Science & Technology Development Journal 2024, 27(1):3339-3346

Open Access Full Text Article

Natural Sciences: Research article

Impacts of basis sets, solvent models, and NMR methods on the accuracy of ^{1}H and ^{13}C chemical shift calculations for biaryls: a DFT study

Thien T. Nguyen[1,2,*]

[1]*Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam*

[2]*Faculty of Natural Science, Duy Tan University, Da Nang 550000, Vietnam*

Correspondence

Thien T. Nguyen, Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

Faculty of Natural Science, Duy Tan University, Da Nang 550000, Vietnam

Email: nguyentrongthien86@gmail.com

History
- Received: 2023-09-21
- Accepted: 2024-03-19
- Published Online: 2024-03-31

DOI :

https://doi.org/10.32508/stdj.v27i1.4182

Check for updates

Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

VNUHCM PRESS

ABSTRACT

Introduction: Biaryls are core structures composed of chiral ligands, organocatalysts, biologically active natural products and biopolymer lignins. In this study, the effects of basis sets, solvent models, and NMR methods on the accuracy of ^{1}H/^{13}C NMR chemical shift calculations for biaryl structures were evaluated. **Methods**: All calculations were performed using Gaussian09. The GIAO NMR results were observed and extracted using GaussView05. To reduce the systematic error of the calculations, linear regression analysis of the calculated chemical shifts versus the experimental shifts was performed. **Results**: The tested basis sets showed good ^{1}H/^{13}C results, with CMAE values as low as 0.0425 ppm and 1.09 ppm for ^{1}H and ^{13}C, respectively. The use of solvent models significantly increased the accuracy of the ^{1}H chemical shift calculations. The GIAO method produced more accurate results than did the IGAIM and CSGT methods. **Conclusion**: This study recommends 6-31G(d,p) and DGDZVP basis sets, IEMPCM and CPCM solvent models, and GIAO NMR methods for the accurate prediction of ^{1}H and ^{13}C chemical shifts for biaryls, assisting in their full structural assignments.

Key words: accuracy, lignin model dimer, NMR, chemical shift, DFT

INTRODUCTION

Biaryls are important core structures present in useful chiral ligands, organocatalysts, biologically active natural products, and biopolymer lignins.[1,2] Typical examples illustrated in Figure 1A are C_2-symmetric binaphthyls (BINOL and BINAPs), which catalyze numerous asymmetric transformations[3,4] the mycotoxin viriditoxin[5] the alkaloid bismurrayaquinone A[6] and 5-5/4-O- lignin substructures[7]. These compounds also possess biaryl linkages, which can give rise to atropisomers that have received significant attention from the synthetic community in the last decade.[1,2] An accurate method for predicting NMR spectra would contribute valuable insights into the conformations of biaryl structures and the local electron environment of each NMR active nucleus. Gauge-independent atomic orbital (GIAO)-DFT NMR calculations have effectively supported the structural assignment and validation of biaryl compounds with accurate predictions at affordable computational costs.[6,8] In general, the accuracy is impacted by optimized geometries, density functional methods, basis sets, solvation models, and NMR methods.[7–9] For the two common nuclei of organic molecules, ^{1}H shift predictions are more challenging than ^{13}C shift predictions due to the significant impact of solvation effects on protons.

Previous studies reported how the use of different density functional methods and basis sets for NMR calculations affected the ^{1}H/^{13}C results for a variety of different organic structures.[10–12] In 2015, Toomsalu reported the use of 18 DFT functionals and 6 basis sets for ^{1}H and ^{13}C calculations of small organic molecules and reported that the best functional/basis set for ^{13}C was PBE1PBE/aug-cc-pVDZ, and those for ^{1}H were HSEH1PBE, mPW1PW91, PBE1PBE, CAM-B3LYP, and B3PW91 functionals and cc-pVTZ for ^{1}H. In 2017, Iron recommended LC-TPSSTPSS/cc-pVTZ among an extensive list of tested functionals and basis sets for ^{13}C predictions. In our continuing interest in the NMR modeling of biaryls, we have recently reported the impact of density functional methods on the accuracy of ^{1}H/^{13}C chemical shift calculations for biaryls.[5,9] Herein, the present study shows how basis sets, solvent models, and NMR methods influence the accuracy of ^{1}H and ^{13}C NMR shift calculations for biaryl **1** (Figure 1B).

COMPUTATIONAL METHODS

All calculations were performed using Gaussian09[13] on a commercial computer with an Intel Core i3-7100

Cite this article : Nguyen T T. **Impacts of basis sets, solvent models, and NMR methods on the accuracy of ^{1}H and ^{13}C chemical shift calculations for biaryls: a DFT study**. *Sci. Tech. Dev. J.* 2024; 27(1):3339-3346.

processor. Geometry optimizations were performed at CAMB3LYP/6-31G(d,p) with default convergence criteria. The integral equation formalism variant of the polarized continuum model (IEFPCM) was incorporated during geometry optimization.[14] Subsequent frequency calculations ensured that a potential energy surface (PES) local minimum was attained during energy minimization.

The following basis sets, solvent models, and NMR methods, which are commonly used for determining $^1H/^{13}C$ NMR chemical shifts, were evaluated:

- Basis sets: Pople's 6-31G, 6-31G(d,p), 6-31G(3d,p), 6-31G(d,3p), 6-31G+(d,p), 6-31G++(d,p), and 6-311G(d,p);[15] Dunning's cc-pVDZ correlation consistent basis set;[16] and DGDZVP.[17]
- Two-solvent models: IEFPCM and CPCM.
- Three NMR methods: GIAO, IGAIM, and CSGT.

Unless specified otherwise, single-point NMR GIAO calculations were carried out at the IEFPCM (DMSO)/ωB97XD/6-31G(d,p) level of theory, which was found to produce computed 1H/13C chemical shifts with high accuracy. The GIAO NMR results were observed and extracted using GaussView05. Each optimized structure was used for computing the corresponding isotropic shielding constants (σ_{cal}). The chemical shifts (δ_{cal}) were obtained using Equation 1. For both the 1H and 13C NMR calculations, an average of the values of equivalent atoms was assumed. For example, a single proton/carbon signal is observed for the two methoxy groups of dimer 1 due to fast rotations of the biaryl linkage and two methyl groups relative to the NMR measurement time scale. To reduce the systematic error of the calculations, linear regression analysis of the calculated chemical shifts versus the experimental shifts (δ_{exp}) (Equation 2)) was performed, and the scaled chemical shifts (δ_{scal}) were computed according to Equation 3. Linear regression was used based on the fitness of the calculated data. As a reference had a negligible impact on the linear regression analysis, fixed values of 197 ppm and 31 ppm were chosen as the TMS shielding constants for 13C and 1H, respectively. The computed results were evaluated using the mean absolute value ($|\Delta\delta|$/ppm, Equation 4), corrected mean absolute error (CMAE/ppm, Equation 5), corrected root mean squared error (CRMSE/ppm, Equation 6), and Pearson correlation coefficient (r^2). Smaller values of CMAE and CRMSE indicate smaller errors, and a larger value of r^2 indicates a stronger correlation between the theoretical and experimental data. Error calculations and linear correlations were performed using Microsoft Excel 2013.

$$\delta_{cal} = \sigma_{TMS} - \sigma_{cal} \qquad (1)$$

$$\delta_{cal} = a\delta_{exp} + b \qquad (2)$$

$$\delta_{scal} = \frac{\delta_{cal} - b}{a} \qquad (3)$$

$$|\triangle\delta| = |\delta_{scal} - \delta_{exp}| \qquad (4)$$

$$CMAE = \Sigma_1^n(|\delta_{scal} - \delta_{exp}|)^2/n \qquad (5)$$

$$CRMSE = \sqrt{\Sigma_1^n(\delta_{scal} - \delta_{exp})^2/n} \qquad (6)$$

Figure 1 shows the numbered dimers used for the proton and carbon atoms in this study. Due to the axial symmetry of biaryl **1**, only one side of the structure was labeled. Compound **1** contains phenolic and carboxylic protons, which typically do not appear in the NMR spectra due to rapid exchanges in DMSO-d_6 or $CDCl_3$. Therefore, these protons were excluded from the calculations in this study. The experimental ^1H and ^{13}C NMR spectra of **1**[9,18] were reported.

RESULTS

Impact of basis sets

9 Basis sets, including 6-31G, 6-31G(d,p), 6-31G(3d,p), 6-31G(d,3p), 6-31G+(d,p), 6-31G++(3d,p), 6-311(d,p), cc-pVDZ, and DGDZVP, were coupled with *w*B97XD and the IEFPCM solvent model (DMSO) for the NMR calculations of compound 1, optimized at the IEFPCM(DMSO)/CAM-B3LYP/6-31G(d,p) level of theory. The calculated $^1H/^{13}C$ shifts, statistical parameters, and absolute deviations are shown in Table 1, Table 2, and Figure 2, respectively. In general, the performance of basis sets depends on the main basis set, diffusion functions, and polarization functions.

Effects of solvent models and NMR methods

NMR calculations with no solvent and with two solvent models, IEFPCM and CPCM, were carried out, and the results are shown in Tables 3 and 4, and Figure 3. It should be expected that the impacts of solvent on protons were less than those on carbon nuclei due to the less crowded environment around protons.

NMR calculations were performed at the IEFPCM(DMSO)/ωB97XD/6-31G(d,p)//IEFPCM(DMSO)CAM-B3LYP/6-31G(d,p) level of theory using three NMR methods, and the results are summarized in Tables 3 and 5, and Figure 4.

Table 1: The $^{1}H/^{13}C$ chemical shifts were calculated using 9 basis sets (all values are in ppm).

	Exp.	6-31G	6-31G(d,p)	6-31G(3d,p)	6-31G(d,3p)	6-31G+(d,p)	6-31G++(d,p)	6-311G(d,p)	cc-pVDZ	DGDZVP
H1	6.78	6.97	6.83	6.74	6.81	6.88	6.86	6.86	6.86	6.92
H5	6.56	6.34	6.50	6.61	6.53	6.48	6.50	6.47	6.46	6.43
H7	2.74	2.60	2.67	2.64	2.67	2.69	2.69	2.66	2.70	2.42
H8	2.49	2.57	2.54	2.57	2.56	2.59	2.59	2.55	2.48	2.16
H10	3.79	3.87	3.82	3.81	3.78	3.71	3.71	3.82	3.86	3.52
C1	126.83	126.73	126.14	127.41	125.67	126.18	127.36	126.53	125.72	128.06
C2	123.62	125.44	125.23	124.69	124.86	126.24	126.14	124.41	125.40	122.77
C3	142.69	141.75	143.78	144.50	144.28	141.76	141.39	144.17	143.59	142.94
C4	148.59	146.80	147.92	147.19	147.89	147.81	147.84	148.34	148.59	146.86
C5	111.76	110.80	110.57	112.79	109.75	108.92	109.16	109.69	110.34	111.64
C6	131.94	131.58	132.25	129.62	132.12	133.30	132.67	133.01	133.34	131.01
C7	31.11	31.93	31.94	32.00	32.15	32.46	32.65	32.12	32.41	30.18
C8	36.66	39.15	38.24	37.58	38.62	37.84	37.32	38.45	38.52	38.88
C9	174.95	177.33	174.95	175.06	175.66	176.05	175.93	174.57	174.28	176.68
C10	56.82	53.43	53.88	54.02	53.87	54.34	54.50	53.61	52.71	55.93

Table 2: Accuracy evaluation of ^{1}H and ^{13}C chemical shift calculations using 9 basis sets

		$\delta(^{1}H)$			$\delta(^{13}C)$		
Entry	Basis set	r2	CMAE	CRMSE	r^2	CMAE	CRMSE
1	6-31G	0.9934	0.141	0.151	0.9985	1.50	1.80
2	6-31G(d,p)	**0.9992**	**0.0519**	**0.0535**	**0.9992**	**1.09**	**1.34**
3	6-31G(3d,p)	0.9988	0.0574	0.0646	0.9990	1.29	1.50
4	6-31G(d,3p)	**0.9993**	**0.0425**	**0.0491**	0.9989	1.35	1.55
5	6-31G+(d,p)	0.9978	0.0837	0.0860	0.9987	1.53	1.71
6	6-31G++(d,p)	0.9983	0.0745	0.0768	0.9989	1.39	1.59
7	6-311G(d,p)	0.9986	0.0664	0.0696	0.9990	1.24	1.52
8	cc-pVDZ	0.9987	0.0591	0.0672	0.9986	1.46	1.78
9	DGDZVP	0.9984	0.0568	0.0735	**0.9993**	**1.09**	**1.26**

*The highest accuracy data are in bold.

Table 3: ^{1}H/^{13}C chemical shifts calculated using different solvent models and NMR methods

	Exp.	No solvent	IEFPCM	CPCM	GIAO	IGAIM	CSGT
H5	6.56	6.41	6.50	6.51	6.50	6.47	6.47
H1	6.78	6.93	6.83	6.83	6.83	6.73	6.73
H7	2.74	2.66	2.67	2.67	2.67	2.51	2.51
H8	2.49	2.57	2.54	2.54	2.54	2.36	2.35
H10	3.79	3.79	3.82	3.82	3.82	4.30	4.30
C1	126.83	126.33	126.14	126.10	126.14	125.43	125.43
C2	123.62	125.16	125.23	125.18	125.23	123.24	123.23
C3	142.69	143.98	143.78	143.75	143.78	145.97	145.96
C4	148.59	148.13	147.92	147.92	147.92	149.45	149.44
C5	111.76	111.17	110.57	110.60	110.57	112.22	112.22
C6	131.94	132.38	132.25	132.27	132.25	128.94	128.93
C7	31.11	31.61	31.94	31.94	31.94	31.84	31.84
C8	36.66	38.68	38.24	38.24	38.24	37.53	37.52
C9	174.95	174.24	174.95	174.97	174.95	174.96	174.94
C10	56.82	53.27	53.88	53.89	53.88	55.40	55.40

Table 4: Accuracy evaluation of ^{1}H and ^{13}C chemical shift calculations using solvent models

		$\delta(^{1}H)$			$\delta(^{13}C)$		
Entry	Solvent model	r^2	CMAE	CRMSE	r^2	CMAE	CRMSE
1	No solvent model	0.9965	0.0940	0.110	0.9990	1.16	1.50
2	IEFPCM	0.9992	0.0519	0.0535	0.9992	1.09	1.34
3	CPCM	0.9992	0.0515	0.0531	0.9992	1.09	1.33

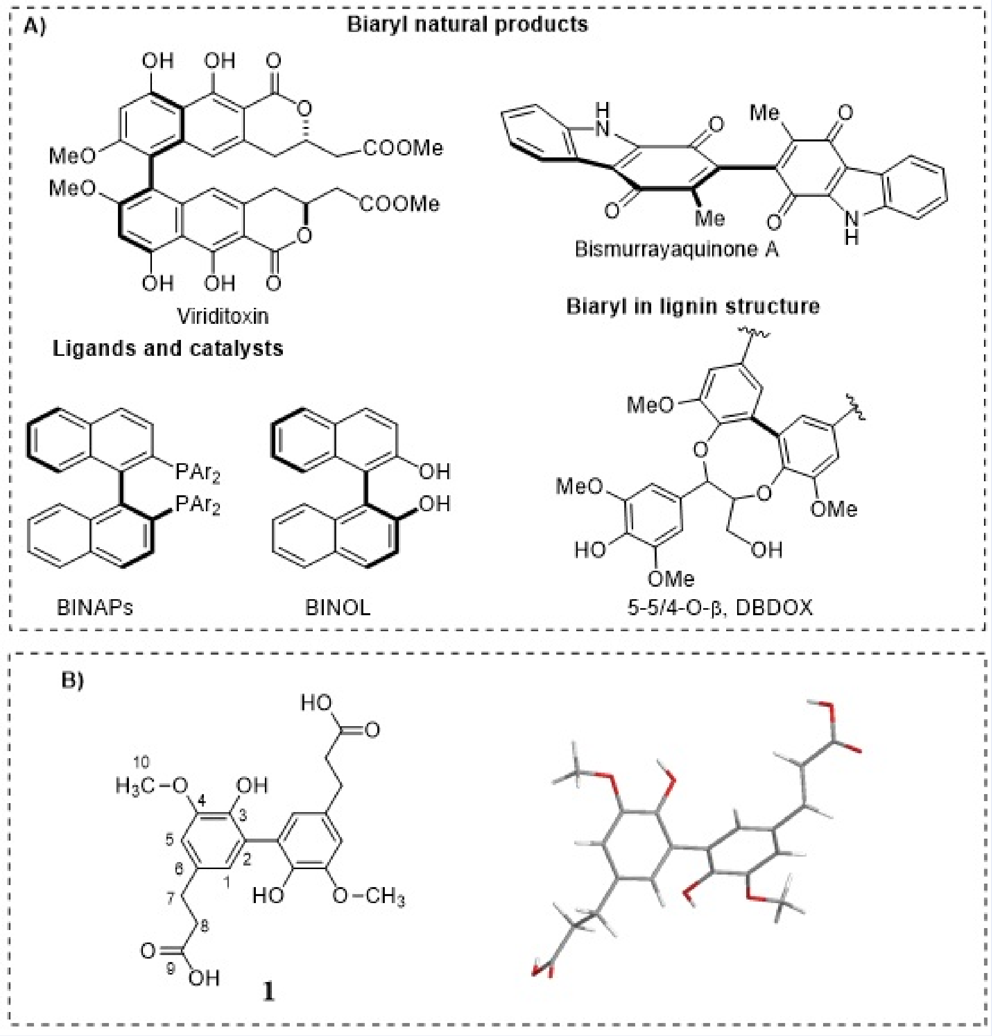

Figure 1: **A)** Biaryls in natural products, ligands, catalysts, and lignins; and **B)** Biaryl **1** with numbering labels and its optimized geometry at the level IEFPCM(DMSO)/CAM-B3LYP/6-31G(d,p) of theory.

Table 5: Accuracy evaluation of ^{1}H and ^{13}C chemical shift calculations using NMR methods

		$\delta(^{1}H)$			$\delta(^{13}C)$		
Entry	NMR method	r^2	CMAE	CRMSE	r^2	CMAE	CRMSE
1	GIAO	0.9992	0.0519	0.0535	0.9992	1.09	1.34
2	IGAIM	0.9806	0.203	0.260	0.9988	1.24	1.62
3	CSGT	0.9800	0.206	0.264	0.9988	1.24	1.62

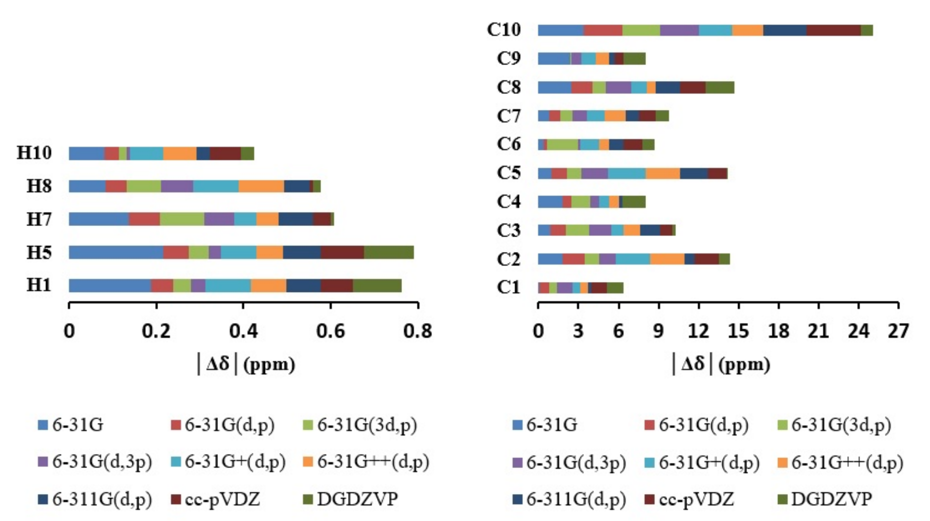

Figure 2: Absolute deviations (ppm) of the $^1H/^{13}C$ chemical shifts calculated using 9 basis sets

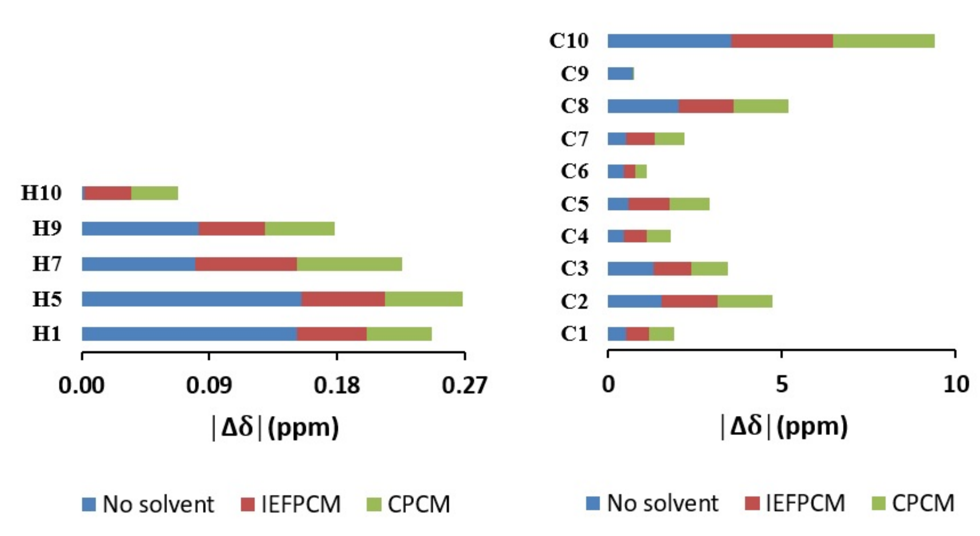

Figure 3: Mean absolute values (ppm) of $^1H/^{13}C$ calculations using two solvent models.

DISCUSSION

Table 2 shows that the lowest errors in the ^1H results were observed for 6-31G(d,p) (Entry 2, CMAE = 0.0519 ppm) and 6-31G(d,3p) (Entry 4, CMAE = 0.0425 ppm), which were slightly better than our previous calculation results[10]. While adding more polarization functions for hydrogen atoms to Pople's basis sets slightly increased the calculation accuracy, polarization functions for heavy atoms (Entry 3) or sets of diffusion functions (Entries 5 and 6) (Table 5, entries 2 and 3) had opposite effects and consumed much more computation time. All tested basis sets produced highly correlated results with $r^2 \geq 0.9934$, which would allow meaningful predictions of ^1H chemical shifts. In terms of the absolute deviations (Figure 2), most protons had relatively low errors. These deviations of protons **H7**, **H8**, and **H10** ($|\Delta\delta| \leq 0.14$ ppm), which are bound to sp^3-hybridized carbons, were smaller than those of protons **H1** and **H5** ($|\Delta\delta| \leq 0.22$ ppm), which are attached to sp^2-hybridized carbons (Figure 2). The difference in deviations for these two proton types could be explained

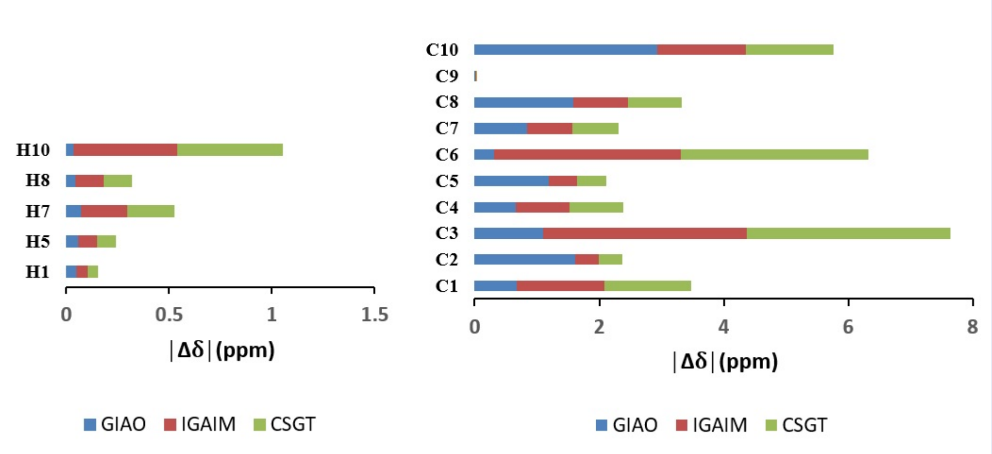

Figure 4: Mean absolute values (ppm) of ^{1}H/^{13}C calculations using three NMR methods

by relatively stronger solvent effects on aromatic protons. For the ^{13}C results, the two best performing basis sets were 6-31G(d,p) (Entry 2, CMAE = 1.09 ppm) and DGDZVP (Entry 9, CMAE = 1.09 ppm), which were well below the acceptable CMAE value of 6 ppm. High correlations with $r^2 \geq 0.9985$ were obtained for the tested basis sets. The computed ^{13}C chemical shifts had low deviations (Figure 2), except for methoxy carbon **C10**, which was consistently observed with relatively large deviations ($|\Delta\delta|$max = 4.11 ppm, cc-pVDZ). Overall, all tested basis sets showed good-to-excellent accuracy, and 6-31G(d,p) was the best basis set for both the ^{1}H and ^{13}C results. The use of either IEFPCM (Entry 2, CMAE = 0.0519 ppm) or CPCM (Entry 3, CMAE 0.0515 ppm) produced much better ^{1}H results than no use of a solvent model, while the ^{13}C results for these three methods yielded similar accuracies. These results could be explained by the fact that the high exposure of protons to solvent molecules is more obvious than that of carbon nuclei, which are well shielded. All protons had relatively close deviations (Figure 3), except for methoxy proton **H10**, which showed a low error when no solvent model was employed. For the computed ^{13}C chemical shifts, noticeable deviations of carbons **C2**, **C8**, and **C10** were consistently observed.

A significantly greater accuracy for ^{1}H results was obtained using the GIAO method than using the IGAIM and CSGT methods. The relatively low absolute deviations for the ^{1}H results obtained using GIAO are clearly observed in Figure 3. Compared to the ^{1}H calculations, the ^{13}C results were not strongly impacted by the three tested NMR methods. This observation was expected due to the relatively low impact of the solvent environment and molecular interactions on the carbon nuclei. The CMAE values ranged from 1.09 to 1.24. The ^{13}C results were obtained with high coefficients of determination ($0.9988 \leq r^2 \leq 0.9992$). Noticeable deviations were observed for methoxy proton **H10 and** carbon atoms **C3**, **C6**, and **C10** (Figure 4).

Overall, the above results relating to the tested basis sets, solvent models, and NMR methods indicated the importance of utilizing specific methods to obtain the desired accuracy of ^{1}H/^{13}C NMR calculations for biaryl **1**. High-accuracy results could be expected when applying these methods to compounds having this core biaryl structure.

CONCLUSION

The influence of 9 common basis sets, solvent models, and NMR methods on the accuracy of ^{1}H/^{13}C chemical shift calculations for biaryl **1** were evaluated. The tested basis sets showed good ^{1}H/^{13}C results, with CMAE values as low as 0.0425 ppm and 1.09 ppm for ^{1}H and ^{13}C, respectively. For the solvent models, the results indicated that solvent incorporation was necessary for improving the accuracy of the ^{1}H chemical shift calculations, while it had little effect on the computed ^{13}C chemical shifts. This is expected because carbon nuclei are less exposed to solvent molecules than to protons. The GIAO method outperformed the IGAIM and CSGT methods. This study highly recommends 6-31G(d,p) basis sets for the effective production of both ^{1}H/^{13}C with high accuracy and low computational cost, IEMPCM and CPCM solvent models for obtaining good ^{1}H results, and GIAO methods for

NMR calculations. This work will be useful for assisting in the full ^{1}H and ^{13}C NMR assignments of similar biaryls. In the near future, NMR calculations for biaryl natural products possessing interesting biological properties will be conducted.

ACKNOWLEDGMENT

This project was supported by the International Foundation for Science (IFS), Stockholm, Sweden, through a Grant No. 3-I-E-6576-1 to TTN.

REFERENCES

1. Nguyen TT. Traceless point-to-axial chirality exchange in the atropselective synthesis of biaryls/heterobiaryls. Org Biomol Chem. 2019;17:6952-6963;Available from: https://doi.org/10.1039/c9ob01304k.
2. Nguyen TT. The Use of Chiral ortho-Auxiliaries/Substituents and Remote Stereogenic Centers in Atropselective Biaryl Synthesis. Eur J Org Chem. 2020;Available from: https://doi.org/10.1002/ejoc.201901557.
3. Watts HD, Mohamed MNA, Kubicki JD. Comparison of multistandard and TMS-standard calculated NMR shifts for coniferyl alcohol and application of the multistandard method to lignin dimers. J Phys Chem B. 2011;115:1958-1970;Available from: https://doi.org/10.1021/jp110330q.
4. Nguyen TT, Le PQ, Helminen J, Sipilä J. The 1H and 13C chemical shifts of 5-5 lignin model dimers: An evaluation of DFT functionals. J Mol Struct. 2021;1226:129300;Available from: https://doi.org/10.1016/j.molstruc.2020.129300.
5. Nguyen TT. 1H/13C chemical shift calculations for biaryls: DFT approaches to geometry optimization. R Soc Open Sci. 2021;8:210954;Available from: https://doi.org/10.1098/RSOS.210954.
6. Elder T. Correlation of experimental and ab initio 13C-NMR chemical shifts for monomeric lignin model compounds. J Mol Struct THEOCHEM. 2000;505:257-267;Available from: https://doi.org/10.1016/S0166-1280(99)00389-9.
7. Iron MA. Evaluation of the Factors Impacting the Accuracy of 13C NMR Chemical Shift Predictions using Density Functional Theory - The Advantage of Long-Range Corrected Functionals. J Chem Theory Comput. 2017;13:5798-5819;Available from: https://doi.org/10.1021/acs.jctc.7b00772.
8. Lodewyk MW, Siebert MR, Tantillo DJ. Computational prediction of 1H and 13C chemical shifts: A useful tool for natural product, mechanistic, and synthetic organic chemistry. Chem Rev. 2012;112:1839-1862;Available from: https://doi.org/10.1021/cr200106v.
9. Benassi E. Benchmarking of density functionals for a soft but accurate prediction and assignment of 1H and 13C NMR chemical shifts in organic and biological molecules. J Comput Chem. 2017;38:87-92;Available from: https://doi.org/10.1002/jcc.24521.
10. Tormena CF, da Silva GVJ. Chemical shifts calculations on aromatic systems: A comparison of models and basis sets. Chem Phys Lett. 2004;398:466-470;Available from: https://doi.org/10.1016/j.cplett.2004.09.103.
11. Jain R, Bally T, Rablen PR. Calculating accurate proton chemical shifts of organic molecules with density functional methods and modest basis sets. J Org Chem. 2009;74:4017-4023;Available from: https://doi.org/10.1021/jo900482q.
12. Toomsalu E, Burk P. Critical test of some computational methods for prediction of NMR 1H and 13C chemical shifts. J Mol Model. 2015;21:1-21;Available from: https://doi.org/10.1007/s00894-015-2787-x.
13. Frisch MJ, Cheeseman JR, Scalmani G, et al. Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT, 2013;.
14. Tomasi J, Mennucci B, Cammi R. Quantum mechanical continuum solvation models. Chem Rev. 2005;105:2999-3093;PMID: 16092826. Available from: https://doi.org/10.1021/cr9904009.
15. Binkley JS, Pople JA, Hehre WJ. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. J Am Chem Soc. 1980;102:939-947;Available from: https://doi.org/10.1021/ja00523a008.
16. Dunning TH. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J Chem Phys. 1989;90:1007-1023;Available from: https://doi.org/10.1063/1.456153.
17. Sosa C, Andzelm J, Elkin BC, et al. A local density functional study of the structure and vibrational frequencies of molecular transition-metal compounds. J Phys Chem. 1992;96:6630-6636;Available from: https://doi.org/10.1021/j100195a022.
18. Russell WR, Scobbie L, Chesson A. Structural modification of phenylpropanoid-derived compounds and the effects on their participation in redox processes. Bioorg Med Chem. 2005;13:2537-2546;Available from: https://doi.org/10.1016/j.bmc.2005.01.047.