SHORT COMMUNICATION

Theoretical study *via* DFT for the prediction of ^{13}C- and ^{1}H-NMR data for two diterpenoids derived from the root of *Salvia grandifolia*

RENATO A. da COSTA[1,2]*, SEBASTIÃO G. SILVA[1], SILVANA de O. SILVA[1], JORDDY N. CRUZ[3], WANESSA A. da COSTA[5], LUCIANE S. N. dos SANTOS BRASIL[6], RAÍ C. SILVA[4], CLEYDSON B. R. SANTOS[4], CLÁUDIO N. ALVES[1] and DAVI de S. BARROS BRASIL[1]

[1]*Program of Post-Graduation in Chemistry, Federal University of Pará, Rua Augusto Corrêa S/N, Guamá, 66075-900 Belém, Pará, Brazil,* [2]*Federal Institute of Pará, Rodovia PA 275, s/n – União, CEP 68515-000, Parauapebas, Pará, Brazil,* [3]*Laboratory of Preparation and Computation of Nanomaterials, Federal University of Pará, C. P. 479, 66075-110, Belém, PA, Brazil,* [4]*Laboratory of Modeling and Computational Chemistry, Federal University of Amapá, Macapá 68902-280, Amapá, Brazil;* [5]*Program of Post-Graduation in Natural Resources Engineering (PRODERNA/ITEC), Federal University of Para, Rua Augusto Corrêa S/N, Guamá, 66075-900 Belém, Pará, Brazil and* [6]*Department of Food Technology Center for Natural Sciences and Technology (CCNT) State University of Pará, Travessa Angustura, 2219, Pedreira, 66087-310, Belém, Pará, Brazil*

Abstract: The DFT method has become a promising alternative in the support of traditional NMR experimental techniques, comparing experimental data with theoretical data, thereby achieving accurate and satisfactory results. In the present study, the experimental data of two diterpenes were compared to the theoretical data obtained by the GIAO method, applying DFT at the B3LYP/cc-pVDZ and B3PW91/DGDZVP levels to verify the degree of correlation, significance and predictability of the models obtained with the purpose of proving which of the computational methods is the most efficient for this class of substances. The theoretical NMR values obtained at B3LYP/cc-pVDZ were the more satisfactory, showing better linear correlations, presenting greater degrees of adjustments, significance and predictability compared to the B3PW91/DGDZVP method. Thus, these values were used to define H_α and H_β 1, 2, 3, 6 and 7 for diterpene **1** and 1, 2, 3, 6 and 7 for diterpene **2**, which was not possible using experimental data. This showed that the quantum method used could help in the structural elucidation of natural products.

Keywords: statistical analysis; B3LYP/cc-pVDZ; linear regression.

591

INTRODUCTION

The density functional theory (DFT) method allows the calculation of useful chemical properties, and structural and electronic analysis of molecules with a high degree of precision when compared with experimental data.[1,2] This method is also used to determine the lowest energy point, data of vibration modes (IR, infrared) and theoretical values of chemical shifts of ^{13}C- and ^{1}H-NMR. Recently, comparing experimental NMR data with theoretical data has become common practice. Papers are found in the literature comparing the experimental data of NMR with theoretical calculations, achieving accurate and satisfactory results.[3–11] The DFT methodology could be applied as a promising alternative in support of traditional NMR experimental techniques.

In this sense, the present study contributes to the state of the art regarding the use of molecular modeling techniques in the spectral identification of natural products *via* NMR. The quantum calculations were performed at the GIAO--DFT/B3LYP/cc-pVDZ and B3PW91/DGDZVP levels to obtain chemical shifts (δ_H and δ_C) data of two diterpenes isolated by Kang *et al.*,[12] grandifolia A (**1**) and grandifolia B (**2**, Fig. 1). The theoretical values of δ_H and δ_C obtained were compared with the experimental ones using simple linear regression, in order to verify the degree of correlation, significance and predictability of the obtained models. The main objective was to prove which of the computational methods is more efficient for this class of substances, and to define Hα and Hβ 1, 2, 3, 6 and 7 for grandifolia A and 1, 2, 3, 6 and 7 for grandifolia B, which was not possible using the experimental data, thus showing that the quantum method used could help in the structural elucidation of natural products.
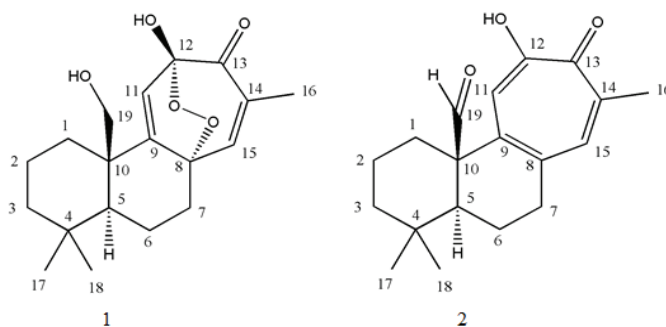

Fig. 1. Structures of compounds **1** and **2**.

EXPERIMENTAL

Computational method

The molecular structures were optimized by the DFT method at the B3LYP/cc-pVDZ and B3PW91/DGDZVP levels, by adjusting such molecules to a minimum local energy conformation. The minimum energy structures are guaranteed by the absence of any imaginary

frequency. With all optimized molecules, the theoretical ^{13}C- and ^{1}H-NMR data were obtained by gauge-independent atomic orbital (GIAO)[13] by applying the B3LYP/cc-pVDZ and B3PW91/DGDZVP levels to all atoms, all being under the effect of the solvent, pyridine and chloroform for **1** and **2**, respectively, by the polarizable continuum model (PCM)[14,15]. To determine the TMS (internal standard) shielding constants, the same protocol of diterpenes was strictly followed. All calculations were performed using the software Gaussian 09.[16]

Statistical analysis

The statistical analysis of this study was based on the paper of Souza *et al.*,[6] in which DFT calculations (B3LYP/cc-pVDZ) were used to determine the ^{13}C- and ^{1}H-NMR chemical shifts for three labdanum diterpenes: ent-13-epi-manoyl oxide, ent-3-oxo-13-epi-manoyl oxide (ribenone) and ribenol. The theoretical NMR data were compared to the experimental ones using linear regression and important statistical parameters such as degree of adjustment (R^2), standard deviation (s), degree of significance (F) and predictability (Q^2 and S_{PRESS}). They were obtained with the software MINITAB (free 390 version, Minitab Inc., State College, PA, USA). Linear equations were obtained for the prediction of the NMR chemical shifts values. The statistical analysis indicated good correlation between the experimental NMR results and the theoretical data, with a high predictive power.

Other statistical parameters were determined to improve the analysis of the obtained theoretical NMR data, such as coefficients a and b of the linear regression $\delta_{calcd} = a + b(\delta_{exptl})$; mean absolute error (*MAE*), defined as $MAE = \Sigma n|\delta_{calcd} - \delta_{exptl}|/n$, corrected mean absolute error (*CMAE*), defined by $C_{MAE} = \Sigma n|\delta_{corr} - \delta_{exptl}|/n$, where $\delta_{corr} = (\delta_{calcd} - a)/b$.[17] In addition, the linear obtained equations were tested for their adjustment, significance and predictability using a cross-validation procedure. Cross-validation is a practical and reliable method to verify its predictive power. In the leave-one-out cross validation, several models are developed with one sample omitted at each step. After developing each model, the omitted data is predicted and the differences between true and predicted values are calculated. The sum of squares of these differences is calculated and finally the performance of the model (its predictive capacity) can be given by the predicted residual error sum of squares (*PRESS*) and standard deviation of cross-validation (S_{PRESS}), as shown in Eqs. (1) and (2):[18–20]

$$PRESS = \sum_{i=1}^{n}(y_i - y)^2 \qquad (1)$$

$$S_{PRESS} = \frac{\sqrt{PRESS}}{n-k-1} \qquad (2)$$

where y_i is the experimental value, y is the predicted value, n is the number of samples used to construct the model and k is the number of NMR parameters. The predictive capacity of the model can also be quantified in terms of Q^2, which is defined as:[18,19]

$$Q^2 = 1.0 - \frac{PRESS}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \qquad (3)$$

where $\overline{y} = y_{mean}$.

RESULTS AND DISCUSSION

The B3LYp/cc-pVDZ computational method proved to be the most efficient in predicting the chemical shifts of the ^{13}C- and ^{1}H-NMR spectra for diterpenes **1** and **2**, exhibiting the smallest parameters of *MAE* and *CMAE* and the highest linear correction coefficients $R^2 > 98$ %. According to the studies of Cimino *et al.*,[17] the most efficient computational method is the one with the lowest values of *MAE* and *CMAE*. This is illustrated in Fig. 2 and is corroborated by the results in Table I.
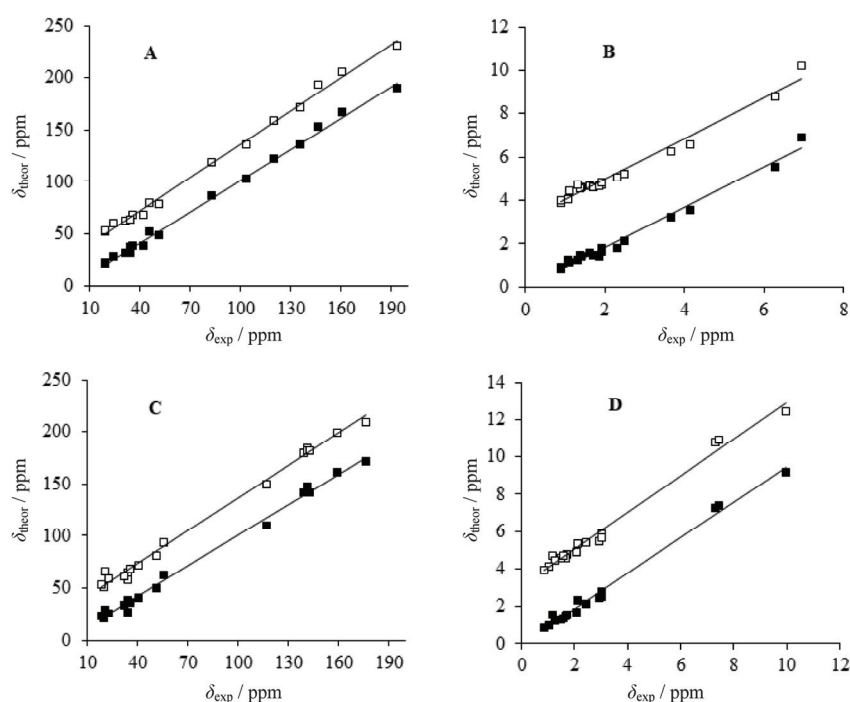

Fig. 2. Correlation between the experimental and calculated chemical shifts of (A) the ^{13}C and (B) ^{1}H of diterpene **1**, and (C) the ^{13}C and (D) ^{1}H of diterpene **2**. Methods: B3LYP/cc-pVDZ (■) and B3PW91/DGDZVP (□). The effect of the solvent, pyridine and chloroform to diterpene **1** and **2**, respectively, by the PCM method.

The most satisfactory results after the cross-validation procedure calculated for diterpenes **1** and **2** were also at B3LYP/cc-pVDZ level ($s = 3.47$; $S_{PRESS} = 0.99$; $F = 4460.74$; $Q^2 = 99.48$ % for δ_C and $s = 0.23$; $S_{PRESS} = 0.08$; $F = 883.49$; $Q^2 = 96.45$ % for δ_H) for diterpene **1**, ($s = 4.35$; $S_{PRESS} = 1.17$; $F = 3547.90$; $Q^2 = 99.41$ % for δ_C and $s = 0.26$; $S_{PRESS} = 0.08$; $F = 1520.57$; $Q^2 = 98.59$ % for δ_H) for diterpene **2**, when compared to the method B3PW91/DGDZVP (See Table I). It is clear after evaluation of the statistical data, the regression model for the chemical shifts of the ^{13}C- and ^{1}H-NMR results of the structures **1** and **2** using

the B3LYP/cc-pVDZ method have the best settings, as well as a higher degree of significance and predictability compared to the B3PW91 / DGDZVP method (Table I). This analysis indicates that there is a linear agreement between the obtained theoretical chemical displacement data of the ^{1}H- and ^{13}C-NMR results in relation to the experimental values for the structures under study, since they are quantitatively similar to the chemical displacements mentioned in the literature, showing that the theoretical method B3LYP/cc-pVDZ is more efficient. According to Gaudio and Zandonade,[18] the greater R^2, the smaller S, the greater F, the Q^2 and S_{PRESS} closer to zero, the more satisfactory method.

TABLE I. Correlations and parameters[a] of the linear fit for the NMR properties calculated for diterpenes **1** and **2**; parameters related to linear $\delta_{calc} = a + b(\delta_{exp})$

| Parameter | B3LYP/cc-pVDZ | | | | B3PW91/DGDZVP | | | |
| | δ / ppm | | | | | | | |
	^{13}C **1**	^{13}C **2**	^{1}H **1**	^{1}H **2**	^{13}C **1**	^{13}C **2**	^{1}H **1**	^{1}H **2**
a	0.99	0.96	0.92	0.95	1.06	1.04	0.95	0.98
b	2.29	3.84	–0.22	–0.04	0.30	2.15	0.10	0.06
MAE[a]	3.36	3.96	0.23	0.27	5.58	6.19	0.20	0.25
$CMAE$[b]	2.76	3.41	0.19	0.22	3.17	3.50	0.20	0.25
S	3.47	4.35	0.23	0.26	4.13	4.97	0.26	0.31
R^2	99.62	99.52	98.23	99.02	99.53	99.46	97.84	98.69
Q^2	99.48	99.41	96.45	98.59	99.38	99.33	95.44	97.81
F	4460.74	3547.9	883.49	1520.57	3605.75	3143.59	656.71	1124.03
$PRESS$	283.81	395.93	1.65	1.44	384.24	522.33	2.25	2.41
S_{PRESS}	0.99	1.17	0.08	0.08	1.15	1.34	0.09	0.10

[a]Mean absolute error: $MAE = \Sigma n|\delta_{calc}-\delta_{exp}|/n$; [b]corrected mean absolute error: $CMAE = \Sigma n|\delta_{corr}-\delta_{exp}|/n$

Experimental and theoretical chemical shifts of ^{13}C- and ^{1}H-NMR[12] peaks at B3LYP/cc-pVD and B3PW91/DGDZVP levels for diterpenes **1** and **2** are given in Tables S-I and S-II of the Supplementary material to this paper, as well as the residues (*RS*) in ppm, for each carbon atom and each hydrogen atom present in the structures, and the values predicted by the methods used. Analysis of the results shown in Tables I and S-I for the calculated δ values for ^{13}C and ^{1}H at the B3LYP/cc-pVDDZ level for diterpenes **1** and **2** indicates that this computational method showed more satisfactory results, since the theoretical data were closer to the experimental ones, presenting low residual values (difference between experimental and theoretical chemical shifts) It is expected that an appropriate method would have its residues randomly scattered around zero.[18] Thus, the theoretical method B3LYP/cc-pVDZ becomes an auxiliary tool to experimental NMR techniques for the elucidation of natural products.

As the results of the statistical analysis confirmed the relevance of the obtained theoretical data and showed that the B3LYP/cc-pVDZ method is more effective, this method was used to confirm the chemical shifts for the non-equi-

valent Hα and Hβ hydrogens at positions 1, 2, 3, 6 and 7 for diterpene **1** and 1, 2, 3, 6 and 7 for diterpene **2**, since the NMR experimental data did not allow this definition because they are diastereotopic hydrogens.

Typically, diastereotopic hydrogens exhibit similar chemical shifts, and sometimes non-equivalence is not observed in the ^{1}H-NMR spectrum. This occurs especially when the diastereotopic hydrogens are away from the stereogenic center. All of the methylene hydrogens of structure **1** and **2** are diastereotopic.

The computational simulation of the ^{1}H-NMR spectra allowed the identification of the non-equivalent hydrogens presenting theoretical values relatively close to the experimental ones, as shown in Table II, which corroborates with works of the literature that used computational simulation to define Hα and Hβ protons of diterpenes that were not elucidated experimentally.[6,9]

TABLE II. Theoretical chemical shifts of the diastereotopic protons of substances **1** and **2**.

Diterpene	1α	1β	2α	2β	3α	3β	6α	6β	7α	7β
1	1.22	2.14	1.58	1.42	1.14	1.44	1.61	1.48	1.42	1.82
2	2.42	1.57	1.47	1.57	1.26	1.31	1.68	2.31	2.50	2.75

CONCLUSIONS

The theoretical NMR values obtained using the B3LYP/cc-pVDZ method were more satisfactory, showed better linear correlations, and presented a higher degree of adjustments, degree of significance and predictability compared to the B3PW91/DGDZVP method. The theoretical values at the B3LYP/cc-pVDZ level were also the closest to the experimental ones, presenting low residue values. In this way, the use of the B3LYP/cc-pVDZ method in diterpenes (that have not been fully elucidated by traditional experimental techniques) similar to those studied in this work is corroborated.

This work contributed to the proposition of chemical shifts of diastereotopic ^{1}H that were not identified at the experimental level. Due to their very similar values, the experimental observation is often impaired, which shows the importance of the use of computational methods to aid the experimental tool.

SUPPLEMENTARY MATERIAL

Additional data are available electronically from http://www.shd.org.rs/JSCS/, or from the corresponding author upon request.

И З В О Д

ТЕОРИЈСКО DFT ПРОУЧАВАЊЕ ПРЕДВИЂАЊА ^{13}C- И ^{1}H-NMR ПОДАТАКА ЗА ДВА ДИТЕРПЕНОИДА ИЗ КОРЕНА БИЉКЕ *Salvia grandifolia*

RENATO A. DA COSTA[1,2], SEBASTIÃO G. SILVA[1], SILVANA DE O. SILVA[1], JORDDY N. CRUZ[3], WANESSA A. DA COSTA[5], LUCIANE S. N. S. BRASIL[6], RAÍ C. SILVA[4], CLEYDSON B. R. SANTOS[4], CLÁUDIO N. ALVES[1] и DAVI S. B. BRASIL[1]

[1]*Program of Post-Graduation in Chemistry, Federal University of Pará, Rua Augusto Corrêa S/N, Guamá, 66075-900 Belém, Pará, Brazil,* [2]*Federal Institute of Pará, Rodovia PA 275, s/n - União. CEP 68515-000, Parauapebas, Pará, Brazil,* [3]*Laboratory of Preparation and Computation of Nanomaterials, Federal University of Pará, C. P. 479, 66075-110, Belém, PA, Brazil,* [4]*Laboratory of Modeling and Computational Chemistry, Federal University of Amapá, Macapá 68902-280, Amapá, Brazil,* [5]*Program of Post-Graduation in Natural Resources Engineering (PRODERNA/ITEC), Federal University of Para, Rua Augusto Corrêa S/N, Guamá, 66075-900 Belém, Pará, Brazil и* [6]*Department of Food Technology Center for Natural Sciences and Technology (CCNT) State University of Pará, Travessa Angustura, 2219, Pedreira, 66087-310, Belém, Pará, Brazil*

DFT метод је постао обећавајућа алтернатива као помоћ традиционалним NMR експерименталним техникама, поређењем експерименталних и теоријских података, постижући поуздане и задовољавајуће резултате. У овој студији су експериментални подаци за два дитерпена поређени са теоријским подацима добијеним GIAO методом, примењујући DFT на B3LYP/cc-pVDZ и B3PW91/DGDZVP нивоима за потврђивање степена корелације, значајности и моћи предвиђања модела добијених у сврху доказивања који је рачунарски модел најефикаснији за ову класу једињења. Теоријске NMR вредности добијене са B3LYP/cc-pVDZ су боље задовољавале: показивале су бољу линеарну корелацију, представљале већи степен слагања, значајности и моћи предвиђања у поређењу са B3PW91/DGDZVP методом и биле су коришћене за дефинисање Hα и Hβ на 1, 2, 3, 6 и 7 за дитерпен **1** и на 1, 2, 3, 6 и 7 за дитерпен **2**, што није било могуће коришћењем експерименталних података, показујући да примењени квантно хемијски модел може помоћи у разјашњавању структура природних производа.

(Примљено 27. септембра, прихваћено 1. октобра 2018)

REFERENCES

1. C. Lee, W. Yang, R. G. Parr, *Phys. Rev., B* **37** (1988) 785 (http://dx.doi.org/10.1103/PhysRevB.37.785)
2. E. J. Barreiro, C. R. Rodrigues, M. G. Albuquerque, C. M. R. de Sant'Anna, R. B. de Alencastro, *Química Nova* **20** (1997) 300 (http://dx.doi.org/10.1590/S010040421997000300011)
3. R. A. Da Costa, F. C. D. A. Nascimento, S. G. Silva, A. D. S. S. Silva, D. D. S. B. Brasil, *J. Eng. Technol. Ind. Appl. (ITEGAM-JETIA)* **3** (2017) 1 (http://dx.doi.org/10.5935/2447-0228.20170003)
4. D. Xin, C. A. Sader, U. Fischer, K. Wagner, P.-J. Jones, M. Xing, K. R. Fandrick, N. C. Gonnella, *Org. Biomol. Chem.* **15** (2017) 928 (http://dx.doi.org/10.1039/C6OB02450E)
5. D. Xin, C. A. Sader, O. Chaudhary, P. J. Jones, K. Wagner, C. S. Tautermann, Z. Yang, C. A. Busacca, R. A. Saraceno, K. R. Fandrick, N. C. Gonnella, K. Horspool, G. Hansen, C. H. Senanayake, *J. Org. Chem.* **82** (2017) 5135–5145 (http://dx.doi.org/10.1021/acs.joc.7b00321)
6. F. S. de Souza, S. de O. Silva, L. J. G. de Faria, C. N. Alves, A. H. Muller, G. M. S. P. Guilhon, D. do S. B. Brasil, *Química Nova* **38** (2015) 645 (http://dx.doi.org/10.5935/0100-4042.20150057)

7. S. de O. Silva, R. N. S. Peixoto, J. R. A. Silva, C. N. Alves, G. M. S. P. Guilhon, L. S. Santos, D. D. S. B. Brasil, *Int. J. Mol. Sci.* **12** (2011) 9389 (http://dx.doi.org/10.3390/ijms12129389)
8. S. deO. Silva, M. C. Corrêa, H. Bitencourt, W. R. Monteiro, J. Lameira, L. Santos, G. M. P. Guilhon, D. B. Brasil, *J. Comput. Theor. Nanosci.* **9** (2012) 953 (http://dx.doi.org/10.1166/jctn.2012.2123)
9. D. S. B. Brasil, A. H. Müller, G. M. S. P. Guilhon, C. N. Alves, G. Peris, R. Llusar, V. Moliner, *J. Brazilian Chem. Soc.* **21** (2010) 731 (http://dx.doi.org/10.1590/S0103-50532010000400021)
10. D. S. B. Brasil, C. N. Alves, G. M. S. P. Guilhon, A. H. Muller, R. de S. Secco, G. Peris, R. Llusar, *Int. J. Quantum Chem.* **108** (2008) 2564 (http://dx.doi.org/10.1002/qua.21673)
11. D. S. B. Brasil, R. Y. O. Moreira, A. H. Müller, C. N. Alves, *Int. J. Quantum Chem.* **106** (2006) 2706 (http://dx.doi.org/10.1002/qua.21031)
12. J. Kang, L. Li, D. Wang, H. Wang, C. Liu, B. Li, Y. Yan, L. Fang, G. Du, R. Chen, *Phytochemistry* **116** (2015) 337 (http://dx.doi.org/10.1016/j.phytochem.2015.03.014).
13. J. R. Cheeseman, G. W. Trucks, T. A. Keith, M. J. Frisch, *J. Chem. Phys.* **104** (1996) 5497–5509 (http://dx.doi.org/10.1063/1.471789)
14. S. Miertuš, E. Scrocco, J. Tomasi, *Chemical Physics* **55** (1981) 117 (http://dx.doi.org/10.1016/0301-0104(81)85090-2)
15. S. Miertuš, J. Tomasi, *Chem. Phys.* **65** (1982) 239 (http://dx.doi.org/10.1016/0301-0104(82)85072-6)
16. *Gaussian 16, Revision B.01*, Gaussian, Inc., Wallingford, CT, 2016
17. P. Cimino, L. Gomez-Paloma, D. Duca, R. Riccio, G. Bifulco, *Magn. Reson. Chem.* **42** (2004) 26 (http://dx.doi.org/10.1002/mrc.1410)
18. A. C. Gaudio, E. Zandonade, *Quimica Nova* **24** (2001) 658 (http://dx.doi.org/10.1590/S0100-40422001000500013)
19. C. B. R. Santos, J. B. Vieira, C. C. Lobato, L. I. S. Hage-Melim, R. N. P. Souto, C. S. Lima, E. V. M. Costa, D. S. B. Brasil, W. J. C. Macêdo, J. C. T. Carvalho, *Molecules* **19** (2014) 367 (http://dx.doi.org/10.3390/molecules19010367)
20. J. Costa, C. Santos, K. Costa, R. Ramos, C. H. Silva, W. Macêdo, *Química Nova* **42(7)** (2018) 732 (http://dx.doi.org/10.21577/0100-4042.20170239).