Structural Determination of Complex Natural Products by Quantum Mechanical Calculations of ^{13}C NMR Chemical Shifts: Development of a Parameterized Protocol for Terpenes

Ana Carolina F. Albuquerque, Daniel J. Ribeiro & Mauro B. Amorim

Introduction

In the last decades, Nuclear Magnetic Resonance (NMR) spectroscopy has established itself as one of the most important techniques for structural determination of natural products.[1] Despite the current advances in multidimensional techniques and in probe technology,[1,2] instances of revision of structures erroneously established for natural products are still common in the literature.[3] With the recent development of quantum mechanical methods and the availability of modern computers, the prediction of chemical shifts through *ab initio* and Density Functional Theory (DFT) calculations has become a very powerful tool for assistance in the assignment of chemical shifts and in the structural determination of complex organic molecules, such as natural products.[4-6]

With the aim to achieve a good ratio between accuracy and computational cost, calculated chemical shifts can be corrected through the use of scaling factor procedures. It has been established in the literature that the application of this approach can reduce systematic errors inherent in theoretical calculations.[6] Even though in recent years there has been the development of several scaling factors,[7-11] there are no studies in the literature reporting the use of scaling factors parameterized for a specific class of natural products.

In this work, we present the development of a calculation protocol for terpenes, a class of natural products with a board distribution among plant species and with great importance due to its biological and pharmacological activities.[12] This protocol consists of GIAO-DFT calculations of chemical shifts and application of a scaling factor parameterized with terpenes, in order to ensure accurate structural determination of this class of natural products.

Methods

In order to reduce the computational costs in our calculations, we parameterized the scaling factor using sesquiterpene molecules, a sub-class of terpenes with structural frameworks containing only 15 carbon atoms.[13]

A set of 10 sesquiterpene molecules (figure 1), whose structures have been reliably elucidated in literature,[14-23] were selected and submitted to randomized conformational searches using Monte Carlo method and MMFF force field. The most significant conformations of each compound, considering an initial energy cutoff of 10 kcal.mol^{-1}, were selected to single-point energy calculations at the B3LYP/6-31G(d) level of theory. All conformations within 5 kcal.mol^{-1} of energy were selected to geometry optimization calculations carried out at the mPW1PW91/6-31G(d) level of theory. Population-averaged ^{13}C nuclear magnetic shielding constants (δ) were calculated using GIAO method at the same level of theory and assuming Boltzmann statistics. Chemical shifts (δ) were obtained as $\delta_{calc} = \delta_{TMS} - \delta$, where δTMS is the shielding constant of the reference compound, tetramethylsilane (TMS), calculated at the same level of theory. All quantum mechanical calculations were performed in gas phase, using Gaussian 09 software package.[24]

The scaling factor was generated by plotting calculated (δ_{calc}) against experimental chemical shifts of the set of sesquiterpenes. Thus, slope (a) and intercept (b) values obtained from this linear regression can be used to generate scaled chemical shifts (δ_{scal}) from as δ_{calc}, using the expression $\delta_{scal} = a . \delta_{calc} + b$.

In order to validate the protocol, the scaling factor was used to obtain δ_{scal} for (4S)-3-ishwarone (11), with a rare ishwarane squeleton,[25] and 9¬-epi-presilphiperfolan-1-ol (13), a triquinane sesquiterpene.[26] Additionally, aiming to evaluate the ability of the method to distinguish possible isomers of the terpenes, we also computed chemical shifts for the (4R)-diastereomer of 3-ishwarone (12),27 and for the known regioisomer of 9-epi-presilphiperfolan-1-ol, presiphiperfolan-9-ol (14)[28] (figure 2).

Finally, with the purpose of demonstrating that this calculation protocol can be applied for the prediction of chemical shifts of larger terpenes, we computed chemical shifts for the triterpene oleana-12(13),15(16)-dienoic acid (15)[29] (figure 2).

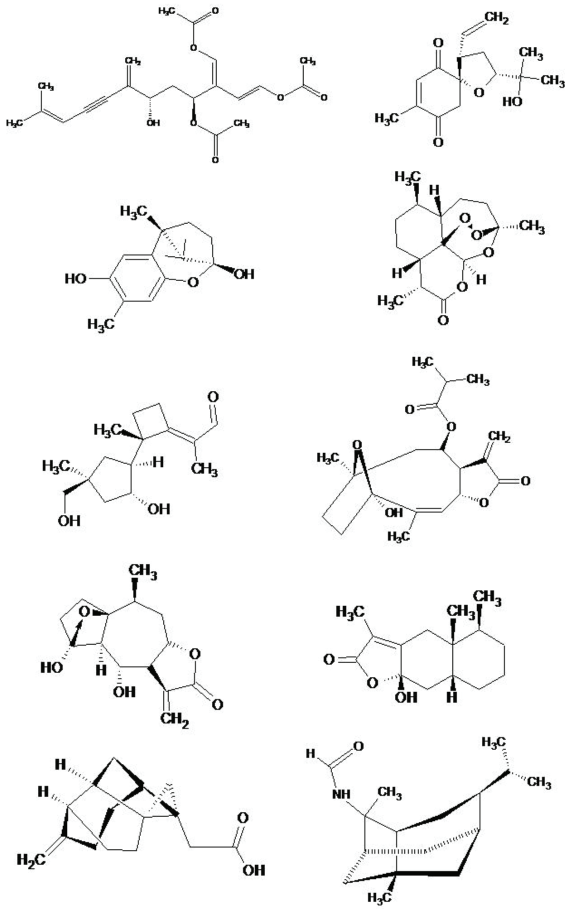

Figure 1. Structures of sesquiterpenes 1-10, used for parameterization.

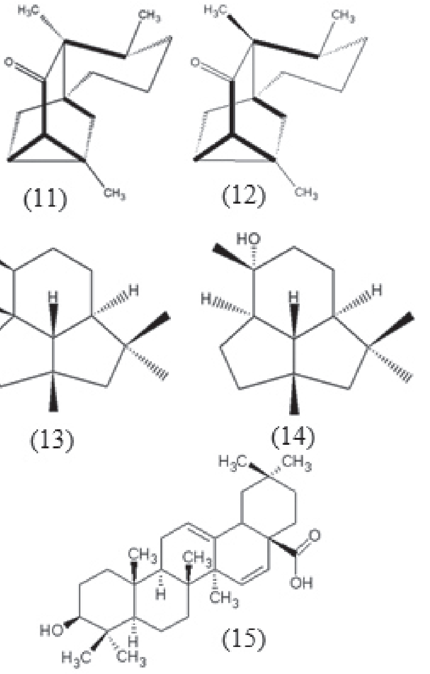

Figure 2. Structures of (4S)-3-ishwarone (11), (4R)-3-ishwarone (12), 9-*epi*-presilphiperfolan-1-ol (13), presiphiperfolan-9-ol (14) and oleana-12(13), 15(16)-dienoic acid (15).

Results and Discussion

With the purpose to ensure that the calculation protocol could be applied to any terpene, it was selected a series of sesquiterpenes with varied and complex structural frameworks (figure 1). Its ^{13}C chemical shifts were calculated according to procedures described in previous section. Statistical parameters, obtained from calculated and experimental chemical shifts of molecules 1-10, are shown in table 1.

Table 1. Statistical parameters obtained from δ_{calc} and δ_{exp} ^{13}C chemical shifts of molecules 1-10.

Statistical parameters	Values
Mean Absolute Deviation (MAD)	2.91 ppm
Root Mean Square Deviation (RMSD)	4.08 ppm
Correlation coefficient (r²)	0.9987
Slope (*a*)	1.0564
Intercept (*b*)	-2.1945

Both MAD and RMSD parameters indicate errors between theoretical and experimental data within the range of expected deviations for ^{13}C nuclei calculated with DFT.6 Furthermore, r² value also indicates an excellent correlation between calculated and experimental chemical shifts.

Thereby, results in table 1 show that the calculation protocol employed in this work, even without the application of a scaling factor, can reproduce experimental ^{13}C chemical shifts with satisfactory accuracy.

Slope (*a*) and intercept (*b*) values in table 1 were used to obtain scaled chemical shifts for terpenes 11-15. Table 2 shows MAD, RMSD, maximum error (E_{max}) and r² values, obtained from theoretical and experimental data, before and after application of the scaling factor, for molecules 11, 13 and 15.

Table 2. MAD, RMSD, Emax and r² values obtained from theoretical and experimental ^{13}C chemical shifts for molecules 11, 13 and 15.

Statistical parameters	(4S)-3-ishwarone (11)	
	Before linear scaling	After linear scaling
MAD	1.33	0.46
RMSD	2.56	0.60
E_{max}	9.30	1.20
r²	0.9999	0.9999
	9-epi-presilphiperfolan-1-ol (13)	
	Before linear scaling	After linear scaling
MAD	1.11	0.54
RMSD	1.49	0.75
E_{max}	2.90	1.85
r²	0.9988	0.9988
	Oleana-12(13),15(16)-dienoic acid (15)	
	Before linear scaling	After linear scaling
MAD	1.71	1.35
RMSD	2.84	1.89
E_{max}	10.96	5.96
r²	0.9982	0.9982

The values of the statistical parameters, shown in table 2, indicate once again that the calculation protocol employed to the three validation molecules (11, 13 and 15) can generate theoretical chemical shifts well converged towards experimental values.

Particularly, the significant reduction in MAD, RMSD and E_{max} values after application of a linear scaling procedure shows that this is a powerful tool to reduce errors from calculated chemical shifts.

Furthermore, the results obtained for triterpene 15 show that the calculation protocol developed in this work can be applied for the prediction of ^{13}C chemical shifts of higher classes of terpenes with reasonable accuracy.

In order to demonstrate that the theoretical method employed in this work is able to accurately distinguish (4S)-3-ishwarone (11) from its diastereomers, we made a comparison of the experimental data obtained for 11 with the theoretical data for its (4R)-diastereomer, 12. Statistical values obtained from this comparison are presented in table 3.

Table 3. MAD, RMSD, E_{max} and r² values obtained from theoretical ^{13}C chemical shifts of molecule 12 and experimental data of molecule 11.

Statistical parameters	**Before linear scaling**	**After linear scaling**
MAD	3.23	2.47
RMSD	4.66	3.68
E_{max}	11.64	10.78
r²	0.9939	0.9939

The comparison between experimental chemical shifts from the (4S)-diastereomer of 3-ishwarone (11) with theoretical chemical shifts from the (4R)-diastereomer (12) yield significantly larger deviations than those shown in table 2. Therefore, these data indicate that the calculation protocol is able to successfully distinguish between diastereomers.

In a similar manner, with the purpose to evaluate the ability of the method to distinguish regioisomers, we also made a comparison between the experimental data from 9-*epi*-presilphiperfolan-1-ol (13) with the theoretical data obtained for one of its regioisomers, presilphiperfolan-9-ol (14). Statistical values obtained from this comparison are presented in table 4.

Table 4. MAD, RMSD, E_{max} and r^2 values obtained from theoretical ^{13}C chemical shifts of molecule 14 and experimental data of molecule 13.

Statistical parameters	Before linear scaling	After linear scaling
MAD	8.45	8.35
RMSD	12.64	12.88
Emax	35.57	37.45
r^2	0.5301	0.5301

Statistical values shown in table 4 indicate a worst correlation and larger deviations between the two data sets than those in table 2. Thereby, we can safely assume that the calculation protocol employed in this work is able to accurately distinguish between two regioisomers.

Conclusions

Considering a set of 10 sesquiterpenes, we developed a parameterized protocol for the calculation of ^{13}C NMR chemical shifts of terpenes. This protocol, consisted of GIAO-DFT calculations and a linear scaling factor, was able to yield calculated chemical shifts with satisfactory accuracy.

Therefore, the calculation protocol developed in this work is a very attractive tool as an alternative to more computationally demanding approaches for the calculation of complex organic structures, such as terpenes.

Acknowledgments

The authors are grateful for the support given from FAPERJ, CAPES and CNPq.

References

1. A. G. Petrovic, A. Navarro-Vázquez, J. L. Alonso-Gómez, Curr. Org. Chem., 14, 1612, **2010**
2. F. E. Koehn and G. T. Carter, Nat. Rev., 4, 206, **2005**
3. K. C. Nicolaou and S. A. Snyder, Ang. Chem. Int. Ed., 44, 1012, **2005**
4. T. Helgaker, M. Jaszunski, K. Ruud, Chem. Rev., 99, 293, **1999**
5. G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, Chem. Rev., 107, 3744, **2007**
6. M. W. Lodewyk, M. R. Siebert, D. J. Tantillo, Chem. Rev., 112, 1839, **2012**
7. N. J. R. van E. Hommes and T. Clark, J. Mol. Model., 11, 175, **2005**
8. G. K. Pierens, J. Comput. Chem., 35, 1388, **2014**
9. A. E. Aliev, D. Courtier-Murias, S. Zhou, J. Mol. Struct. (THEOCHEM), 893, 1, **2009**
10. F. L. P. Costa, A. C. F. de Albuquerque, R. M. Borges, F. M. dos S. Junior, M. B. de Amorim, J. Comput. Theor. Nanosci., 11, 1, **2014**
11. F. L. P. Costa, A. C. F. de Albuquerque, F. M. dos S. Junior, M. B. de Amorim, J. Phys. Org. Chem., 23, 972, **2010**
12. P. M. Dewick, Medicinal Natural Products: A Biosynthetic Approach, 3 ed., Wiley, **2007**
13. B. M. Fraga, Nat. Prod. Rep., 30, 1226, **2013**
14. L. Commeiras, J. Thibonnet, J. L. Parrain, Org. Biomol. Chem., 7, 425, **2009**
15. C. Huang and B. Liu, Chem. Commun., 46, 5280, **2010**
16. M. Yoshida, Y. Shoji, K. Shishido, Org. Lett., 11, 1441, **2009**
17. J. S. Yadav, B. Thirupathaiah, P. Srihari, Tetrahedron, 66, 2005, **2010**
18. G. R. Pettit, Y. Meng, R. K. Pettit, D. L. Herald, F. Hogan, Z. A. Cichacz, Bioorg. Med. Chem., 18, 4879, **2010**
19. T. Nakamura, K. Tsuboi, M. Oshida, T. Nomura, A. Nakazaki, S. Kobayashi, Tetrah. Lett., 50, 2835, **2009**
20. S. Rosselli, A. Maggio, R. A. Raccuglia, M. Bruno, Eur. J. Org. Chem., 2010, 3093, **2010**
21. A. Bodensieck, W. M. F. Fabian, O. Kunert, F. Belaj, S. Jahangir, W. S. Hly, R. Bauer, Chirality, 22, 308, **2010**
22. T. Magauer, J. Mulzer, K. Tiefenbacher, Org. Lett, 11, 5306, **2009**
23. A. Srikrishna, G. Ravi, D. R. C. V. Subbaiah, Synlett, 1, 32, **2009**
24. Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT, **2009**
25. J. H. G. Lago, A. Oliveira, E. F. Guimarães, M. J. Kato, J. Braz. Chem. Soc., 18, 638, **2007**
26. P. Joseph-Nathan, S. G. Leitão, S. C. Pinto, G. G. Leitão, H. R. Bizzo, F. L. P. Costa, M. B. de Amorim, N. Martinez, E. Dellacassa, A. Hernández-Barragán, N. Pérez-Hernández, Tetrah. Lett., 51, 1963, **2010**
27. F. M. dos Santos Jr., L. S. M. Velozo, E. M. de Carvalho, A. M. Marques, R. M. Borges, A. P. F. Trindade, M. I. S. dos Santos, A. C. F. de Albuquerque, F. L. P. Costa, M. A. C. Kaplan, M. B. de Amorim, Molecules, 18, 13520, **2013**
28. J. A. Marco, J. F. Sanz-Cervera, M. D. Morante, V. García-Lliso, J. Vallés-Xirau, J. Jakupovic, Phytochem., 41, 837, **1996**
29. J. Bhattacharyya and E. V. L. Cunha, Phytochem., 31, 2546, **1992**

Ana Carolina Ferreira de Albuquerque[a*], Daniel Joras Ribeiro[a] & Mauro Barbosa de Amorim[a]

[a]Instituto de Pesquisas de Produtos Naturais – Universidade Federal do Rio de Janeiro

[*]E-mail: maurobdeamorim@gmail.com