Proceeding Paper

Communication: DFT Approach towards Accurate Prediction of ¹H/¹³C NMR Chemical Shifts for Dipterocarpol Oxime †

Phong Q. Le [1]**, Nhu Q. Nguyen** [1] **and Thien T. Nguyen** [2,3,]*****

[1] School of Biotechnology, International University, VNU HCM, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam; email1@email.com (P.Q.L.); email2@email.com (N.Q.N.)

[2] Faculty of Pharmacy, College of Medicine and Pharmacy, Duy Tan University, Da Nang 550000, Vietnam

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

* Correspondence: nguyentrongthien@duytan.edu.vn

† Presented at the 27th International Electronic Conference on Synthetic Organic Chemistry, 15–30 November 2023; Available online: https://ecsoc-27.sciforum.net/.

Abstract: A computational NMR approach for accurate predicting the ¹H/¹³C chemical shifts of triterpenoid oximes featuring the screening of 132 DFT methods was demonstrated. Efficiently synthesized dipterocarpol oxime was employed as a model compound. The six highest accurate methods from the screening generated root-mean-square-error (RMSE) values in the range of 0.84 ppm (0.55%) to 1.14 ppm (0.75%) for calculated ¹³C shifts. For ¹H results, simple, economical 6-31G basis set unexpectedly outperformed other more expensive basic sets; and the couple of it with selected functionals provided high accuracy shifts (0.0617 ppm (1.49%) ≤ RMSE ≤ 0.0870 ppm (2.04%)). These computational results strongly supported the proton and carbon assignments of the oxime including the difficult ones of diastereotopic methyl groups, the methyl groups attached to an internal olefin, and diastereotopic-protons.

Keywords: dipterocarpol oxime; diastereotopic methyl groups; NMR; chemical shift; DFT

1. Introduction

Oxime **1** (MW = 457.74) containing a dammarane skeleton (6/6/6/5-fused tetracyclic ring system) [1–3] is a potential intermediate for the syntheses of many functional groups [4], such as lactams via Beckmann rearrangement [5], azetidines via Kulinkovich-type mechanism [6], and pyrroles via N-alkylation-aza-Cope rearrangement (Figure 1b) [7]. The structure elucidation tasks of this triterpenoid derivative using 1D and 2D NMR, including the assignments of diastereotopic methyl groups (**C29** and **C30**), the two methyl groups (**C26** and **C27**) at the internal olefin, and diastereotopic-protons (**H2a** and **H2b**) (Figure 1a), are typically challenging due to similar coupling patterns and a broad overlapping of signals. Therefore, high accuracy ¹H/¹³C calculations would significantly support the full structure interpretation of the oxime including those difficult assignments. Herein, we demonstrate the computational approach to accurate ¹H/¹³C computation for oxime **1** (Figure 2) featuring the screening of 132 combinations from 12 density functional methods and 11 basis sets and the utility of best performing combinations for ¹H/¹³C chemical shift calculations.

(a)

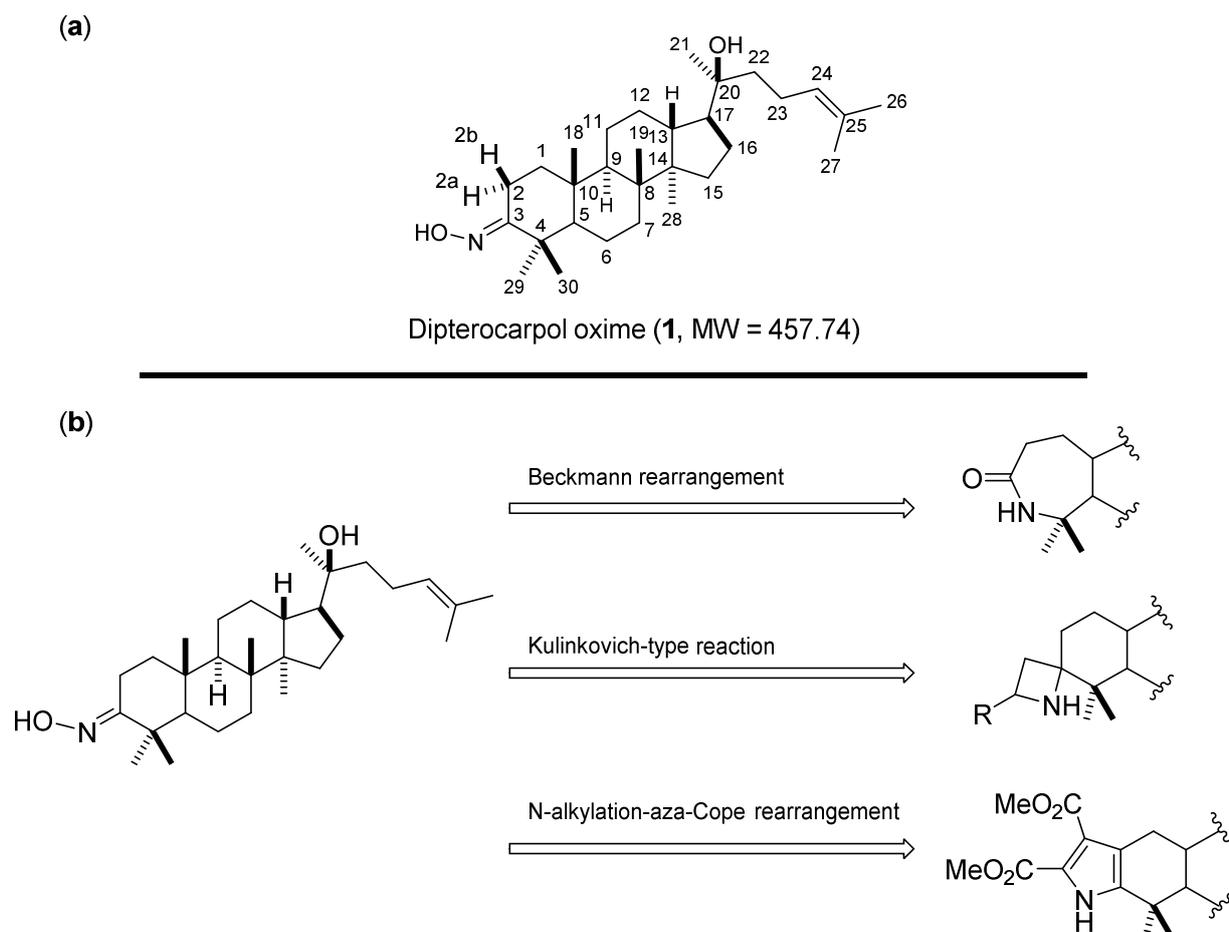Dipterocarpol oxime (**1**, MW = 457.74)

(b)

Figure 1. (**a**) Structures of oxime **1** and (**b**) synthetic potentials of the oxime functional group.

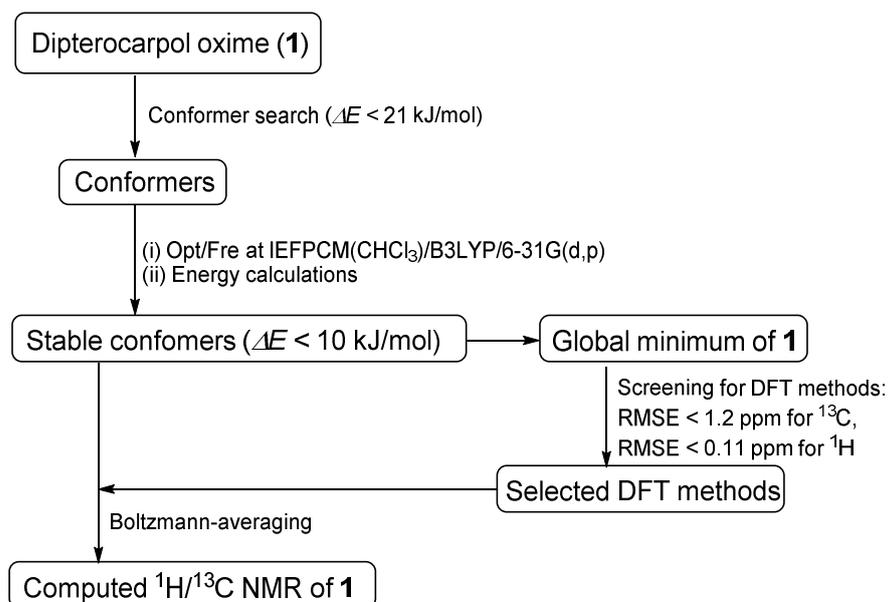**Figure 2.** DFT approach toward accurate $^1H/^{13}C$ NMR chemical shift computation.

2. Computational Methods

Conformer searches of oxime **1** were performed using Macromodel software. The MMFFs force field in gas phase was used due to the presence of a sp^2-hybridized nitrogen in **1**. 10^5 steps of Mixed Torsional Low-Mode sampling algorithm with a maximum

number of 1000 steps per rotable bond were used. A Polak-Ribier Conjugate Gradient minimization method with a maximum number of 2500 iterations and a convergence threshold of 0.05 were applied. A RMSE cutoff of 1.0 Å was used to reduce redundant conformations. Extended non-bonded cutoff distances with Van der Waals cutoff of 8.0 Å and an electrostatic cutoff of 20.0 Å were applied. All local minima within 21 kJ/mol of the global minimum were saved. These conformers were optimized at the level IE-FPCM(CHCl₃)/B3LYP/6-31G(d,p) of theory using Gaussian09 [8]. Subsequent frequency calculations ensured that potential energy surface (PES) local minima were attained. Single-point energies were re-calculated for optimized geometries at the same level of theory with the grid size of ultrafine.

For the screenings of functional and basis set performances, the global minimum was employed. The 12 functionals and 11 basis sets chosen for this investigation due to their common uses for ¹H/¹³C calculations were shown in Tables 1 and 2.

Single-point NMR GIAO calculations were carried out using the above density functional methods and basis sets. Integral equation formalism variant of the polarized continuum model (IEFPCM) was incorporated during NMR calculations [9,10] The GIAO NMR results were observed and extracted using GaussView05. Each optimized structure was used for computing the corresponding isotropic shielding constants (σ_{cal}). To reduce the systematic error of the calculations, the linear regression analysis of calculated shielding constants versus the experimental ones (δ_{exp}) (Equation (1)) were performed and the chemical shifts (δ_{cal}) were computed according to Equation (2). The deviations between computed and experimental chemical shifts were given in the SI. For the ¹H calculations, due to the overlapping proton signals in the experimental spectrum (SI), only assignable protons were considered for the calculations. An average of values of equivalents atoms was assumed. For example, single proton signals are experimentally observed for the methyl groups of **1** due to fast rotations around C-C bonds relative to the NMR measurement time scale. Computed results were evaluated using absolute deviations ($|\Delta\delta|$/ppm, Equation (3)); mean absolute error (MAE/ppm, Equation (4)); root mean squared error (RMSE/ppm, Equation (5)); and the coefficient of determination (r^2). The smaller values of MAE and RMSE indicate smaller errors and the larger value of r^2 means a stronger correlation between theoretical and experimental data. Error calculations and linear correlations were performed using Microsoft Excel 2013.

$$\sigma_{cal} = a\delta_{exp} + b \tag{1}$$

$$\delta_{cal} = (\sigma_{cal} - b)/a \tag{2}$$

$$|\Delta\delta| = |\delta_{scal} - \delta_{exp}| \tag{3}$$

$$\text{MAE} = \sum_{1}^{n} |\delta_{scal} - \delta_{exp}|/n \tag{4}$$

$$\text{RMSE} = \sqrt{\sum_{1}^{n}(\delta_{scal} - \delta_{exp})^2/n} \tag{5}$$

Major contributing conformers with a 10 kJ/mol energy window were used for NMR calculations. Boltzmann weighing average was calculated according to Equation (6), in which σ^i is the shielding constant, E_i is the relative energy of conformer i, R is the molar gas constant (8.3145 J K⁻¹ mol⁻¹), T is the temperature used for the calculation (298 K).

$$\sigma_{cal} = \sum_i \sigma_i\, e^{-(E_i/RT)} / \sum_i e^{-(E_i/RT)} \tag{6}$$

3. Results and Discussion

Oxime **1** was efficiently prepared in 87% yield by the condensation of dipterocarpol (Figure 3A), whose structure was confirmed by XRD analysis [11], with hydroxylamine. With a set of experimental ^{1}H/^{13}C NMR chemical shifts of **1** in hands, we proceeded with the computational study. Initially, the conformer searches of oxime **1** generated 38 conformers. After the optimizations and energy calculations, 15 conformers within the energy window of 10 kJ/mol from the global minimum were located. Figure 3b showed the optimized geometry of the global minimum, which was adopted the chair conformation for all six-member rings including those containing a sp^2-hybridized carbon.
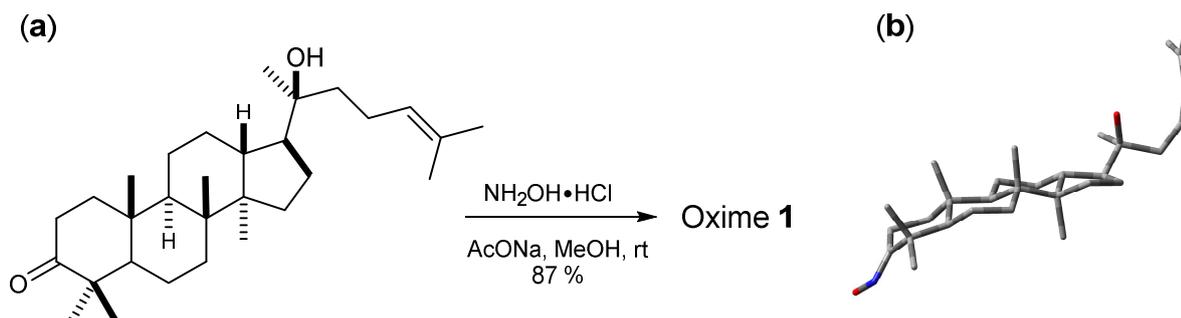

Figure 3. (**a**) Synthesis of oxime **1** and (**b**) its global minimum geometry optimized at IE-FPCM(CHCl$_3$)/B3LYP/6-31G(d,p).

Screening of DFT methods. 132 DFT methods from the combinations of 12 functionals and 11 basis sets were tested for the ^{1}H/^{13}C NMR calculations of **1**. For ^{13}C calculations, B3LYP, B3PW91, CAM-B3LYP, HSEH1PBE, mPW1PW91, and B97XD were more effective with r^2 and RMSE values in the ranges of 0.9899 to 0.9992 and 1.03 ppm (0.68%) to 3.61 ppm (2.38%), respectively (Table 1). CAM-B3LYP was consistent in providing high accuracy results. Among tested basis sets, 6-31G(d,p) (Entry 2), 6-31+G(d,p) (Entry 3), DGDZVP (Entry 10), and DGDZVP2 (Entry 11) showed lowest errors (1.03 ppm (0.68%) ≤ RMSE ≤ 1.57 ppm (1.03%)). 6-31++G(d,p) (Entry 4), 6-311G(d,p) (Entry 5), cc-pVDZ (Entry 9) had relatively good results (1.18 ppm (0.78%) ≤ RMSE ≤ 3.61 ppm (2.38%)) but these basis sets were computationally expensive. It was obvious that too many sets of polarity functions lead to lower accuracy and more computation costs. As showed in Table 1, the use of 6-31G(3d,3p) yielded much larger errors (1.96 ppm (1.29%) ≤ RMSE ≤ 3.61 ppm (2.38%)) in comparison to 6-31G(d,p).

For the predictions of ^{1}H chemical shifts, only assignable protons from ^{1}H and COSY NMR spectra were considered. Those are the protons of 8 methyl groups, an olefinic proton **H24**, diastereotopic protons **H2a** and **H2b**, and allylic protons **H23**, in which the COSY correlation signals of (**H2a**, **H2b**) and (**H23**, **H24**) were unambiguously assigned. The results of B3LYP, LSDA, HSEH1PBE, HCTH, BPV86, PBEPBE (Table 2) were better than those of the remaining density functional methods (Table 2). Simple, economical 6-31G (Entry 1) unexpectedly outperformed the other basis sets with r^2 and RMSE values in the ranges of 0.9915 to 0.9936 and 0.098 ppm (2.30%) to 0.108 ppm (2.54%), respectively. Either adding diffusion or polarity function to the Pople's basis sets or using triple-zeta version (Entry 2–8) significantly lowered the accuracy. The uses of cc-pVDZ, DGDZVP, and DGDZVP2 (Entry 9–11) also generated larger errors. These results (Table 2) would strongly suggest that the screening for accurate ^{1}H calculation is a necessary step for identifying suitable density functional methods and basis sets. Highest performing combinations, including B3LYP/6-31G, LSDA/6-31G, HSEH1PBE/6-31G, HCTH/6-31G, BPV86/6-31G, and PBEPBE/6-31G, were employed for the next step of ^{1}H NMR calculations.

Table 1. Screening density functional methods and basis sets for ^{13}C calculations.

Entry	Basis Set	B3LYP		HSEH1PBE		B3PW91		CAM-B3LYP		mPW1PW91		B97XD	
		r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE
1	6-31G	0.9981	1.54	0.9988	1.24	0.9985	1.37	0.9990	1.16	0.9988	1.24	0.9989	1.20
2	6-31G(d,p)	0.9983	1.50	0.9988	1.24	0.9985	1.37	0.9992	1.03	0.9989	1.20	**0.9991**	**1.08**
3	6-31+G(d,p)	0.9987	1.31	0.9990	1.12	0.9989	1.17	0.9991	1.06	**0.9991**	**1.08**	0.9987	1.27
4	6-31++G(d,p)	0.9988	1.25	**0.9991**	**1.10**	0.9990	1.13	0.9991	1.09	0.9990	1.13	0.9988	1.26
5	6-311G(d,p)	0.9983	1.49	0.9986	1.36	0.9984	1.43	0.9989	1.18	0.9987	1.30	0.9988	1.24
6	6-31G(d,3p)	0.9977	1.72	0.9985	1.37	0.9982	1.51	0.9989	1.20	0.9986	1.32	0.9988	1.21
7	6-31G(3d,p)	0.9930	3.01	0.9951	2.51	0.9939	7.72	0.9963	2.19	0.9950	2.54	0.9976	1.76
8	6-31G(3d,3p)	0.9899	3.61	0.9934	2.91	0.9915	3.31	0.9944	2.68	0.9930	3.00	0.9970	1.96
9	cc-pVDZ	0.9975	1.79	0.9983	1.48	0.9980	1.60	0.9989	1.20	0.9985	1.40	0.9989	1.19
10	DGDZVP	**0.9989**	**1.19**	0.9990	1.12	**0.9990**	**1.13**	0.9992	1.03	0.9990	1.14	0.9990	1.11
11	DGDZVP2	0.9981	1.57	0.9987	1.31	0.9984	1.43	0.9992	1.03	0.9988	1.26	0.9990	1.12

Best performing combinations for each functional are in bold. RMSE values are in ppm.

Table 2. Screening density functional methods and basis sets for ^1H calculations.

Entry	Basis Set	B3LYP		LSDA		HSEH1PBE		HCTH		BPV86		PBEPBE	
		r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE
1	6-31G	**0.9915**	**0.108**	**0.9931**	**0.098**	**0.9915**	**0.108**	**0.9930**	**0.098**	**0.9926**	**0.101**	**0.9926**	**0.101**
2	6-31G(d,p)	0.9881	0.129	0.9897	0.120	0.9874	0.132	0.9886	0.126	0.9892	0.122	0.9889	0.124
3	6-31+G(d,p)	0.9882	0.128	0.9895	0.120	0.9881	0.128	0.9879	0.130	0.9892	0.122	0.9890	0.123
4	6-31++G(d,p)	0.9838	0.219	0.9855	0.142	0.9848	0.146	0.9813	0.161	0.9843	0.148	0.9841	0.149
5	6-311G(d,p)	0.9865	0.137	0.9882	0.128	0.9866	0.136	0.9875	0.132	0.9880	0.129	0.9875	0.132
6	6-31G(d,3p)	0.9864	0.137	0.9878	0.130	0.9863	0.138	0.9875	0.132	0.9879	0.130	0.9876	0.131
7	6-31G(3d,p)	0.9791	0.171	0.9786	0.173	0.9804	0.166	0.9796	0.169	0.9782	0.175	0.9782	0.175
8	6-31G(3d,3p)	0.9782	0.175	0.9775	0.178	0.9790	0.171	0.9756	0.185	0.9770	0.180	0.9769	0.180
9	cc-pVDZ	0.9856	0.141	0.9875	0.132	0.9851	0.144	0.9870	0.134	0.9876	0.131	0.9871	0.134
10	DGDZVP	0.9866	0.136	0.9890	0.124	0.9847	0.146	0.9871	0.134	0.9881	0.128	0.9877	0.130
11	DGDZVP2	0.9886	0.125	0.9907	0.113	0.9879	0.129	0.9901	0.117	0.9903	0.116	0.9900	0.118

Best performing combinations for each functional are in bold. RMSE values are in ppm.

¹H/¹³C NMR calculations. The conformer searches followed by optimizations and energy calculations resulted in 15 geometries within 10 kJ/mol energy window from the global minimum. Selected DFT methods and Boltzmann averaging were used for the ¹H/¹³C NMR calculations. Table 3 showed the high accuracy results using six combinations for ¹³C shifts ($0.9990 \leq r^2 \leq 0.9994$, 0.62 ppm (0.41%) $\leq$ MAE $\leq$ 0.89 ppm (0.59 %), and 0.84 ppm (0.55 %) $\leq$ RMSE $\leq$ 1.14 ppm (0.75 %)). Among them, CAM-B3LYP/DGDZVP (Entry 3) was the best performing methods with the lowest errors. Except HSEH1PBE/6-31G++(d,p), the other methods had $|\Delta\delta_{max}|$ below 3.00 ppm. These results would allow meaningful predictions for all carbon nuclei as showed in Figure 4. All carbon atoms had the absolute deviation averages of the six calculation methods below 1.50 ppm, except **C14** ($|\Delta\delta|$ = 1.83 ppm). The high accuracy NMR calculations would strongly support the assignments of ambiguous carbon chemical shifts including those for diastereotopic methyl carbons (**C29** and **C30**) and the two methyl carbons (**C26** and **C27**) attached to the internal olefin. The calculation results showed that carbons **C30** and **C27** were more shielded than carbons **C29** and **C26**, respectively. It should be noted that even with the careful analysis of 1D & 2D NMR the assignments of carbons **C29** and **C30** would be still challenging.

Table 3. ¹³C NMR calculations using selected combinations .

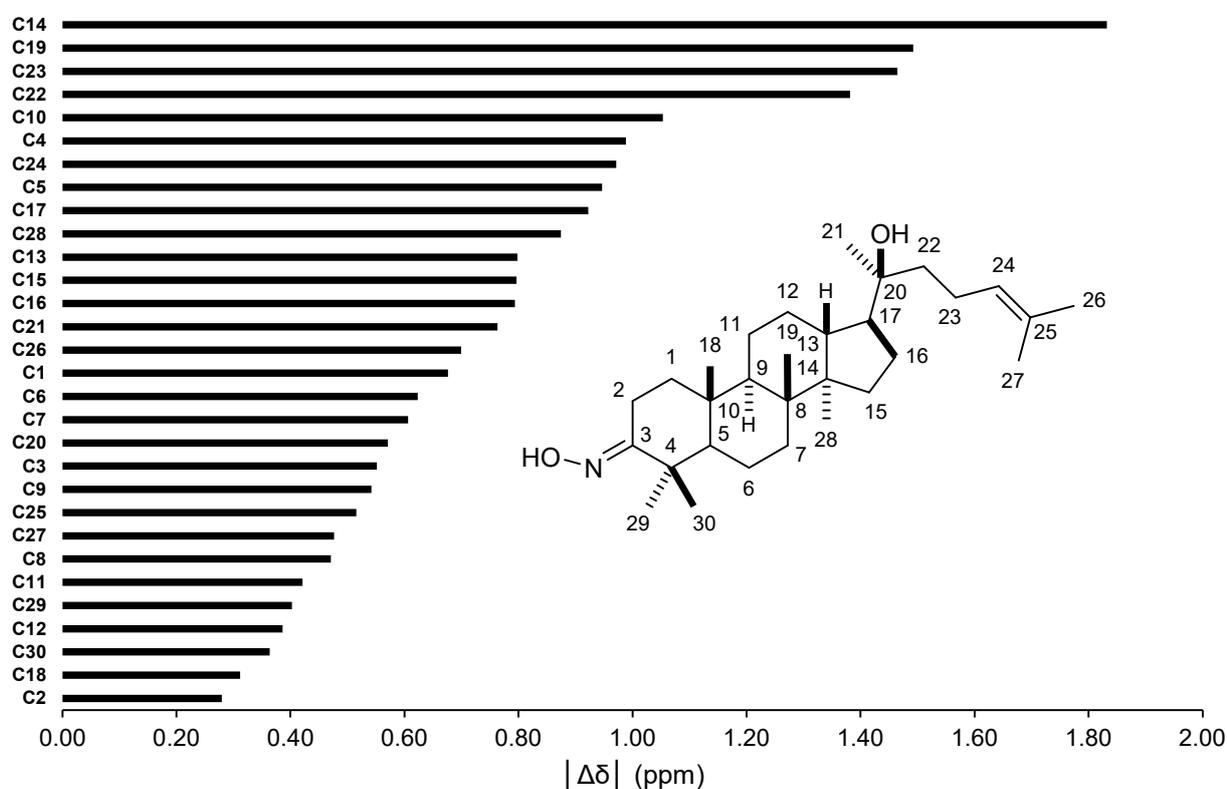| Entry	Combination	r^2	MAE (ppm)	RMSE (ppm)	$|\Delta\delta_{max}|$ (ppm)
1	B3LYP/DGDZVP	0.9993	0.75	0.95	2.09
2	B3PW91/DGDZVP	0.9993	0.71	0.93	2.21
3	CAM-B3LYP/DGDZVP	0.9994	0.62	0.84	2.15
4	HSEH1PBE/6-31++G(d,p)	0.9991	0.89	1.09	3.17
5	mPW1PW91/6-31G+(d,p)	0.9990	0.89	1.14	2.73
6	B97XD/6-31G(d,p)	0.9994	0.72	0.90	2.08

Figure 4. Averages of absolute deviations for ¹³C chemical shifts using selected combinations.

High accuracy ^{1}H results (Table 4, $0.9945 \le r^2 \le 0.9972$, 0.0554 ppm (1.30 %) ≤ MAE ≤ 0.0765 ppm (1.80 %), and 0.0617 ppm (1.45 %) ≤ RMSE ≤ 0.0870 ppm (2.04 %)) were also obtained for the selected methods. The best performing combination was HSEH1PBE/6-31G (Entry 3). Noticeable absolute deviation ($\left|\Delta\delta\right|$ = 0.125 ppm) was observed for proton **H2b** (Figure 5). This result can be explained by the electronic impact of the neighboring oxime group and the solvent effects, which is not sufficiently modelled by the selected DFT methods. The remaining protons had the errors below 0.100 ppm. The high accuracy calculations showed the average shift difference of 0.751 ppm between diastereotopic protons **H2a** and **H2b**, in which proton **H2b** were more shielded than proton **H2a**. This allowed the assignments of proton **H2a** (2.96 ppm) and proton **H2b** (2.27 ppm). The more crowded **H2b**-face of the six-membered ring could be resulted in its lower chemical shift.

Table 4. ^{1}H NMR calculations using selected combinations .

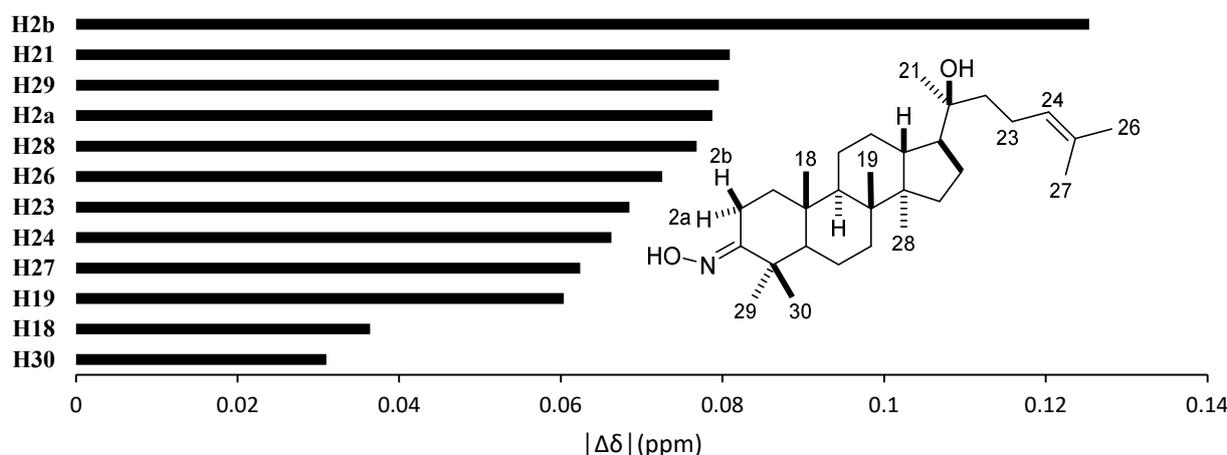| Entry	Combination	r^2	MAE (ppm)	RMSE (ppm)	$\left|\Delta\delta_{max}\right|$ (ppm)
1	B3LYP/6-31G	0.9969	0.0594	0.0647	0.107
2	LSDA/6-31G	0.9965	0.0614	0.0691	0.115
3	HSEH1PBE/6-31G	0.9972	0.0554	0.0617	0.109
4	HCTH/6-31G	0.9966	0.0646	0.0681	0.0940
5	BPV86/6-31G	0.9970	0.0606	0.0643	0.0940
6	PBEPBE/6-31G	0.9945	0.0765	0.0870	0.161

Figure 5. Averages of absolute deviations for ^{1}H chemical shifts using selected combinations.

4. Conclusions

The high accurate ^{1}H/^{13}C chemical shift calculations for dipterocarpol oxime **1** featuring the screening of 132 DFT methods from 12 density functional methods and 11 basis sets were achieved with the lowest RMSE values of 0.84 ppm (0.55%) and 0.0617 ppm (1.49%) for all 30 carbon atoms and 29 considered protons, respectively. B3LYP/DGDZVP, B3PW91/DGDZVP, CAM-B3LYP/DGDZVP, and B97XD/6-31G(d,p) were recommended for ^{13}C calculations. Simple, economical 6-31G basis set coupled with HSEH1PBE, BPV86, and B3LYP unexpectedly provided highest accuracy ^{1}H results. These results strongly supported the challenging assignments of diastereotopic methyl carbons **C29** and **C30**, the methyl carbons **C26** and **C27** attached to the olefin, and diastereotopic protons **H2a** and **H2b**. The further investigation of this project will be conducted in the near future.

5. Experimental Section

Chemical reaction was carried out at International University, VNU HCM. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker 400 MHz spectrometer at ambient temperature at University of Science, VNU HCM. ^{1}H and ^{13}C chemical shifts were reported in ppm using residual solvent peaks as an internal reference (CDCl$_3$: 7.27 ppm for ^{1}H NMR and 77.16 ppm for ^{13}C NMR).

Author Contributions: P.Q.L. & N.Q.N.: Performing the experimental section and experimental NMR analysis. T.T.N.: Conceptualization, Methodology, Software, NMR analysis, Investigation, Writing—original draft, Writing—review & editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This project received no financial support.

Conflicts of Interest: There is no conflict to declare.

References

1. Yang, J.L.; Ha, T.K.Q.; Dhodary, B.; Kim, K.H.; Park, J.; Lee, C.H.; Kim, Y.C.; Oh, W.K. Dammarane triterpenes as potential SIRT1 activators from the leaves of Panax ginseng. *J. Nat. Prod.* **2014**, *77*, 1615–1623. https://doi.org/10.1021/np5002303.
2. Homhual, S.; Bunyapraphatsara, N.; Kondratyuk, T.; Herunsalee, A.; Chaukul, W.; Pezzuto, J.M.; Fong, H.H.S.; Zhang, H.J. Bioactive dammarane triterpenes from the mangrove Plant Bruguiera gymnorrhiza. *J. Nat. Prod.* **2006**, *69*, 421–424. https://doi.org/10.1021/np058112x.
3. Pakhathirathien, C.; Karalai, C.; Ponglimanont, C.; Subhadhirasakul, S.; Chantrapromma, K. Dammarane triterpenes from the hypocotyls and fruits of Ceriops tagal. *J. Nat. Prod.* **2005**, *68*, 1787–1789. https://doi.org/10.1021/np0502793.
4. Bolotin, D.S.; Bokach, N.A.; Demakova, M.Y.; Kukushkin, V.Y. Metal-Involving Synthesis and Reactions of Oximes. *Chem. Rev.* **2017**, *117*, 13039–13122. https://doi.org/10.1021/acs.chemrev.7b00264.
5. Kaur, K.; Srivastava, S. Beckmann rearrangement catalysis: A review of recent advances. *New J. Chem.* **2020**, *44*, 18530–18572. https://doi.org/10.1039/d0nj02034f.
6. Behnke, N.E.; Lovato, K.; Yousufuddin, M.; Kürti, L. Titanium-Mediated Synthesis of Spirocyclic *NH*-Azetidines from Oxime Ethers. *Angew. Chemie Int. Ed.* **2019**, *58*, 14219–14223. https://doi.org/10.1002/anie.201909151.
7. Liao, J.; Ouyang, L.; Jin, Q.; Zhang, J.; Luo, R. Recent advances in the oxime-participating synthesis of isoxazolines. *Org. Biomol. Chem.* **2020**, *18*, 4709–4716. https://doi.org/10.1039/d0ob00963f.
8. Frisch, M.J.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.P.; et al. *09, Revision D.01*; Gaussian, Inc.: Wallingford, CT, USA, 2013.
9. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3093. https://doi.org/10.1021/cr9904009.
10. JTomasi; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct. THEOCHEM* **1999**, *464*, 211–226. https://doi.org/10.1016/S0166-1280(98)00553-3.
11. Smirnova, I.E.; Thu, H.D.T.; Kazakova, O.B.; Tolstikov, G.A.; Kukovinets, O.S.; Lobov, A.N.; Suponitskii, K.Y. Ozonolysis of dipterocarpol and its derivatives. *Russ. J. Org. Chem.* **2012**, *48*, 1370–1376. https://doi.org/10.1134/S1070428012100193.