal of Organic Chemistry

Article

pubs.acs.org/joc

Beyond DP4: an Improved Probability for the Stereochemical Assignment of Isomeric Compounds using Quantum Chemical Calculations of NMR Shifts

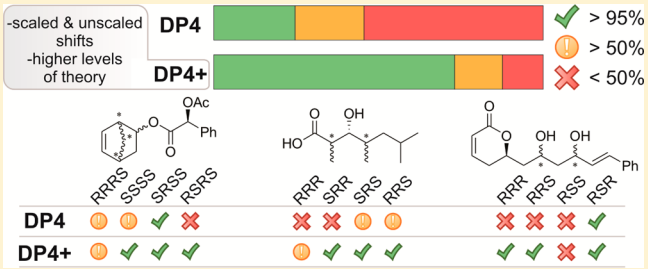
Nicolás Grimblat,[†] María M. Zanardi,[†,‡] and Ariel M. Sarotti*[,†]

[†]Instituto de Química Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario 2000, Argentina

[‡]Facultad de Química e Ingeniería "Fray Rogelio Bacón", Pontificia Universidad Católica Argentina, Av. Pellegrini 3314, Rosario 2000, Argentina

Supporting Information

ABSTRACT: The DP4 probability is one of the most sophisticated and popular approaches for the stereochemical assignment of organic molecules using GIAO NMR chemical shift calculations when only one set of experimental data is available. In order to improve the performance of the method, we have developed a modified probability (DP4+), whose main differences from the original DP4 are the inclusion of unscaled data and the use of higher levels of theory for the NMR calculation procedure. With these modifications, a significant improvement in the overall performance was achieved, providing accurate and confident results in establishing the stereochemistry of 48 challenging isomeric compounds.



INTRODUCTION

The total synthesis of natural products is one of the most beautiful and exciting chemistry areas, lying somewhere between a fine art and a hard science.[1] Apart from mimicking nature's ability to build complex molecular architectures, the development of innumerable synthetic strategies, methodologies, and new chemical transformations have been made possible by total synthesis. The enterprises are often hard and fraught with difficulties and detours.[1] Eventually, after a considerable investment of time, money, and man-power the synthetic target is accomplished. However, surprisingly often such an exciting moment quickly moves to frustration once a mismatch between the NMR data of the synthesized compound and the natural product is detected.[2]

Incorrectly assigned natural products are not uncommon,[2] even in the golden age of NMR.[3] High molecular complexity, human errors, signal ambiguity and sample impurities can be pointed to as the most common sources of misassignments.[2] Hundreds of structural revisions have been published in the last decades, ranging from profound connectivity to subtle (but not least) stereochemical errors.[2] Considering that the discrepancies are often detected after total synthesis of the originally proposed (wrong) structure, it is not unreasonable to assume that the real molecular architecture of many reported natural products remains unknown.

Modern computational chemistry has significantly contributed to prevent these misinterpretations. Recent years have witnessed an increase in the use of quantum chemistry approaches in solving structural validation problems,[4] facilitated by the capability of most computational chemistry software packages to compute NMR parameters in a user-friendly environment.[5] After seminal contributions of Bagno[6] and Bifulco,[7] numerous reports have tackled the successful application of NMR calculations in the assignment or reassignment of complex molecular structures.[8]

In principle, there are two main strategies concerning the use of quantum chemical calculations of NMR shifts in structural elucidation. On one side, the correctness of a given putative structure (only one) is assessed with the only information provided by the experimental NMR recorded for that compound and the chemical shifts computed for the structural proposal. We have recently proved that pattern recognition analysis via artificial neural networks resulted in a promising method in this field.[4] On the other hand, in a conceptually different approach, two or more candidates are evaluated and correlated with at least one set of experimental shifts following a comparison-based methodology. In this regard, Smith and Goodman have made a major breakthrough by first introducing the CP3 parameter and later the DP4 probability.[9,10] The CP3 parameter was designed to assign two sets of experimental data (a common situation found in stereoselective reactions) to two possible structures by comparing the differences in calculated shifts between the two isomers with the corresponding differences in the experimental shifts of both.[9] A much more complex situation arises when only one set of experimental data

Received: October 15, 2015
Published: November 18, 2015

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

is available, where the CP3 parameter cannot be computed. This is often the case of natural products, but also of organic reactions with perfect levels of stereoselectivity. To determine the correct structure among many plausible isomers, the DP4 probability was introduced as a powerful tool.[10] The level of correct assignment of DP4 has been shown to be significantly better than those computed based on other statistical parameters (correlation coefficient, MAE, CMAE, etc.). In the recent past, DP4 has been used extensively to confirm or correct the structural identification of several complex molecules,[11] emerging as one of the best methods to tackle this important and difficult task. In addition, it can be also used coupled with other experimental techniques, such as residual dipolar coupling (RDC) by Gil and co-workers.[11a,h] Nevertheless, considering the challenge involved in correlating computational data of closely related isomeric compounds with only one set of experimental values, DP4 shows a modest performance in many cases by pointing the incorrect isomer or affording inconsistent and unreliable results (vide infra).[10,11b,i,j] In this sense, DP4 guidance can be decisive to determine the relative configuration of a complex natural product, encouraging its publication or, even more important, its total synthesis. Based on this fact, we have been encouraged to build an improved probability upon DP4 foundations.

■ RESULTS AND DISCUSSION

Smith and Goodman showed that the errors e between experimental, δ_{exp}, and calculated (scaled) chemical shifts, δ_s, ($e = \delta_s - \delta_{exp}$) for a set of organic molecules obeys a t distribution with mean $\mu = 0$ (as consequence of the linear scaling procedure), standard deviation σ, and degrees of freedom ν. Then, for a given molecule with N nuclei, the probability of each ith error can be computed. Assuming that the error set is an independent random variable, the multiplication of the individual ith probabilities gives the total probability of that candidate structure. Next, a set of percentage probabilities that each candidate is the correct isomer are obtained using Bayes's theorem. Mathematically, the DP4 probability was defined as follows (eq 1):[10]

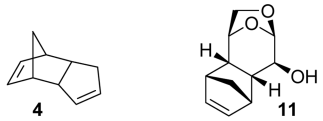
Original DP4

$$P(i) = \frac{\prod_{k=1}^{N}(1-T^{\nu}(\delta^{i}_{s,k} - \delta_{exp,k})/\sigma)}{\sum_{j=1}^{m}[\prod_{k=1}^{N}(1-T^{\nu}(\delta^{i}_{s,k} - \delta_{exp,k})/\sigma)]} \quad (1)$$

T^{ν}, σ and δ^{i}_{s} computed from MMFF geometries

where $P(i)$ is the probability that structure i (from m plausible candidates) is the correct one. T^{ν} gives the cumulative t distribution function with ν degrees of freedom and variance σ. $\delta_{exp,k}$ is the experimental chemical shift for nucleus k (running over N) and $\delta_{s,k}$ represents the calculated shift for nucleus k (running over N) after the scaling procedure to remove systematic errors. This is done according to $\delta_s = (\delta_{calc} - b)/m$, where b and m are the intercept and slope of a plot of δ_{calc} against δ_{exp}.[10] From eq 1, σ and ν are the key terms in the calculation of the DP4 probability and must be determined by fitting the data (errors between scaled and experimental shifts) of a large data set to a t distribution using specific statistical programs. In particular, Smith and Goodman computed 1717 ^{13}C shifts and 1794 ^{1}H shifts from 117 known organic molecules, using GIAO NMR calculations at the B3LYP/6-31G**//MMFF level of theory (gas phase).[10]

After careful analysis of the method, we identified two potential drawbacks in the DP4 architecture: the level of theory and the exclusive use of scaled shifts. Regarding the first issue, the B3LYP/6-31G**//MMFF level was used in the original DP4 formulation for providing good results in the NMR shift calculation at low computational cost (mainly avoiding expensive *ab initio* or DFT treatments for the geometry optimization step). However, from our experience it is rather far from being the most accurate method for NMR calculations, mainly in the prediction of ^{1}H shifts. Despite the fact that a full account for this observation is beyond the scope of this article, two representative examples are given in Figure 1. In these



level of theory	CMAE ^{1}H (4)	CMAE ^{13}C (4)	CMAE ^{1}H (11)	CMAE ^{13}C (11)
B3LYP/6-31G** //MMFF	0.12	1.5	0.21	1.9
B3LYP/6-31G** //B3LYP/6-31G*	0.08	0.9	0.12	1.2
PCM/mPW1PW91/6-31+G** //B3LYP/6-31G*	0.06	0.6	0.09	1.1

Figure 1. Corrected mean average errors (CMAE) for compounds **4** and **11** computed at different levels of theory.

cases, the accuracy of the NMR shift calculations in both conformationally rigid molecules significantly increase when passing from MMFF to B3LYP/6-31G* geometries, as well as computing the shielding tensors at higher levels (for example, PCM/mPW1PW91/6-31+G**). This is of vital importance considering the growing support to the claim that proton data makes the most decisive contribution (i.e., are more effective discriminators than ^{13}C data) in stereochemical assignment.[12] The sometimes modest performance of the level of theory employed by Smith and Goodman for the DP4 development can be attributed to the geometry optimization step (MMFF) rather than the method to perform the NMR calculations, as also suggested by other authors.[11i,j] It is well-known that even small errors in the starting geometry can lead to significant errors in the computed chemical shifts.[4b,5] Thus, we considered that computing the NMR shifts at higher levels of theory would afford more accurate predictions and, therefore, more reliable DP4 probabilities.[13]

On the other hand, the exclusive use of scaled shifts to compute the DP4 probability can be also challenged. Linear scaling is typically employed to remove systematic errors, leading to corrected shifts that are closer to the experimental values.[5−10] However, this practice assumes that the magnitude of an error is independent from the chemical environment (for example, the ^{13}C hybridization), which in general it does not. Moreover, there is always a risk of false positives when correlating data of isomers with similar computed chemical shifts, because one of the incorrect candidates might afford a (fortuitously) better fitting than the correct isomer (*vide infra*).[14] Thus, we thought that adding unscaled ("pure") shift values would emphasize the environmental contrast between the plausible structures.

On the grounds discussed above, we considered that using more accurate levels of theory for the NMR calculation procedure and including unscaled shifts might lead to an

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

improved DP4-like probability. The basic formulation of the new DP4+ probability to assign one set of experimental data to one of many different structures is given in eq 2. The

more accurate NMR shifts ← **DP4+** → scaled & unscaled shifts

$$P(i) = f\,[P(i)_s;\ P(i)_u] \qquad (2)$$

$P(i)_s = f\,[\mu_s, \sigma_s, \nu_s]$ from scaled errors
$P(i)_u = f\,[\mu_u, \sigma_u, \nu_u]$ from unscaled errors

probability for which candidate i (out of m isomers) represents the correct structure, $P(i)$, is given as a function of the corresponding probabilities computed using scaled and unscaled shifts, $P(i)_s$ and $P(i)_u$, respectively, that in turn can be computed using the standard DP4 formalism (see eq 1).

To build the DP4+ probability, the terms T^{ν}_s, σ_s, T^{ν}_u, σ_u, and μ_u (that in principle depend on the level of theory employed in the NMR calculation procedure) must be computed for both ^{13}C and ^{1}H data (note that μ_s is zero as a consequence of the scaling procedure). To accomplish this task, we selected a big data set comprised of 72 small-to-medium sized organic compounds (Figure 2 shows some representative examples) with a wide variety of functional groups and molecular complexity for which the ^{1}H and ^{13}C spectra are known and fully assigned (that is, having all resonances assigned to the corresponding nuclei in the structure).[15] The chemical shifts from B3LYP/6-31G* geometries were next computed at 24 different levels of theory, combining 2 functionals (B3LYP and mPW1PW91) and 6 basis sets (6-31G*, 6-31G**, 6-31+G**, 6-311G*, 6-311G**, and 6-311+G**) for the GIAO single-point NMR calculations, that were also computed in gas phase and in solution (PCM, solvent: chloroform).[16] In the case of conformationally flexible molecules, all conformations within 2 kcal/mol from the B3LYP/6-31G* global minima were taken into consideration for further NMR analysis, and the contribution of each conformer was weighted using Boltzmann averaging.

Once the calculations were done, 1219 and 1123 individual ^{13}C and ^{1}H errors, respectively, were computed at each level of theory (both scaled and unscaled) by subtracting the GIAO scaled and unscaled NMR shifts to the experimental values. With these sets of errors in hand, we next evaluated whether they obey a t distribution, a primary requirement for the DP4+ probability. Despite the fact that this was the case for the scaled errors (Figure 3A), we found that the unscaled errors did not follow a Student's distribution. In contrast, the corresponding histograms seemed to be formed by overlapping of two normally distributed series (Figure 3B). Considering that the performance of TMS as reference standard depends mainly on the hybridization of the nuclei in question (the origins of the multistandard approach),[17] we speculated that the series could be derived from the errors of sp^2 and sp^3 carbons (or protons attached to sp^2 and sp^3 hybridized carbons). In fact, after separating the data we were glad to find that each sp^2- and sp^3-derived series smoothly fitted into two t distributions (Figure 3B). It is important to point out that this behavior was noted both for ^{1}H and ^{13}C at the 24 levels of theory under study.

This finding allowed us to postulate the mathematical formulation of our new DP4+ probability, depicted in eq 3.

DP4+

$$P(i) = \frac{\prod_{k=1}^{N}[1-T^{\nu}_s(e^i_{s,k}/\sigma_s)][1-T^{\nu}_{u\text{-}spx}((e^i_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{\nu}_s(e^j_{s,k}/\sigma_s)][1-T^{\nu}_{u\text{-}spx}((e^j_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]} \qquad (3)$$

T^{ν}, σ, μ, and e^i computed from B3LYP/6-31G* geometries

sDP4+

$$\frac{\prod_{k=1}^{N}[1-T^{\nu}_s(e^i_{s,k}/\sigma_s)]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{\nu}_s(e^j_{s,k}/\sigma_s)]}$$

uDP4+

$$\frac{\prod_{k=1}^{N}[1-T^{\nu}_{u\text{-}spx}((e^i_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{\nu}_{u\text{-}spx}((e^j_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}$$

Under the assumption that the putative structure i is correct, the probability to obtain a given set of scaled ($e_s = \delta_s - \delta_{exp}$) and unscaled ($e_u = \delta_u - \delta_{exp}$) errors is given by the multiplication of each independent probability $[1 - T^{\nu}(e - \mu)/\sigma)]$ term, computed for every scaled and unscaled chemical shift (numerator of eq 3). Then, and assuming that the correct structure is among the m candidates, the probability that i is the correct isomer, $P(i)$, is obtained by dividing by the sum of the probabilities of all m candidates (denominator of eq 3). A simple glance of eq 3 reveals that the DP4+ probability can be

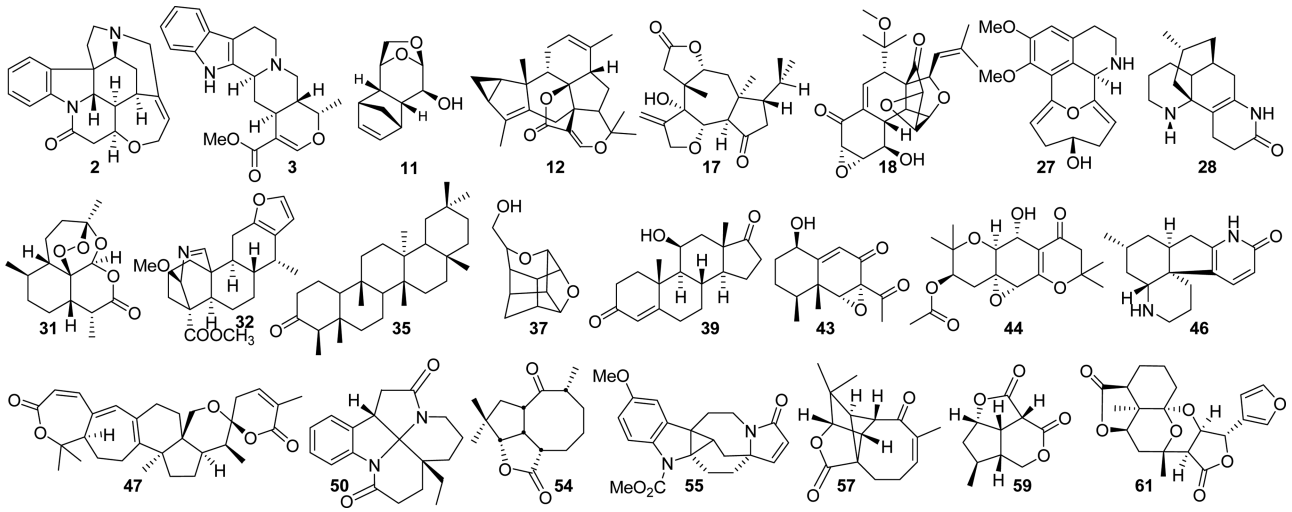


Figure 2. Selected representative examples of the compounds used to compute the DP4+ statistical parameters. For the full test set, see the Supporting Information.

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

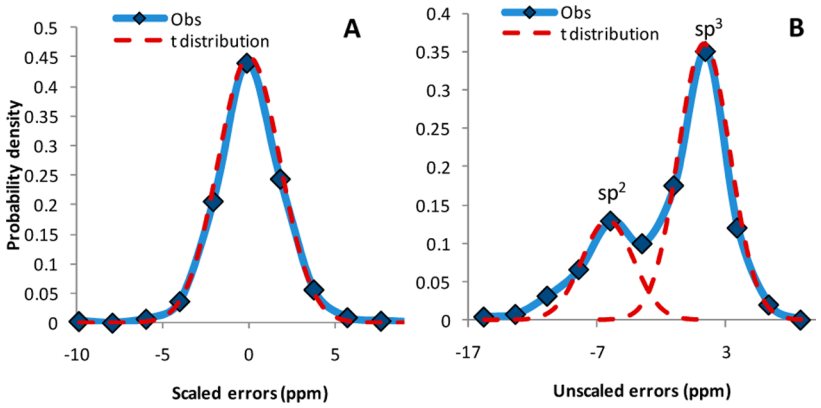


Figure 3. Error distribution plot of the scaled (A) and unscaled (B) ^{13}C chemical shifts computed at the B3LYP/6-31G*//B3LYP/6-31G* level of theory.

decomposed into two main contributions: *s*DP4+ and *u*DP4+. The first term is the probability obtained when using exclusively scaled shifts (as in the case of DP4), whereas the second affords the probability when only unscaled data is used. Note that in this case, the $\mu_{u\text{-}spx}$, $\sigma_{u\text{-}spx}$, and $\nu_{u\text{-}spx}$ values depend upon the hybridization of the nuclei. Then, $T^{\nu}_{u\text{-}sp2}$, $\mu_{u\text{-}sp2}$, and $\sigma_{u\text{-}sp2}$ are the cumulative *t* function with ν degrees of freedom centered on μ and variance σ corresponding of the unscaled sp^2 carbons (or hydrogens attached to sp^2 carbons), and $T^{\nu}_{u\text{-}sp3}$, $\mu_{u\text{-}sp3}$, and $\sigma_{u\text{-}sp3}$ correspond to the analogous parameters of the distribution corresponding to sp^3 nuclei. Therefore, in order to build our DP4+ probability, 16 parameters must be defined at each level of theory: ν_s, σ_s, $\nu_{u\text{-}sp2}$, $\mu_{u\text{-}sp2}$, $\sigma_{u\text{-}sp2}$, $\nu_{u\text{-}sp3}$, $\mu_{u\text{-}sp3}$, and $\sigma_{u\text{-}sp3}$ for the ^{13}C distributions, and the corresponding eight parameters for the ^{1}H series. Despite the DP4+ can be computed "by hand", to facilitate the overall process an Excel spreadsheet is given as part of the Supporting Information (or from the authors at sarotti-NMR.weebly.com) that considerably simplifies the calculation.

As expected, the σ_s values (ranging from 0.09 to 0.14 ppm for ^{1}H and 1.23−2.09 ppm for ^{13}C) were considerably lower than those obtained in the original paper using MMFF geometries (0.185 and 2.306 ppm, respectively).[10] Such sharpening of the error distributions is a reflect of the higher accuracy of the NMR calculations used in this study, though the effect of using a different test set should not be neglected.

The performance of our improved DP4+ probability was evaluated with a challenging set of 48 examples (Figure 4) for which the original DP4 afforded unsatisfactory results. Compounds **73**−**80** were included in the original work of Smith and Goodman,[10] whereas compounds **81**−**90** were taken from recent publications.[18]

Figure 5 shows the performance of the DP4+ probabilities (computed at the 24 levels of theory discussed above using both proton and carbon data) on the stereoassignment of the 48 examples shown in Figure 4. The corresponding DP4 values (also shown in Figure 5) of compounds **73**−**80** were directly taken from ref 10, whereas in the case of compounds **81**−**90** the DP4 probabilities were computed as originally described.[10] To facilitate further discussion, a simple scoring system was arbitrarily introduced based on the DP4+ probability value calculated for a given compound. Depending on the confidence in the correct assignment, three main intervals were identified: > 95% (good), 50%−95% (medium) and <50% (bad), and each was given a different score: 3, 1, and 0 points, respectively.

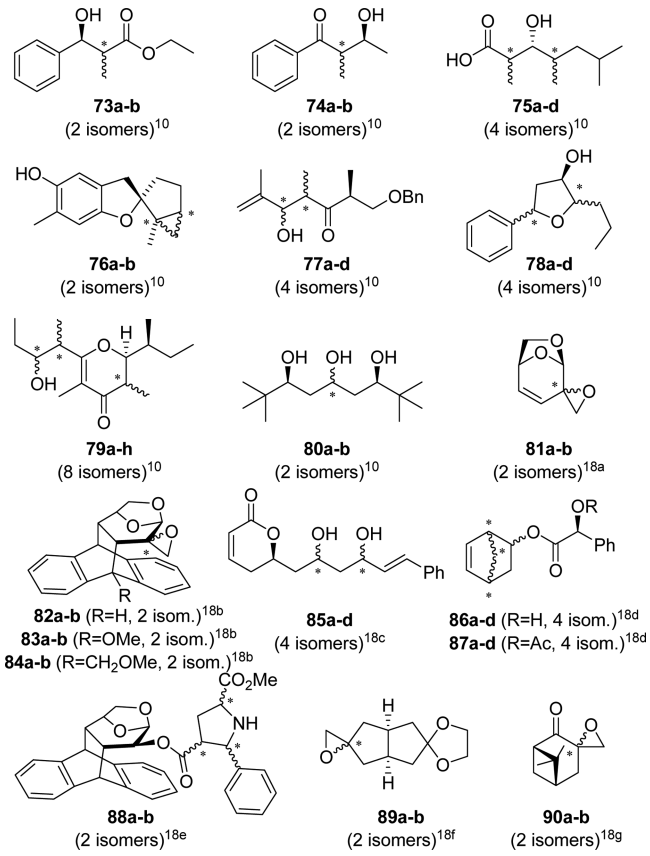


Figure 4. Test set of molecules used to evaluate the performance of the DP4+ probability.

This "three points for a win" standard, inspired in many sports leagues, was implemented to reward highly confident correct assignments, which represents the optimal scenario.[19]

From the data shown in Figure 5 several conclusions can be drawn: (a) All the 24 new DP4+ probabilities performed better than the original DP4 (up to 2.4 times), indicating a clear superiority in the stereochemical assignment of isomeric compounds. It is interesting to note that, in general, when the original DP4 successfully points to the correct isomer, our modified probability also does. On the opposite hand, many cases of incorrect assignment by DP4 could be reverted by DP4+ (for example, compounds **73b**, **75b**, **76a**, **77b**, **78c**, **82b**, and **86d**). In a few examples, both DP4 and DP4+

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

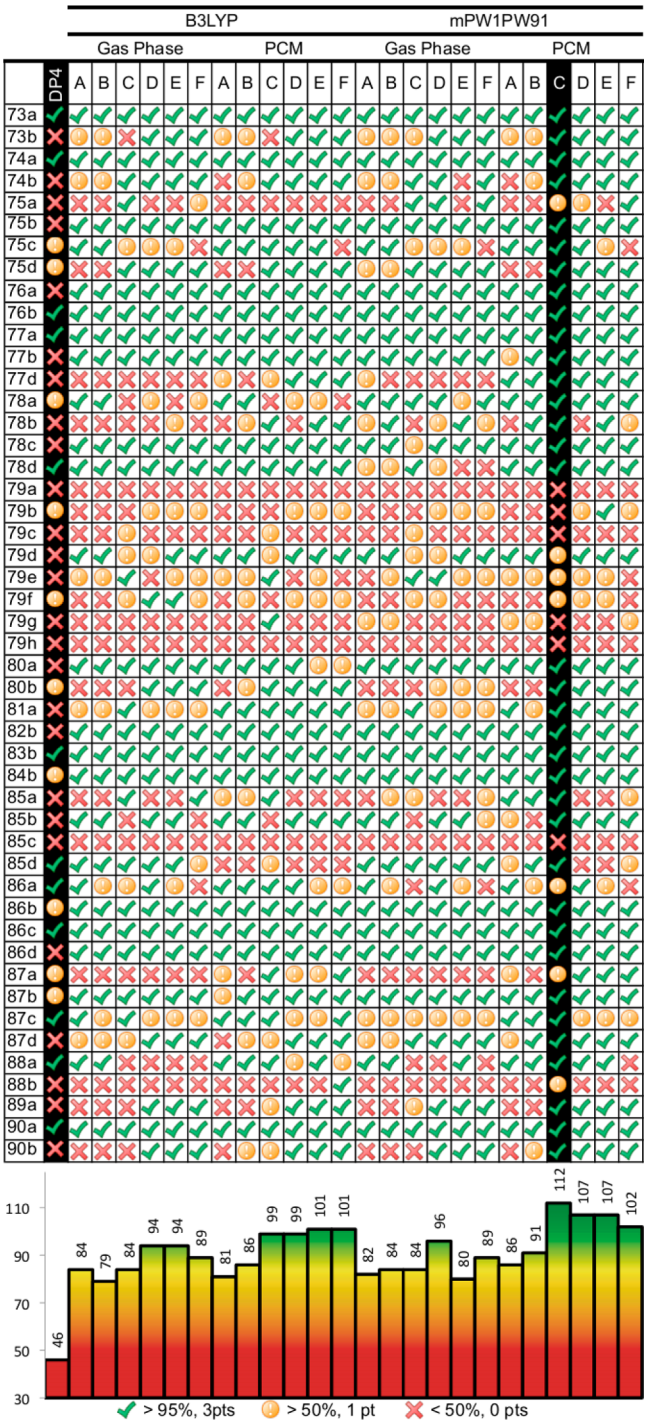


Figure 5. Overall performance of original DP4 (first column) and the new DP4+ probabilities computed for compounds **73**–**90** (Figure 4) at the 24 levels of theory under study (columns 2 to 25). In black shading are highlighted the original DP4 and the level that afforded the best DP4+ result (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*). Basis sets: A: 6-31G*; B: 6-31G**; C: 6-31+G**; D: 6-311G*; E: 6-311G**; F: 6-311+G**.

systematically failed in pointing toward the correct isomer (for example, compounds **79a**, **79h**, and **85c**).

However, even in these cases DP4+ performed better by reducing the known tendency of DP4 to overstate the probability when making incorrect assignments in high probability.[10] For instance, 11 of the 26 incorrectly assigned isomers by DP4 (42%) were made in >90% confidence, whereas none of the few wrong assignments made by DP4+ took place in such high certainty. This indicates that when the correlation between experimental and computed data cannot allow a fairly certain assignment, DP4+ does not advocate for any specific option. This important effect can be clarified in the example shown in Figure 6, in which both DP4 and DP4+ fail in correctly identifying **85c** as the correct isomer, but only DP4 is confident about the incorrect assignment.

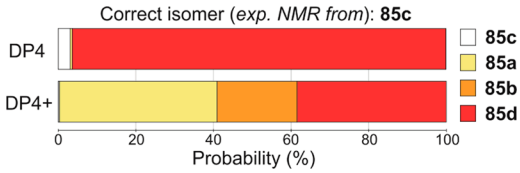


Figure 6. Graph of DP4 (B3LYP/6-31G**//MMFF) and DP4+ (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*) probabilities obtained by correlating the experimental NMR of **85c** with the calculated data of **85a**–**d**. The probability for the correct assignment is shown in white.

The level of theory used in the shift calculation procedure also displayed an interesting effect. In general, best results were obtained in solution with triple-ζ or double-ζ polarized basis sets including diffuse functions. Interestingly, these last levels afforded the sharper t (lower σ) series for the ^{1}H error distributions, but not necessarily for the ^{13}C series, indicating a clear relationship between the accuracy of proton NMR prediction with the DP4+ performance (one of the main hypothesis formulated in this work). Moreover, mPW1PW91 performed slightly better than B3LYP and coupled with the 6-31+G** basis set (in solution) was the best among the 24 levels of theory under study.

The parameters used to calculate DP4+ were taken from the test set shown in Figure 2. To support that choice, the 16 statistical terms [μ,σ,ν] obtained at the optimal level (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*) were recalculated by adding to the original set the 578 and 545 individual ^{13}C and ^{1}H errors, respectively, computed from the validation set (Figure 4) at the same level of theory. The new [μ,σ,ν] set closely matched the original values, as also did the corresponding DP4+ probabilities computed from this new set of parameters (in fact, both cases afforded the same scoring).

Case Study. The improved performance of DP4+ can be clearly seen in a recent case of natural products stereochemical uncertainty. In 2000, Cavalheiro and Yoshida reported the isolation of cryptomoscatone D1 and D2 from the bark of *Cryptocarya mandiocanna*. The absolute configuration of the dihydropyran-2-one center was set as (*R*) on the basis of positive Cotton effect on CD spectra, but they were unable to unambiguously define the absolute configurations of the two remaining stereocenters (though suggested a 1,3-*anti* and 1,3-*syn* relationships for cryptomoscatone D1 and D2, respectively).[20] Total synthesis of cryptomoscatone D2 by Yadav questioned the original assignment,[21] and the issue was finally resolved when Pilli and co-workers synthesized the four candidates and matched cryptomoscatone D1 and D2 with compounds **85b** and **85a**, respectively (Figure 7).[18c] In this case, our DP4+ probability could have been useful to settle the correct stereochemistry of the two natural products in high

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

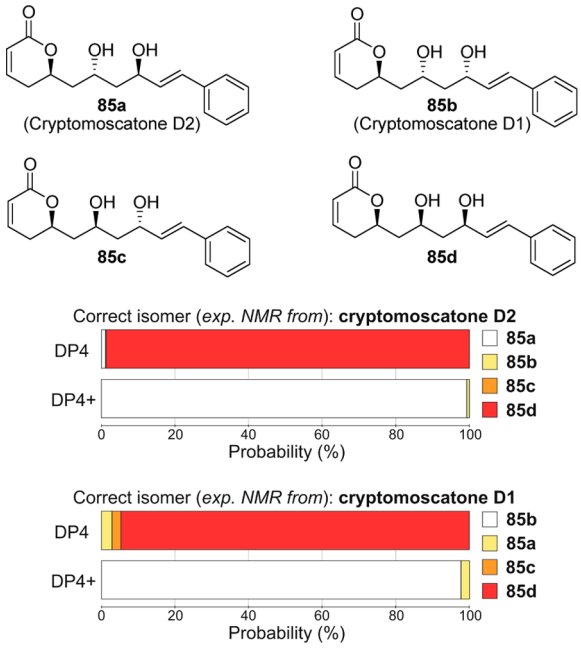


Figure 7. Graph of DP4 (B3LYP/6-31G**//MMFF) and DP4+ (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*) probabilities obtained by correlating the experimental NMR of cryptomoscatone D2 (**85a**) and D1 (**85b**) with the calculated data of **85a−d**. The probability for the correct assignment is shown in white.

confidence. On the other hand, original DP4 failed by systematically diagnosing **85d** as the correct isomer.

Origins of the Improved Performance of DP4+. The present work was founded on the main hypotheses that a better DP4-like statistical analysis could be developed by using more accurate NMR calculations and including unscaled shifts in the probability equations. Having established the clear improvement of the resulting method (Figure 5), we next aimed to better understand the actual contribution of each factor in the DP4+ outcome. First, we recomputed the probabilities of the 48 validation examples shown in Figure 4 using only scaled or unscaled shifts from the *s*DP4+ and *u*DP4+ terms, respectively, of eq 3. Figure 8 shows the performance scores of the resulting calculations averaged over the 24 levels of theory under study, along with the corresponding DP4 and DP4+ values.

Comparing the results obtained with *u*DP4+ and *s*DP4+ with the corresponding DP4+ values, it comes clear that the combination of both scaled and unscaled NMR shifts affords the highest assignment capacity. Second, taking into account

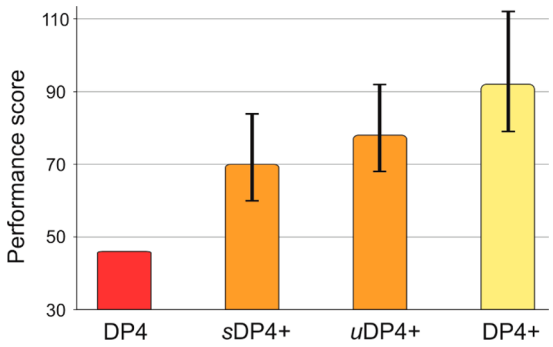


Figure 8. Overall performance scores of DP4, *s*DP4+, *u*DP4+, and DP4+, averaged over the 24 levels of theory used in this study.

that *s*DP4+ performs better than DP4 regardless the level of theory (~1.5 times in the average), it becomes evident that computing the NMR shifts from more robust geometries (B3LYP/6-31G* vs MMFF) resulted in a significant effect in the probability outcome. Moreover, the influence of the level of theory in the [σ, ν] values was also investigated. The DP4 probabilities of compounds **73−90** (Figure 4) were computed from eq 1 (using the original [σ, ν] parameters reported by Smith and Goodman), and the scaled NMR shifts (δ^i_s terms) from B3LYP/6-31G* geometries at the 24 levels under study. The resulting DP4 probabilities were next compared with the corresponding *s*DP4+ analogues, and found that the performance of the formers was always lower (up to 30%). Considering that both methods differ only in the [σ, ν] values, on the basis of the presented evidence it can be concluded that, even under the DP4 mathematical architecture, best results are obtained with the [σ, ν] terms computed at the same level of theory employed to obtain the NMR chemical shifts.[13] Finally, another relevant observation is that the use of unscaled shifts by *u*DP4+ affords even better results (~1.7 times in the average) that those obtained with *s*DP4+, suggesting that during the shift scaling by linear regression some valuable data to differentiate between the candidate structures is lost. This effect can be better understood with a particular example. As shown in Figure 9, aldol **74b** was incorrectly assigned as **74a** by the

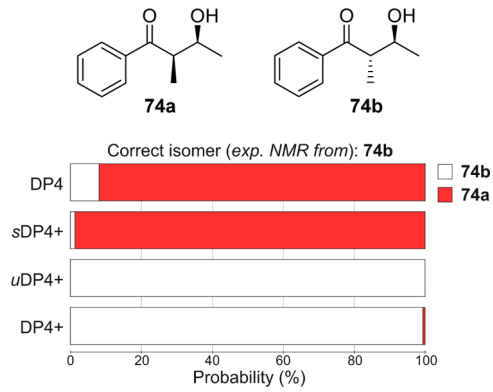


Figure 9. Graph of DP4 (B3LYP/6-31G**//MMFF) and DP4+ (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*) probabilities obtained by correlating the experimental NMR of aldol **74b** with the calculated data of **74a−b**. The probability for the correct assignment is shown in white.

original DP4 in high confidence (92%). The same was observed with scaled shifts (*s*DP4+) at the PCM/mPW1PW91/6-31+G**//B3LYP/6-31G* level. However, a neat inversion in the probability was observed when using unscaled shifts (*u*DP4+). This probability overrides the former misassignment made by *s*DP4+ and the combined DP4+ method confidently (>99%) identifies **74b** as the right isomer. Inspection of the scaled and unscaled ^{13}C shifts computed for the **74ab** pair clarifies this interesting behavior (Table 1).

The scaled shifts of **74a** are closer to the experimental values than those compued for **74b** (CMAE 1.2 vs 1.5 ppm, respectively). As the individual errors from **74a** are generally lower, the associated probabilities are higher resulting in the (incorrect) *s*DP4+ assignment depicted in Figure 9. Interestingly, the unscaled shifts also match better for **74a** (MAE 1.3 vs 2.1 ppm, respectively). Shouldn't then **74a** be the correct isomer? To understand why *u*DP4+ points in the opposite

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

Table 1. Experimental ^{13}C NMR Shifts of Aldol 74b, along with the Corresponding Unscaled and Scaled Chemical ^{13}C Shifts Computed for Isomers 74a and 74b at the PCM/mPW1PW91/6-31+G//B3LYP/6-31G* Level of Theory**

atom N°	δ_{exp} (74b)	δ_u (unscaled shifts)	unscaled error ($\delta_u - \delta_{exp}$)	δ_s (scaled shifts)	scaled error ($\delta_s - \delta_{exp}$)
		Compound **74a** (incorrect isomer)			
C-1	205.2	207.0	1.8	207.3	2.1
C-2	47.6	47.9	0.3	48.9	1.3
C-3	69.7	68.9	−0.8	69.8	0.1
C-4	20.7	21.2	0.5	22.3	1.6
C-5	15.2	12.8	−2.4	13.9	−1.3
C-6	136.4	132.8	−3.6	133.4	−3.0
C-7	128.6	128.3	−0.3	128.9	0.3
C-8	128.3	126.5	−1.8	127.1	−1.2
C-9	133.2	132.6	−0.6	133.2	0.0
		MAE	***1.3***	***CMAE***	***1.2***
		Compound **74b** (correct isomer)			
C-1	205.2	206.0	0.8	208.3	3.1
C-2	47.6	50.0	2.4	48.2	0.6
C-3	69.7	72.1	2.4	70.9	1.2
C-4	20.7	22.3	1.6	19.8	−0.9
C-5	15.2	19.5	4.3	16.9	1.7
C-6	136.4	133.0	−3.4	133.4	−3.0
C-7	128.6	127.7	−0.9	128.0	−0.6
C-8	128.3	126.5	−1.8	126.8	−1.5
C-9	133.2	132.2	−1.0	132.6	−0.6
		MAE	***2.1***	***CMAE***	***1.5***

direction (toward **74b**), it must be recalled that the *t* distribution of errors might not be centered at zero (in fact, they commonly don't). For instance, at the actual level of theory, $\mu_{u\text{-}sp2} = -0.9$ ppm and $\mu_{u\text{-}sp3} = -2.9$ ppm for sp^2- and sp^3-hybridized carbon atoms, respectively. As a consequence, a high individual probability can be computed from large errors (and vice versa). For example, the experimental ^{13}C shift of C-5 is 15.2 ppm, and the corresponding computed (unscaled) values are 12.8 ppm (**74a**) and 19.5 ppm (**74b**). Despite the fact that the error is smaller for the former (−2.4 ppm vs 4.3 ppm), the "real" error (that is, the difference between the computed error and the center of the distribution) is only 1.4 ppm (4.3−2.9) for **74b** and −5.3 ppm (−2.4−2.9) for **74a**. Therefore, the associated probability is, in fact, much higher for **74b** (0.8% vs 20.5%, respectively).

Proton or Carbon Data? At this stage of the study we computed the DP4+ probabilities using both ^{1}H and ^{13}C data. The arguments supporting this choice rest on the main assumption that a better assignment could be made with the more information available. However, in this work we found that the levels of theory with sharper *t* distribution of ^{1}H errors (but not necessarily for the ^{13}C series) led to better DP4+ probabilities in terms of correct stereochemical assignment. To determine the relative importance of each nucleus (^{1}H and ^{13}C), we broke down the DP4+ probabilities computed for the 48 compounds of Figure 4 into the corresponding ^{1}H and ^{13}C probabilities. The performance scores (averaged over the 24 levels of theory) are shown in Figure 10.

Interestingly, the assignment ability of DP4+ using only ^{1}H or ^{13}C shifts is actually better than that of DP4 with all data, providing additional evidence of the improvement exerted by our modifications in the original formulation. In the average, ^{1}*H*-DP4+ and ^{13}*C*-DP4+ displayed similar overall results,

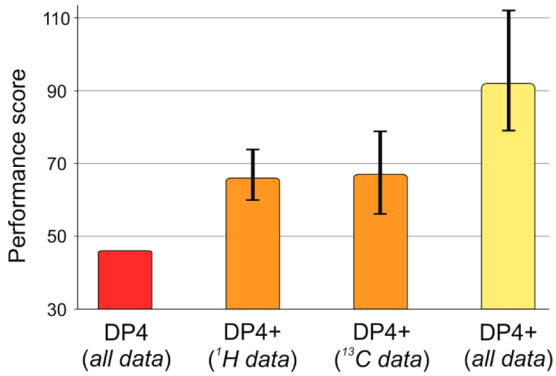


Figure 10. Overall performance scores of DP4, ^{1}*H*-DP4+, ^{13}*C*-DP4+, and DP4+, averaged over the 24 levels of theory used in this study.

suggesting that neither nuclei is more discriminating that the other. Naturally, the combination of both affords a clear enhancement in the resulting DP4+ performance, indicating that all data is important and must be used when available. Careful analysis of the disaggregated results (see Supporting Information) allowed us to observe that an eventual stereochemical misassignment made by ^{1}*H*-DP4+ can be often corrected by ^{13}*C*-DP4+ (and vice versa). This can be illustrated by the following example (Figure 11). The stereoassignments

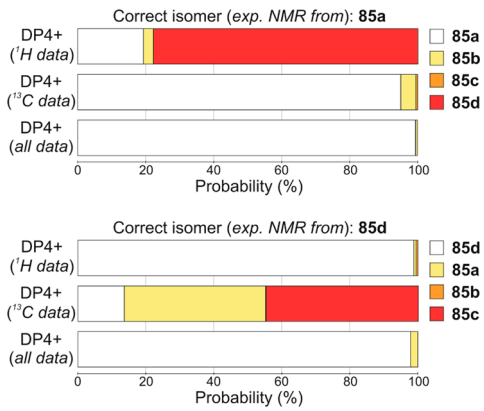


Figure 11. Graph of ^{1}*H*-DP4+, ^{13}*C*-DP4+, and DP4+ (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*) probabilities obtained by correlating the experimental NMR of compounds **85a** and **85d** with the calculated data of **85a−d**. The probability for the correct assignment is shown in white.

of **85a** and **85d** are wrong on the basis of ^{1}H and ^{13}C data, respectively. In the first case, compound **85d** is strongly identified as the correct isomer by ^{1}*H*-DP4+ (78%), whereas in the second case the highest ^{13}*C*-DP4+ probabilities are computed for **85a** (42%) and **85c** (45%). However, such misassignments are corrected when including all the data, leading to the right DP4+ assignment of **85a** (99%) and **85d** (98%), respectively. This example strengthens the argument in favor of using ^{1}H as well as ^{13}C shifts to arrive at confident conclusions, providing a useful illustration of the different influence of proton and carbon data in the stereoassignment of similar molecules even at the same level of theory.

Returning to the original discussion, Figure 12 shows the relative effect exerted by each nucleus (in terms of DP4+ assignment capacity) by subtracting the performance scores computed using proton data to the corresponding values obtained from carbon data (^{13}*C*-DP4+ score - ^{1}*H*-DP4+ score).

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

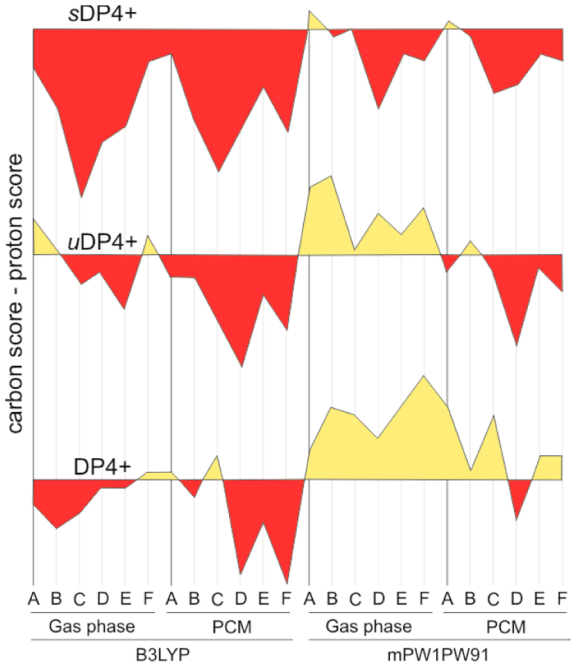


Figure 12. Difference between ^{13}C-DP4+ and ^{1}H-DP4+ performance scores computed at all levels of theory. Negative values (in red) or positive values (in green) indicate that the most important contribution is made by proton or carbon data, respectively.

Interestingly, such difference is almost always negative when using scaled shifts, indicating that in these cases proton data makes the most important contribution. However, this predominance becomes diffuse when including unscaled data in the DP4+ calculations. In particular, ^{1}H NMR seems to be more influential when computing the shielding tensors with B3LYP, while the opposite is observed for the mPW1PW91 functional.

Recently, in an interesting debate about what nucleus is more relevant for stereochemical assignment, proton was found as the most discriminating one.[12] However, from the data herein presented, both nuclei are important and must be used when possible to compute the DP4+ probability.

■ CONCLUSION

We have developed a new probability (DP4+) as a tool for the important and difficult task of GIAO NMR-based structural or stereochemical assignment of organic molecules with only one set of experimental data available. Inclusion of unscaled shifts in the probability formulation and using higher levels of theory in the NMR calculation procedure resulted in a significant improvement in the overall performance of the DP4+ probability, providing accurate and confident results in establishing the stereochemistry of 48 challenging isomeric compounds. To simplify the DP4+ calculation procedure, an Excel file is given in the Supporting Information (or from the authors at sarotti-NMR.weebly.com).

■ EXPERIMENTAL SECTION

Computational Methods. All the quantum mechanical calculations were performed using Gaussian 09.[22] In the case of conformationally flexible compounds, the conformational search was done in the gas phase using the MM+ force field (implemented in Hyperchem),[23] and/or the MMFF force field (implemented in Spartan 08).[24] All conformers within 5 kcal/mol of the lowest energy conformer were subjected to further reoptimization at the B3LYP/6-31G* level of theory. The choice for the 5 kcal/mol of cutoff was set as a balance between reducing the overall CPU calculation time and minimizing the possibility of losing further contributing conformers. The conformations within 2 kcal/mol from the B3LYP/6-31G* global minima were subjected to NMR calculations. The magnetic shielding constants (σ) were computed using the gauge including atomic orbitals (GIAO) method,[25] the method of choice to solve the gauge origin problem, with two different DFT functionals: B3LYP and mPW1PW91. The calculations were carried out both in the gas phase and in solution (using the polarizable continuum model, PCM,[26] with chloroform as the solvent), and six different basis sets: 6-31G*, 6-31G**, 6-31+G**, 6-311G*, 6-311G**, and 6-311+G**). The unscaled chemical shifts (δ_u) were computed using TMS as reference standard according to $\delta_u = \sigma_0 - \sigma^x$, where σ^x is the Boltzmann averaged shielding tensor (over all significantly populated conformations) and σ_0 is the shielding tensor of TMS computed at the same level of theory employed for σ^x. The Boltzmann averaging was done according to eq 4:

$$\sigma^x = \frac{\sum_i \sigma_i^x e^{(-E_i/RT)}}{\sum_i e^{(-E_i/RT)}} \quad \text{(eq 4)}$$

where σ_i^x is the shielding constant for nucleus x in conformer i, R is the molar gas constant (8.3145 J K^{-1} mol^{-1}), T is the temperature (298 K), and E_i is the energy of conformer i (relative to the lowest energy conformer), obtained from the single-point NMR calculation at the corresponding level of theory. The scaled chemical shifts (δ_s) were computed as $\delta_s = (\delta_u - b)/m$, where m and b are the slope and intercept, respectively, resulting from a linear regression calculation on a plot of δ_u against δ_{exp}. The $[\mu,\sigma,\nu]$ terms were obtained by fitting the errors to a t distribution using MATLAB 7.0.22.[27]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.joc.5b02396.

Instructions for DP4+ Excel file, full list of compounds, experimental shifts, GIAO isotropic shielding tensors of all compounds, and all computational data associated with this paper (PDF)

Excel file for DP4+ calculations (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sarotti@iquir-conicet.gov.ar

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by UNR (BIO 316), ANPCyT (PICT 2011-0255 and PICT-2012-0970), and CONICET (PIP 11220130100660CO). N.G. and M.M.Z. thank CONICET for the award of a fellowship.

■ REFERENCES

(1) Nicolaou, K. C.; Vourloumis, D.; Winssinger, N.; Baran, P. S. *Angew. Chem., Int. Ed.* **2000**, *39*, 44.

(2) For leading reviews, see: (a) Nicolaou, K. C.; Snyder, S. A. *Angew. Chem., Int. Ed.* **2005**, *44*, 1012. (b) Suyama, T. L.; Gerwick, W. H.; McPhail, K. L. *Bioorg. Med. Chem.* **2011**, *19*, 6675. (c) Maier, M. E. *Nat. Prod. Rep.* **2009**, *26*, 1105. For recent references, see: (d) Nicolaou, K. C.; Shah, A. A.; Korman, H.; Khan, T.; Shi, L.; Worawalai, W.; Theodorakis, E. A. *Angew. Chem., Int. Ed.* **2015**, *54*, 9203. (e) Zhu, L.; Liu, Y.; Ma, R.; Tong, R. *Angew. Chem., Int. Ed.* **2015**, *54*, 627. (f) Xiao, Q.; Young, K.; Zakarian, K. *J. Am. Chem. Soc.*

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534

2015, *137*, 5907. (g) Willwacher, J.; Fürstner, A. *Angew. Chem., Int. Ed.* **2014**, *53*, 4217. (h) Nicolaou, K. C.; Hale, C. R. H.; Nilewski, C.; Ioannidou, H. A.; ElMarrouni, A.; Nilewski, L. G.; Beabout, K.; Wang, T. T.; Shamoo, Y. *J. Am. Chem. Soc.* **2014**, *136*, 12137.

(3) Gil, R. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 7222. (b) Breton, R. C.; Reynolds, W. F. *Nat. Prod. Rep.* **2013**, *30*, 501.

(4) (a) Sarotti, A. M. *Org. Biomol. Chem.* **2013**, *11*, 4847. (b) Zanardi, M. M.; Sarotti, A. M. *J. Org. Chem.* **2015**, *80*, 9371.

(5) For recent reviews, see: (a) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839. (b) Bagno, A.; Saielli, G. *Wiley Interdisciplinary Reviews: Computational Molecules Science* **2015**, *5*, 228. (c) Tantillo, D. J. *Nat. Prod. Rep.* **2013**, *30*, 1079. (d) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107*, 3744−3779. (e) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. *Nat. Protoc.* **2014**, *9*, 643.

(6) (a) Bagno, A.; Rastrelli, F.; Saielli, G. *Chem. - Eur. J.* **2006**, *12*, 5514−5525. (b) Bagno, A.; Rastrelli, F.; Saielli, G. *J. Phys. Chem. A* **2003**, *107*, 9964. (c) Bagno, A. *Chem. - Eur. J.* **2001**, *7*, 1652.

(7) (a) Barone, G.; Gomez-Paloma, L.; Duca, D.; Silvestri, A.; Riccio, R.; Bifulco, G. *Chem. - Eur. J.* **2002**, *8*, 3233−3239. (b) Barone, G.; Duca, D.; Silvestri, A.; Gomez-Paloma, L.; Riccio, R.; Bifulco, G. *Chem. - Eur. J.* **2002**, *8*, 3240−3245.

(8) For leading references, see: (a) Li, Y. *RSC Adv.* **2015**, *5*, 36858. (b) Cen-Pacheco, F.; Rodríguez, J.; Norte, M.; Fernández, J. J.; Hernández Daranas, A. *Chem. - Eur. J.* **2013**, *19*, 8525. (c) Lodewyk, M. W.; Soldi, C.; Jones, P. B.; Olmstead, M. M.; Rita, J.; Shaw, J. T.; Tantillo, D. J. *J. Am. Chem. Soc.* **2012**, *134*, 18550. (d) Quasdorf, K. W.; Huters, A. D.; Lodewyk, M. W.; Tantillo, D. J.; Garg, N. K. *J. Am. Chem. Soc.* **2012**, *134*, 1396. (e) Saielli, G.; Nicolaou, K. C.; Ortiz, A.; Zhang, H.; Bagno, A. *J. Am. Chem. Soc.* **2011**, *133*, 6072. (f) Lodewyk, M. W.; Tantillo, D. J. *J. Nat. Prod.* **2011**, *74*, 1339.

(9) Smith, S. G.; Goodman, J. M. *J. Org. Chem.* **2009**, *74*, 4597.

(10) Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946.

(11) For leading references, see: (a) Riveira, M. J.; Trigo-Mouriño, P.; Troche-Pesqueira, E.; Martin, G. E.; Navarro-Vázquez, A.; Mischne, M. P.; Gil, R. R. *J. Org. Chem.* **2015**, *80*, 7396. (b) Willwacher, J.; Heggen, B.; Wirtz, C.; Thiel, W.; Fürstner, A. *Chem. - Eur. J.* **2015**, *21*, 10416. (c) Wyche, T. P.; Piotrowski, J. S.; Hou, Y.; Braun, D.; Deshpande, R.; McIlwain, S.; One, I. M.; Myers, C. L.; Guzei, I. A.; Westler, W. M.; Andes, D. R.; Bugni, T. S. *Angew. Chem., Int. Ed.* **2014**, *53*, 11583. (d) Cen-Pacheco, F.; Norte, M.; Fernández, J. J.; Daranas, A. H. *Org. Lett.* **2014**, *16*, 2880. (e) Cen-Pacheco, F.; Rodríguez, J.; Norte, M.; Fernández, J. J.; Hernández Daranas, A. *Chem. - Eur. J.* **2013**, *19*, 8525. (f) Rodriguez, J.; Nieto, R. M.; Blanco, M.; Valeriote, F. A.; Jimenez, C.; Crews, P. *Org. Lett.* **2014**, *16*, 464. (g) 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. *Angew. Chem., Int. Ed.* **2011**, *50*, 3219. (h) Riveira, M. J.; Gayathri, C.; Navarro-Vázquez, A.; Tsarevsky, N. V.; Gil, R. R.; Mischne, M. P. *Org. Biomol. Chem.* **2011**, *9*, 3170. (i) Nazarski, R. B.; Pasternak, B.; Lesniak, S. *Tetrahedron* **2011**, *67*, 6901. (j) Shepherd, D. J.; Broadwith, P. A.; Dyson, B. S.; Paton, R. S.; Burton, J. W. *Chem. - Eur. J.* **2013**, *19*, 12644.

(12) (a) Chini, M. G.; Riccio, R.; Bifulco, G. *Eur. J. Org. Chem.* **2015**, *6*, 1320. (b) Marell, D. J.; Emond, S. J.; Kulshrestha, A.; Hoye, T. R. *J. Org. Chem.* **2014**, *79*, 752.

(13) We noticed that DP4 is frequently computed using other levels of theory (such as PBE0/pcS-2//OPBE/6-31G* [ref 11a] or mPW1PW91/6-31G**//B3LYP/6-31G** [ref 11e]), for which the statistical parameters that define the probability (σ, T^{υ}) were collected (B3LYP/6-31G**//MMFF). Despite the conclusions arrived in those studies are indisputable, it should be noted that the DP4 values are strongly related to the distribution parameters set (σ, T^{υ}), that in turn depend upon the actual level of theory employed in the NMR calculation procedure. This issue is also highlighted in ref 11i.

(14) Andrews, K. G.; Spivey, A. C. *J. Org. Chem.* **2013**, *78*, 11302.

(15) The experimental NMR shifts and the full references of the original papers are provided in the Supporting Information.

(16) It is known that the B3LYP functional is not the optimal for many purposes Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670 Nevertheless, it has been widely used in quantum chemistry calculation of NMR parameters. For that reason, we choose B3LYP/6-31G* to perform the geometry optimization step, providing better geometries than MMFF or related force fields..

(17) (a) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254. (b) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2012**, *77*, 6059.

(18) (a) Gelas-Mialhe, Y.; Gelas, J. *Carbohydr. Res.* **1990**, *199*, 243. (b) Zanardi, M. M.; Suárez, A. G. *Tetrahedron Lett.* **2014**, *55*, 5832. (c) Toneto Novaes, L. F.; Drekener, R. L.; Avila, C. M.; Pilli, R. A. *Tetrahedron* **2014**, *70*, 6467. (d) Pisano, P. L.; Sarotti, A. M.; Pellegrinet, S. C. *Tetrahedron Lett.* **2009**, *50*, 6121. (e) Llompart, D. F.; Sarotti, A. M.; Corne, V.; Suárez, A. G.; Spanevello, R. A.; Echeverría, G. A.; Piro, O. E.; Castellano, E. E. *Tetrahedron Lett.* **2014**, *55*, 2394. (f) Leonard, J.; Hewitt, J. D.; Ouali, D.; Bennett, L. R.; Mahmood, A.; Simpson, S. J. *Tetrahedron* **2002**, *58*, 4681. (g) Badjah-Hadj-Ahmed, A. Y.; Meklati, B. Y.; Waton, H.; Pham, Q. T. *Magn. Reson. Chem.* **1992**, *30* (9), 807.

(19) Considering that 48 examples have been used in the DP4+ evaluation, 144 points represent the highest score achievable, corresponding to a successful identification of the correct structure in all cases with more than 95% probability.

(20) Cavalheiro, A. J.; Yoshida, M. *Phytochemistry* **2000**, *53*, 811.

(21) Yadav, J. S.; Ganganna, B.; Bhunia, D. C. *Synthesis* **2012**, *44*, 1365.

(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr., Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, 2009.

(23) *Hyperchem Professional Release 7.52*; Hypercube, Inc., 2005.

(24) Spartan'08; Wavefunction:. Irvine, CA.

(25) (a) Ditchfield, R. *J. Chem. Phys.* **1972**, *56*, 5688. (b) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789. (c) Rohlfing, C. M.; Allen, L. C.; Ditchfield, R. *Chem. Phys.* **1984**, *87*, 9. (d) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251.

(26) For a review on continuum solvation models, see: Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(27) MATLAB; MathWorks: Natick, MA, 2007.

DOI: 10.1021/acs.joc.5b02396
J. Org. Chem. 2015, 80, 12526−12534