DFT calculations of $^1$H- and $^{13}$C-NMR chemical shifts of 3-methyl-1-phenyl-4-(phenyldiazenyl)-1H-pyrazol-5-amine in solution

Zaki S. Safi* & Nuha Wazzan2

Geometries of the 3-methyl-1-phenyl-4-(phenyldiazenyl)-1H-pyrazol-5-amine azo-dye compound and its tautomer were optimized using B3LYP and M06-2X functionals in coupling with TZVP and 6–311 + G(d,p) basis sets. The $^1$H- and $^{13}$C-NMR chemical shifts of all species were predicted using 13 density functional theory (DFT) approaches in coupling with TZVP and 6–311 + G(d,p) basis sets at the different optimized geometries by applying the using GIAO method using the eight geometries. The selected functionals are characterized by having different amount of Hartree–Fock exchange. The selected DFT methods were B3LYP, M06-2X, BP86, B97XD, TPSSPSS, PBE1PBE, CAM-B3LYP, wB97XD, LSDA, HSEH1PBE, PW91PW91, LC-WPBE, and B3PW91. The results obtained were compared with the available experimental data using different statistical descriptors such as root mean square error (RMSE) and maximum absolute error (MAE). Results revealed that the prediction of the $^1$H-NMR chemical shifts has more significant dependence on the applied geometry than that of the prediction of the $^{13}$C-NMR chemical shifts. Among all the examined functionals, B97D and TPSSPSS functionals were found to be the most accurate ones, while the M06-2X functional is the least accurate one. Results also revealed that the prediction of NMR chemical shifts using TZVP basis sets results is more accurate results than 6–311 + G(2d,p) basis set.

Azo-Dyes, consisting of 60–70% of dyes, have a great importance in science and technology, such as their applicability in optical materials, molecular data processing, ion sensors, and biological applications. Simultaneously, most of the azo dyes are tautomeric or potentially tautomeric. The tautomerism phenomenon, which plays a significant role in the stability and the color of the dyes, has been extensive studied during the last decades. The intramolecular 1,3-proton transfer process, which is very fast at the NMR time scale, occurs by the migration of a proton from the amino nitrogen atom to the carbonyl oxygen atom. Thus, to overcome this problem, an averaged signal is typically measured. Therefore, reference compounds are needed to evaluate the proportion of the azo and hydrazo tautomers. The selection of these references that represent the pure azo and hydrazo tautomers is necessary. Thus, there is a need for such reference compounds to be identified and used in the NMR investigations of tautomeric azo dyes.

Up to now, nuclear magnetic resonance (NMR) spectroscopy has been one of the most powerful and reliable methods for structure elucidation. $^1$H- and $^{13}$C-NMR are the most widely studied nuclei. However, other nuclei such as $^{15}$N, $^{31}$P, $^{18}$O, etc., are also under consideration, which can be attributed to the corresponding routinely available techniques readily yielding an overabundance of structural and kinetical information for the proper identification of novel compounds. Unfortunately, complex compounds suffer from the main problem in the full signal assignment, which is sometimes impossible, even with the utilization of multidimensional NMR experiments.

Calculating the nuclear magnetic shielding tensor (NMST) is required for the theoretical NMR investigations. For this reason, several computational methods were developed such as the GIAO (gauge-independent atomic orbitals), the IGLO (an individual gauge for localized orbital) and the CGST (continuous set of gauge transformation) methods. In general, the GIAO method has been shown to achieve faster convergence of the shielding value with respect to the size of the basis set and, consequently, has been the most commonly employed procedure.

---

1Department of Chemistry, Faculty of Science, Al Azhar University-Gaza, P.O. Box 1277, Gaza, Palestine. 2Chemistry Department, Faculty of Science, King Abdulaziz University, P.O. Box 42805, Jeddah 21589, Saudi Arabia. *email: zaki.safi@gmail.com; z.safi@alazhar.edu.ps
Last past years have witnessed a great development in the growth of quantum mechanical calculations quickly, which aids the determination of compound structures\cite{22-24}. Different levels of theory can be effectively used to compute the NMR shielding constants. These levels include coupled-cluster theories (e.g., coupled-cluster singles–doubles (CCSD) and coupled-cluster singles–doubles–perturbative-triples, CCSD(T))\cite{23} and Möller-Plesset (MP) second-order perturbation theory (MP2)\cite{22,25}. Exactly, the most accurate sets and highly comparable with the experimental data of benchmark NMR shielding constants were performed using MP2, CCSD, and CCSD(T) methods in various basis sets. Unfortunately, these methods are highly expensive and computationally time consumable, and they are only exclusively applicable to small model molecules and not for larger ones\cite{22-25}.

In spite of being less accurate, DFT methods, on the opposing, can handle structures composed of several hundreds of atoms\cite{26-31}. DFT methods allow the calculation of useful chemical properties, structural and electronic analysis of molecules and the theoretical values of the $^{13}$C- and $^1$H-NMR chemical shifts with a high degree of accuracy compared to the experimental data\cite{32,33}. Bango et al.\cite{27} determined the $^{13}$C-NMR chemical shifts experimentally and theoretically by DFT calculations B3LYP with the 6–31G(d,p) basis sets. Bally and Rablen\cite{28} proved that B3LYP/6-31G(d,p)u + 1 s (one of the most economical methods) and B3LYP/cc-pVTZ basis sets. Therefore, the optimization process, which are: (i) the triple-ζ Pople basis sets augmented with additional polarization functions, for a more accurate prediction of trends in the experimental $^1$H-NMR chemical shifts using the BMK method with different basis sets in toluene-$_d_8$ solvent using IEFPCM solvation model. The agreement between the experimental data and the DFT computed results was improved by inclusion of vibrational corrections for DFT methods, which contrasts the behavior of the coupled-cluster methods\cite{23,24,29,34-40}. This methodology could be applied as a promising alternative in support of traditional NMR experimental techniques. Furthermore, previous studies\cite{40,41} showed that DFT theory brings a respected tool for the prediction of $^{13}$C-NMR properties, even with a small basis set. Additionally, it was recommended to use a larger basis set, including polarization functions, for a more accurate prediction of trends in the experimental $^1$H-NMR chemical shifts\cite{41}. Furthermore, most of studies that related to the computational of $^1$H- and $^{13}$C-NMR were mainly focused on DFT and ab initio performance for structures in CDCl$_3$, which expected due to the fact that a large number of experimental spectra are taken in deuterated chloroform and there is an abundance of data for numerous molecules and functional groups.

**Test set.** The 3-methyl-1-phenyl-4-(phenylazanyl)-1H-pyrazol-5-amine (T1) may have another conformer (T2) that can rapidly interconvert at the NMR time scale with relative energy of about 9.3 kcal/mol\cite{10} (Fig. 1). The $^{1}$H- and $^{13}$C-NMR experimental chemical shifts of the main species (T1) in CDCl$_3$ were taken exclusively from Deneva et al.\cite{10}. All protons that are bonded to the nitrogen atoms were not included in the validation test because they are strongly affected by the ability to form hydrogen bonds as well as the solvent environment\cite{14}. This study is undertaken to present and discuss the following issues: (i) computation of $^1$H- and $^{13}$C-NMR chemical shifts for the species under probe at different geometries that optimized using B3LYP and M06-2X approaches with TZVP and 6–311 + G(2d,p) basis sets within the DFT framework using the GIAO\cite{42} method and (ii) statistical validation of the theoretical NMR chemical shifts based on the available experimental data.

**Computational methodology.** Geometries of the investigated species were optimized using two-hybrid DFT approaches, B: B3LYP\cite{33,44} and the M: M06-2X (meta-GGA)\cite{35,46} in acetonitrile as a solvent using the Polarizable Continuum Model (the integral equation formalism variant, IEFPCM)\cite{48}. Two basis sets were included in the optimization process, which are: (i) the triple-$ζ$ Pople basis sets augmented with additional polarization functions 6–311 + G(2d,p)\cite{46} and (ii) triple zeta for valence electrons plus polarization function, TZVP\cite{47}. Therefore, To ensure that all geometries are minima to the electronic potential energy surface, the vibrational frequencies of each species were calculated at the same levels of theories, and the absence of any imaginary frequencies was noted. The isotropic magnetic shielding tensors were computed using thirteen different XC DFT approaches, applying the gauge-independent atomic orbital (GIAO)\cite{42} approach in coupling with the same basis sets CHCl$_3$ solvent. The classification of the thirteen XC different functionals are given in supplementary materials. In order to simplify the following of our study, DFT approaches B3LYP and M06-2X are given the abbreviations T and P, respectively. The two bases sets TZVP and B3LYP/6–311 + G(2d,p) are given the abbreviation T and P, respectively. So that, BT, BP, MT and MP optimization levels were considered. To calculate the NMR chemical shifts,
XC DFT approach is coupled with the two bases sets at the considered geometries. In total, eight combinations of theoretical approach will be used to compute the NMR chemical shifts (see Table S1 of the supporting information). All quantum chemical calculations were performed using Gaussian 09 computational software package49.

Because NMR chemical shifts result from a dynamic process, the isotropic magnetic shielding tensors were averaged over all symmetry related carbons and hydrogens where applicable. For each $^1$H and $^{13}$C nucleus, the average isotropic magnetic tensor of the two tautomeric forms $\delta_{\text{cal}}$ was taken. Then, the isotropic chemical shifts $\delta_{\text{cal}}$ was defined as $\delta_{\text{cal}} = \sigma_{TMS} - \sigma_{TMS}$, where $\sigma_{TMS}$ is the anisotropic shielding constant of $^1$H and $^{13}$C in trimethylsilane (TMS). Notice that $\sigma_{TMS}$ values of $^1$H- and $^{13}$C- NMR of the TMS were computed using the same strategy that used for the azo-dye species and based on the MP2 geometry (all $\sigma_{TMS}$ are listed in Table S2 of the supporting information). Also, it should be noted that the $\sigma_{TMS}$ for the $^1$H- and $^{13}$C-NMR were calculated as the average for the 12 hydrogen nuclei and 4 carbon atoms, respectively. The $\delta_{\text{cal}}$ values were collected in Tables S3 and S4 of the supporting information.

Due to the possibility of the tautomerization of the investigated species at the NMR time, the scaled $^1$H- and $^{13}$C-NMR chemical shifts of the hydrogen and carbon nuclei of the T1 and T2 species were averaged. The averaged $^1$H- and $^{13}$C-NMR chemical shifts were tested for possible outliers using 1.5xIQR rule, and no outliers were detected. Computation of the different statistical descriptors was performed using Microsoft Excel 2016 using the following equations:

\[
\text{(Mean signed error) ME} = \frac{1}{N} \sum (\delta_{\text{cal}} - \delta_{\text{exp}})
\]  (1)

\[
\text{Absolute difference(error) } \Delta \delta = |\delta_{\text{cal}} - \delta_{\text{exp}}|
\]  (2)

\[
\text{Maximum Absolute Error (Mean unsigned error) : } \text{MAE} = \frac{1}{N} \sum \Delta \delta
\]  (3)

\[
\text{Maximum difference : } \text{MAX} = \max \Delta \delta
\]  (4)

\[
\text{Mean square error : } \text{MSE} = \frac{1}{N} \sum \Delta \delta^2
\]  (5)

\[
\text{Root mean square error : } \text{RMSE} = \sqrt{\frac{1}{N} \sum \Delta \delta^2} = \sqrt{\text{MSE}}
\]  (6)

\[
\text{Mean absolute percentage error : } \text{MAPE} = \frac{\sum_{i=1}^{n} |\delta_{\text{cal}} - \delta_{\text{exp}}|}{n} \times 100
\]  (7)

Results and discussion

The optimized structures of the two envisaged tautomers of the azo-dye species as obtained using B3LYP/TZVP level of theory in acetonitrile are shown in Figure S1 of the supporting information. The statistical parameters (MAE, MAX, RMSE and MAPE) are measured for suitable errors. Among all of the statistical descriptors, RMSE is the most powerful descriptor that can be used to compare the different DFT methods. The smaller the RMSE parameter, the more accurate the DFT approach to estimate the $^1$H- and $^{13}$C-NMR chemical shifts.

Benchmark of $^1$H NMR chemical shifts. Tables S5 and S6 of the supporting information list all the descriptors (RMSE, MSE, MAE, MAX, MAPE, $R^2$, slope, and intercept) computed for $^1$H-NMR deviations. Excellent linear regression correlation coefficient ($R^2 > 0.997$–0.999) is obtained. However, the slope ($0.906$–$1.118$ ppm) and the intercept ($−0.456$–$0.534$ ppm) have slightly deviated from the ideal values for all combinations (Fig. 2 and Tables S5 and S6). The poorest regression was obtained for M06-2X functional, while the best regression was obtained for B3LYP/TZVP. The reason may be attributed to the possible interaction (by non-covalent van der Waal) between the hydrogen atom (van der Waal radii = 2.02 Å) and the N(β) (van der Waal radii = 2.25 Å) of the azo-dye group, resulting in a weak hydrogen bond. The geometrical structure of the T1 and T2 tautomers show that the distance between the hydrogen atom and N(β) atom is an average of 2.450 Å, which strongly confirms our suggestion.

The data given in Tables S2 and S5, show that the signed deviations (ME) are typically positive, indicating that the computed $^1$H NMR chemical shifts are exclusively higher than that of the experimental ones. It is also found that ~ 93.5% of the ME values are positive. The number of negative ME values are ~ 96 out of 1456, with a percentage of ~ 6.5%. Moreover, most of the negative values belong to the hydrogen atom attached to C10 atom, which may be attributed to the consideration of the average value of the isotropic magnetic tensors of the three hydrogen atoms attached to C10.
Basis set effect. In order to analyze the effect of the basis set that couples with the XC functional on the 1H-NMR chemical shift, two basis sets TZVP and 6–311 + G(2d,p) were selected. For the sake of comparison, the eight combinations (Table S4) were divided into two parts based on the basis sets, which coupled with NMR XC functional, as follows:

- Part (1) in which the NMR functional combined with TZVP basis set, and it includes C1, C3, C5 and C7 combinations, and
- Part (2) in which the NMR functional combined with 6–311 + G(d,p) basis set, includes C2, C4, C6, and C8 combinations.

To simplify the discussion, the average absolute differences in RMSE values were considered and plotted as pie charts (Fig. 3). In this case, changing the basis set coupled with the XC functionals to compute the 1H-NMR chemical shifts is considered. Taking into account the same geometry level, the results show that the accuracy of the 1H-NMR chemical shifts as computed by TZVP basis set is higher than that computed by 6–311 + (2d,p) basis set. Actually, it is found that the results obtained by TZVP basis set are 0.069–0.101 ppm higher than those obtained by 6–311 + G(2d,p) one. Moreover, the highest differences are found in the case of using 6–311 + G(2d,p) basis set in the optimization process (combinations C3, C4, C7 and C8). These results led us to conclude that the
The accuracy of the computed ¹H-NMR values using TVZP basis set is much higher than that of 6–311 + G(2d,p) one, and this can be considered as the best choice to compute the ¹H-NMR chemical shifts.

**Geometry level effect.** Four different geometrical structures were optimized using BT, MT, BP, and MP levels of theories (Table S1). Figure 4 shows the graphical representation by columns for the average differences of the RMSE values obtained by taking into account the same basis sets that are coupled with the different XC functionals. Computation of the ¹H-NMR chemical shifts using DFT/TZVP at B3LYP/6–311 + G(2d,p) geometries (Combination 3) is more accurate than those at B3LYP/TZVP geometries (C1) (Fig. 4a). Whereas the computed ¹H-NMR chemical shifts at M06-2X/6–311 + G(2d,p) geometry (combination 5) are more accurate than those at the M06-2X/TZVP geometry (Combination 7). On the other hand, using B3LYP/6–311 + G(2d,p) geometry can compute more accurate ¹H-NMR chemical shifts than the M06-2X/TZVP geometries (combinations C3 and C5). Similar results can also be found for the case of DFT/6–311 + G(2d,p) NMR level at the four geometries (Fig. 4b).

Based on the above results, one may conclude that taking into account either M06-2X or B3LYP functionals, using of the 6–311 + G(2d,p) basis set in the optimization steps computes accurate more accurate ¹H-NMR chemical shifts than the TZVP basis sets.

**XC functionals effect.** Besides to the basis set and geometry level effects, one of the main questions of this study is whether the DFT functionals can compute the most accurate ¹H-NMR chemical shifts? The answer of this question can be achieved by a careful looking at the results obtained indicates, in general, that M06-2X, LC-WPBE, and wB97XD functionals compute the poorest ¹H-NMR chemical shifts with very high RMSE values.
Combination effect. Eight combinations (C1–C8) have been considered (Table S1) to compute the $^1$H-NMR chemical shifts. To shed some light on the effect of the several combinations on the accuracy of the computed $^1$H-NMR chemical, Fig. 5a,b shows the RMSE% of the B97D and TPSSTPSS functionals. It is clearly seen that the most accurate results are obtained for the DFT approach with the TBP combination (C3). Indeed, the RMSE% are 9.5% and 10.8% for B97D and TPSSTPSS functionals, respectively. Whereas the least accurate results were computed using the DFT approach with PBT combination (C4), with RMSE% of 15.2% and 14.9% for B97D and TPSSTPSS functionals, respectively. These results led us to safely conclude that the $^1$H-NMR chemical shifts, with a high degree of accuracy, can be computed by either B97D or TPSSTPSS functionals in coupling with TBP.

To get an exclusive order for the different DFT approaches, the computed descriptors obtained for the TBT, TBP, TMT and TMP combinations were averaged, and the results are listed in Table 1. The most important statistical descriptors (RMSE, MAX, MAE, and MAPE) are graphically shown in Figure S3. It is clearly obvious from Table 1 and Figure S3 that the corresponding RMSE, MAX, MAE, ME and MAPE descriptors are 0.006 (1.6%), 0.020 (3.5%), 0.006 (1.6%), 0.004 (2.8%), 0.008 (2.7%) and 0.002 (3.4%) ppm, respectively. These results indicate that there are no significant differences in the computed $^1$H-NMR chemical shifts based on using either B3LYP or M06-2X geometry. These findings led us to conclude that the main factors that can influence the calculation of the $^1$H-NMR chemicals are the DFT approach and the selected basis sets, regardless of the geometry level used.

Taken into account the B3LYP geometry, the accuracy of the different functionals can be ranked as follows: TPSSTPSS > B97D > BP86 > PW91PW91 > B3LYP > B3PW91 > PBE1PBE > HSEH1PBE > LSDA > CAM-B3LYP > wB97XD > LC-WPBE > M06-2X. Whereas, taken into account the M06-2X geometry, the accuracy of the different functionals can be ranked as follows: B97D > TPSSTPSS > BP86 > PW91PW91 > B3LYP > B3PW91 > PBE1PBE > HSEH1PBE > LSDA > CAM-B3LYP > wB97XD > LC-WPBE > M06-2X.

Benchmark of $^{13}$C-NMR chemical shifts. In the following, we are going through the analysis of the results obtained for the prediction of $^{13}$C-NMR. Tables S7 and S8 list all the statistical descriptors (RMSE, MSE, MAE, MAPE, and R$^2$, slope and intercept) computed for $^{13}$C-NMR deviations. An excellent linear regression correlation coefficient (R$^2$ > 0.986–0.993) was obtained. It is also found that, however, the slope (0.963–1.149 ppm), 84% of these values are in agreement with the ideal values of slope (Fig. 6). The intercept (0.342–3.786 ppm) differed slightly from the ideal values for all combinations (see Tables S5 and S6). The best regressions are obtained for TPSSTPSS and B97D functionals. Whereas, the least accurate results are obtained using M06-2X, PW91PW91, CAM-B3LYP and HSEH1PBE functional.

Figure 7 shows only the statistical descriptors (RMSE, MAX, MAE, and MAPE) for the best $^{13}$C-NMR results (B97D, TPSSTPSS, BP86, B3PW91, PBE1PBE and PW91PW91 functionals), which agrees with the previous results. With the exception to the results of wB97XD and CAM-B3LYP functionals, the accuracies of the different DFT approaches that were used to compute the $^1$H- and $^{13}$C-NMR chemical shifts are almost the same. Therefore, the most accurate $^{13}$C-NMR chemical shifts are computed using B97D, TPSSTPSS, and BP86 functionals.

Whereas the least accurate results are obtained using M06-2X, LC-WPBE, and CAM-B3LYP functionals.
Combination effect. To analyze the effect of the different combinations (C1-C8) on the accuracy of the $^{13}$C-NMR chemical shifts, we apply the same scenario we used for $^1$H-NMR results. In comparison with the experimental results, Fig. 8 shows the 3D color pie charts of the RMSE computed for the $^{13}$C-NMR chemical shifts obtained using B97D and TPSSTPSS, BP86, and PW91PW91. For both functionals, the least deviations of 12.13 and 12.22% (TPSSTPSS) were computed using the B97D/TZVP//M06-2X/TZVP and TPSSTPSS/TZVP//M06-2X/TZVP levels, respectively, while the highest deviations (least accurate) were computed at 12.83 and 12.76% were computed using the B97D/6–311 + G(2d,p)//M06-2X/6–311 + G(2d,p) and TPSSTPSS/TZVP//B3lYP/6–311 + G(2d,p) levels, respectively. Based on these results, one concludes that the prediction of the $^{13}$C NMR chemical shifts can be accurately computed by using the D97D and/or TPSSTPSS in coupling with the TZVP basis set at the M06-2X/TZVP geometry.

To rank the different DFT approaches that we have used to compute the $^{13}$C-NMR chemical shifts, the data in Tables S7 and S8 were averaged and classified into two main combinations. The first one averages the computed descriptors based on the B3LYP geometry with the TZVP and 6–311 + G(2d,p) basis sets, while the second one averages the computed descriptors based on M06-2X geometry with the TZVP and 6–311 + G(2d,p) basis sets. The results are listed in Table 2, and the data are sorted by RMSE descriptors.

For the B97D, TPSSTPSS, BP86, PW91PW91 and B3PW91, the average statistical errors will be as follows: RMSE = 4.74 and 4.32 ppm, MAX = 9.60 and 9.10 ppm, MAE = 3.83 and 3.50 ppm, ME = 2.10 and 1.31 ppm, and MAPE = 4.55 and 4.16% for B3LYP and M06-2X geometries, respectively. These results strongly confirm that the prediction of the $^{13}$C-NMR chemical shifts on the base of the M06-2X geometry, regardless of the NMR functional and basis set, is more accurate than the B3LYP geometry.

Finally, the data given in Tables S7 and S8, show that the signed deviations (ME) are typically positive, indicating that the computed $^{13}$C NMR chemical shifts are exclusively higher than the experimental ones. It is found that > 87% of the ME values are positive. Indeed, the number of negative ME equals approximately 340 out of 2496 values, with a percentage of ~13%). The maximum number of the negative ME values was computed for TPSSTPSS functionals with a percentage of ~ 40%.

Table 1. The averaged descriptors RMSE, MAE, ME, MAX, MAPE, ME, R², slope, and intercept computed for $^1$H-NMR deviations for each combination of geometry level, basis set, and functional. The table is sorted by RMSE column.
Conclusions

Eight different combinations of DFT functionals, geometry level, and basis sets have been examined for the prediction of 1H and 13C-NMR chemical shifts in 3-methyl-1-phenyl-4-(phenyldiazenyl)-1H-pyrazol-5-amine. The computed results have been predicted using experimental shifts as a reference using several statistical descriptors. From the calculated data, the following remarked conclusion can be extracted:

- Excellent linear correlations have been obtained between the DFT-calculated and experimental 1H- and 13C-NMR chemical shifts for all functional, basis set, and geometry level tested.
- The slopes of the linear correlations are within the ideal value for the 13C-NMR chemical, and it is slightly deviated in the case of 1H-NMR chemical shifts.
- The accuracy of the 1H-NMR chemical shifts computed using the TVZP basis set on the four suggested geometries was higher than those obtained by the 6–311 + G(2d,p) basis set.
- The B97D, TPSSTPSS, and BP86 functionals can compute the 1H- and 13C-NMR chemical shifts with a reasonable degree of accuracy.
- For 13C-NMR chemical shifts, the least errors have been obtained using the combination of B97D and TPSS-TPSS functionals with TZVP basis sets with taken into account M06-2X/TZVP geometry.

Figure 6. Correlation between the experimental 13C-NMR chemical shifts and the computed 13C-NMR chemical shifts as calculated using six different DFT methods (BP86, B97D, HSEH1PBE, PBE1PBE, B3PW91, and TPSSTPSS) with TZVP basis set at the B3LYP/TZVP geometry (TBT combination).
B97D, TPSSTPSS, and BP86 functionals are excellent choices to calculate the NMR chemical shifts with highly reasonable accuracy.

M06-2X, LC-WPBE, WB97XD, and CAM-B3LYP are the poor functional choice to predict either 1H- or 13C- NMR chemical shifts.

Both TZVP and 6–311 + G(2d,p) basis sets are very good choices to predict the NMR chemical shift. However, the TZVP basis set can predict the NMR shifts with a higher degree of accuracy.

Figure 7. Graphical representation of statistical descriptors ((a) RMSE, (b) MAX, (c) MAE and (d) MAPE) computed for $^1$C deviations versus each combination of geometry and basis set. Notice: T and P denote TZVP basis sets, B and M correspond to B3LYP and M06-2X DFT functionals used for the geometry optimization process.
Data availability

All data generated or analyzed during this study are included in this published article [supporting information-updated]. Additionally, the datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 29 August 2022; Accepted: 20 October 2022
Published online: 22 October 2022
References

1. Zollinger, H. Color chemistry: Syntheses, properties, and applications of organic dyes and pigments (John Wiley & Sons, Hoboken, 2003).
2. Kelemen, J. Azo-hydrazine tautomerism in azo dyes. I. A comparative study of 1-phenylazo-2-naphthyl and 1-phenylazo-2-naphthalamine derivatives by electronic spectroscopy. Dyes Pigm. 2, 73–91 (1981).
3. Kelemen, J., Moss, S., Sauter, H. & Winkler, T. Azo—hydrazine tautomerism in azo dyes. II. Raman, NMR and mass spectrometric investigations of 1-phenylazo-1-naphthalamine and 1-phenylazo-2-naphthalamine derivatives. Dyes Pigm. 3, 27–47 (1982).
4. Kelemen, J., Korman, G. & Rihs, G. Azo—hydrazine tautomerism in azo dyes. III. The tautomeric structure of 1-(4-nitrophenylazo)-2-naphthalamine from crystal structure determination. Dyes Pigm. 3, 249–271 (1982).
5. Kelemen, J., Moss, S. & Glitsch, S. Azo-hydrazine tautomerism in azo dyes. IV. Colour and tautomeric structure of adsorbed 1-phenylazo-2-naphthalamine and 1-phenylazo-2-naphthalamine dyes. Dyes Pigm. 5, 83–108 (1984).
6. Safi, Z. S. & Al-Hendawi, H. M. Tautomerization and substituent effects on the intramolecular hydrogen bonding in 4-formyl-3-methylpyrazol-5-ol. A density functional theory. Asian J. Chem. 21, 4772 (2009).
7. Kleinmeyer, E. NMR spectroscopic study of tautomerism in solution and in the solid state. Tautomerism: methods and theories, 103–143 (2013).
8. Faris, W. M. & Safi, Z. S. Theoretical investigation of tautomerism stability of hydantoin in the gas phase and in the solution. Orient. J. Chem. 30, 1045–1054 (2014).
9. Safi, Z. S. Conceptual density functional theory and its application in the chemical domain (Apple academic press, New Jersey, 2018).
10. Deneva, V. et al. Tautomerism in azo dyes: Border cases of azo and hydrazo tautomers as possible NMR reference compounds. Dyes Pigm. 165, 157–163 (2019).
11. Wazzan, N., Safi, Z. & Al-Qurashi, O. DFT investigation on the linear and nonlinear optical properties of the tautomers and derivatives of 2-aminobenzothiazole (ABT) in the gas phase and different solvents. J. King Saud Univ.-Sci. 32, 2710–2716 (2020).
12. Nedelcheva, D., Antonov, L., Lycka, A., Damyanova, B. & Popov, S. Chemometric models for quantitative analysis of tautomeric Schiff bases and azo dyes. Curr. Org. Chem. 13, 217–240 (2009).
13. Pankratyev, E. Y., Tulyabaev, A. R. & Khalilov, L. M. How reliable are GIAO calculations of 1H and 13C NMR chemical shifts? A benchmarking study. J. Phys. Chem. A 115, 141–147 (2011).
14. Maurer, M. & Ochsenfeld, C. A linear-and sublinear-scaling method for calculating NMR shieldings in atomic orbital-based second-order Muller-Plesset perturbation theory. J. Chem. Phys. 130, 041113 (2013).
15. Lee, C., YANG, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron correlation energy. Phys. Rev. B 38, 785–795 (1988).
16. Wazzan, N., Safi, Z. & Al-Qurashi, O. DFT investigation on the linear and nonlinear optical properties of the tautomers and derivatives of 2-aminobenzothiazole (ABT) in the gas phase and different solvents. J. King Saud Univ.-Sci. 32, 2710–2716 (2020).
17. Nedelcheva, D., Antonov, L., Lycka, A., Damyanova, B. & Popov, S. Chemometric models for quantitative analysis of tautomeric Schiff bases and azo dyes. Curr. Org. Chem. 13, 217–240 (2009).
18. Wiitala, K. W., Hoye, T. R. & Cramer, C. J. Hybrid density functional methods empirically optimized for the computation of 13C and 1H chemical shifts in chloroform solution. J. Chem. Theory Comput. 2, 1085–1092 (2006).
19. Wolinski, K., Hinton, J. F. & Pulay, P. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. J. Am. Chem. Soc. 112, 8251–8260 (1990).
20. Pankratyev, E. Y., Tulyabaev, A. R. & Khalilov, L. M. How reliable are GIAO calculations of 1H and 13C NMR chemical shifts? A statistical analysis and empirical corrections at DFT (PBE/3z) level. J. Chem. Comput. Sci. 32, 993–1997 (2011).
21. Pulay, P., Hinton, J. & Wolinski, K. Nuclear magnetic shieldings and molecular structure 234–236 (Springer, Dordrecht, 1993).
22. Grimblat, N. & Sarotti, A. M. Computational chemistry to the rescue: Modern toolboxes for the assignment of complex molecules by NMR GIAO calculations. Chem.-A Eur. J. 22, 12246–12261 (2016).
23. Schmidt, M. & Rutzenlng, W. Theory of magnetic susceptibilities and NMR chemical shifts in terms of localized quantities: IV. Some small molecules with multiple bonds (N2, HCN, CO, CH3O, N2O, O3, FNO). Mol. Phys. 48, 781–798 (1983).
24. Keith, T. & Bader, R. Calculation of magnetic response properties using atoms in molecules. Chem. Phys. Lett. 194, 1–8 (1992).
25. Keith, T. A. & Bader, R. F. Calculation of magnetic response properties using a continuous set of gauge transformations. Chem. Phys. Lett. 210, 223–231 (1993).
26. Claridge, T. D. High-resolution NMR techniques in organic chemistry Vol. 27 (Elsevier, Amsterdam, 2016).
27. Lobl, S. & Schütz, M. NMR shielding tensors for density fitted second-order Muller-Plesset perturbation theory using gauge including atomic orbitals. J. Chem. Phys. 117, 084107 (2002).
28. Teale, A. M., Lutnaes, O. B., Helgaker, T., Tozer, D. J. & Gauss, J. Benchmarking density-functional theory calculations of NMR shielding, spin–rotation constants, and spin–rotation coupling constants using accurate coupled-cluster calculations. J. Chem. Phys. 138, 024113 (2013).
29. Lee, C., Wang, Y. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B 37, 785 (1988).
30. Barreiro, E. J., Rodrigues, C. R., Albuquerque, M. G., Sant’Anna, C. M. R. D. & Alencastro, R. B. D. Modelagem molecular: Uma ferramenta para o planejamento racional de fármacos em química medicinal. J. Org. Chem. 79, 73–91 (2016).
31. Konstantinov, I. A. & Broadbelt, L. J. Regression formulas for density functional theory calculated 1H and 13C NMR chemical shifts in toluene-d8. J. Phys. Chem. A 115, 12364–12372 (2011).
32. Stoyanov, S. R., Gusarov, S., Kuznitski, S. M. & Kovalenko, A. Theoretical modeling of zeolite nanoparticle surface acidity for heavy oil upgrading. J. Phys. Chem. C 112, 6794–6810 (2008).
33. Griessen, R. et al. Fully automated quantum-chemistry-based computation of spin–spin-coupled nuclear magnetic resonance spectra. Angew. Chem. Int. Ed. 56, 14763–14769 (2017).
34. Stochev, G. L., Auer, A. A. & Neese, F. Efficient and accurate prediction of nuclear magnetic resonance shielding tensors with double-hybrid density functional theory. J. Chem. Theory Comput. 14, 4756–4771 (2018).
35. Iron, M. A. 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. 13, 5798–5819 (2017).
36. Pipers, G. K., Venkatachal, T. & Reutens, D. C. NMR and DFT investigations of structure of colchicine in various solvents including density functional theory calculations. Sci. Rep. 7, 1–9 (2017).
37. Touw, S. L., de Groot, H. J. & Buda, F. DFT calculations of the 1H chemical shifts and 13C chemical shift tensors of retinal isomers. J. Mol. Struct. (Theochem) 711, 141–147 (2004).
42. Ditchfield, R. Self-consistent perturbation theory of diamagnetism: I. A gauge-invariant LCAO method for NMR chemical shifts. *Mol. Phys.* **27**, 789–807 (1974).
43. Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).
44. Parr, R. & Yang, W. *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
45. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoret. Chem. Acc.* **120**, 215–241 (2008).
46. Frisch, M. J., Pople, J. A. & Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **80**, 3265–3269 (1984).
47. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
48. Tomasi, J., Mennucci, B. & Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999–3094. https://doi.org/10.1021/cr9900909 (2005).
49. Gaussian09, R. A. 1, mj frisch, gw trucks, hb schlegel, ge scuseria, ma robb, jr cheeseman, g. Scalmani, v. Barone, b. Mennucci, ga petersson et al., gaussian. *Inc., Wallingford CT* **121**, 150–166 (2009).
50. Chamkin, A. A. Benchmarking DFT calculations of 1H and 13C chemical shifts in monosubstituted ferrocenes. *Int. J. Quantum Chem.* **121**, e26456 (2021).

**Acknowledgements**
Nuha Wazzan gratefully acknowledges the High-Performance Computing Centre (Aziz Supercomputer) at King Abdulaziz University (http://hpc.kau.edu.sa) for assisting with the calculations for this study.

**Author contributions**
Z.S. Conceptualization; Data curation; Formal analysis; Investigation; Methodology; statistical analysis Writing—original draft. N.W. Carrying out DFT calculations, Formal analysis, review, and editing. All authors read, revised, and approved the final manuscript.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22900-y.

**Correspondence** and requests for materials should be addressed to Z.S.S.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022