Evaluating the accuracy of density functional theory for calculating 1H and 13C NMR chemical shifts in drug molecules

ARTICLE *in* COMPUTATIONAL AND THEORETICAL CHEMISTRY · NOVEMBER 2014

Impact Factor: 1.55 · DOI: 10.1016/j.comptc.2014.11.007

3 AUTHORS, INCLUDING:

Jason P Holland
Harvard Medical School
64 PUBLICATIONS **1,440** CITATIONS

SEE PROFILE

Computational and Theoretical Chemistry 1051 (2015) 161–172

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

journal homepage: www.elsevier.com/locate/comptc

Evaluating the accuracy of density functional theory for calculating ^{1}H and ^{13}C NMR chemical shifts in drug molecules

David E. Hill [a,b], Neil Vasdev [a], Jason P. Holland [a,*]

[a] Division of Nuclear Medicine and Molecular Imaging, Massachusetts General Hospital, Department of Radiology, Harvard Medical School, 55 Fruit St., White 427, Boston, MA 02114, United States

[b] Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, United States

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 7 November 2014
Accepted 7 November 2014
Available online 20 November 2014

Keywords:
NMR
Shielding constants
Chemical shifts
Density functional theory
Ispinesib
Kinesin spindle protein inhibitors

ABSTRACT

The accuracy of different DFT methodologies for calculating ^{1}H and ^{13}C NMR chemical shifts in (*R*)-ispinesib, a complex drug molecule with multiple chemical groups and one stereocentre, has been evaluated. The accuracy of 6 basis sets and 16 different XC functionals was tested. In addition, we present a detailed study on the role of geometry optimisation (in gas and solution phase using a chloroform polarisable continuum model) on the accuracy of the calculated NMR spectra. NMR calculations using the double-ζ basis sets DGDZVP and 6-31++G(d,p) were found to be more accurate (and computationally more efficient) than those using larger triple-ζ basis sets. The O3LYP/DGDZVP methodology in solution phase using a geometry optimised at the same level of theory was found to be the most accurate method with mean absolute errors (MAEs) for ^{1}H and ^{13}C chemical shifts of 0.174 ppm and 3.972 ppm, respectively. Irrespective of the choice of XC or basis set used, complete geometry optimisation in either gas or solvent phase was found to be essential for attaining the highest accuracy in both ^{1}H and ^{13}C calculated chemical shifts. Finally, the role of molecular conformation was examined by calculating the Boltzmann-weighted ^{1}H and ^{13}C chemical shifts. Overall, we demonstrate that DFT shows exceptional promise for use in calculating the NMR chemical shifts in complex drug molecules. In the future, DFT calculations of NMR parameters are set to play an increasingly important role in drug discovery and chemical optimisation.

1. Introduction

NMR spectroscopy is one of the most widely used characterisation tools for monitoring chemical reactions and assigning the molecular structure of isolated compounds. Indeed, the application of NMR spectroscopy, in combination with other techniques including high resolution mass spectrometry and liquid chromatography, has in many ways circumvented the need for chemists to characterise their products using traditional methods such as elemental analysis and chemical functional group reactivity. Given the central role of NMR spectroscopy in modern chemistry, it is not surprising that theoreticians and computational chemists have had a longstanding interest in developing advanced tools to predict *ab initio* the local magnetic properties of small molecules including isotropic shielding constants (σ_{iso}) and spin–spin coupling constants (J/Hz) [1,2], as well as other properties like Electron Spin Resonance (ESR) *g*-tensors and hyperfine coupling constants [3–5]. Detailed expositions of the various theories that underpin the calculation of magnetic properties have been the subject of several excellent articles [6–10]. Examples of diverse chemical problems which have been addressed using NMR calculations include: analysis of rotational barriers in neutral and protonated furfurylidenanilines [11], studies on vanadium(V) dipicolinate complexes [12,13], work on proton exchange reactions in porphyrins [14], intramolecular hydrogen bonding [15], studies on the ^{1}H and ^{13}C NMR chemical shifts in transition metal hydrides [16], substituted hydrocarbons [17] and fullerenes [18,19], configurational assignments in cellulose [20], 6β-hydroxyhyoscyamine diastereoisomers [21], lactones [22] and lactams [23], and chemical shift assignment in natural products [24] and in drug molecules like Taxol [25].

Early efforts using quantum chemistry to predict NMR chemical shifts and spin–spin coupling constants employed the *Individual Gauges for Localised Orbitals* (IGLO) method [26]. Malkin et al. [27–29] were amongst the first to report chemically insightful results in calculating NMR properties of small molecules with up to four heavy atoms ($Z > 1$) using density functional theory combined with the IGLO method. Subsequent benchmark studies [9,30–35] employed the *Gauge-Including Atomic Orbitals* (GIAO)

* Corresponding author. Tel.: +1 (617) 726 6107; fax: +1 (617) 726 6165.
E-mail addresses: holland.jason@mgh.harvard.edu, jasonpholland@gmail.com (J. P. Holland).

http://dx.doi.org/10.1016/j.comptc.2014.11.007
2210-271X/

formalism – a concept originally propose by London [36] in 1937, and developed further in the 1990s [37,38]. Although other methods including the *Continuous Set of Gauge Transformations* (CSGT), the Single Origin method, and IGAIM (a variant of CSGT) are available in commercial software such as Gaussian09 [39], GIAO combined with DFT has emerged as a valuable method for computing NMR properties of small organic and inorganic molecules [40]. However, methods including CCSD(T) or MP2 methods with complete basis sets remain the gold standard for NMR calculations [41].

In the last two decades, several groups have reported DFT studies on the accuracy of exchange–correlation functionals (XC), basis sets [35], geometry and the inclusion of solvation models [42] in calculating chemical shifts and coupling constants for ^{1}H, ^{13}C, and other NMR active nuclei including metals [9,30–34]. A preponderance of these reported benchmark studies typically focused on the calculation of small organic molecules with ca. <10 heavy atoms. However, it is important to note that technological advances in computational hardware and software mean that it is now feasible to perform NMR calculations on molecules with MW > 500 Da and more than 50 heavy atoms. In addition, specialised tools like the relatively small pc-S basis sets [41,43] have expedited the use of NMR calculations. As a consequence, NMR calculations are now a practical tool in synthetic and medicinal chemistry that can be applied to resolve experimental problems such as configurational assignment of natural products or optimisation of molecular geometry for improving the stability, binding affinity and specificity of drugs [44,45]. In light of this, addressing the question of which DFT methodologies are optimal for calculating NMR properties, particularly of larger, structurally complex drug molecules, is paramount.

Recently, Willoughby et al. [46] published, "*A guide to small-molecule structure assignment through computation of* (^{1}H and ^{13}C) *NMR*". This protocol provides an excellent overview to the steps required for calculating NMR spectra and will undoubtedly encourage the wider chemical community to explore the use of DFT to solve experimental problems of increasing complexity.

Here, we present a systematic evaluation of the accuracy of DFT calculated ^{1}H and ^{13}C NMR chemicals shifts in (*R*)-ispinesib – an anti-mitotic kinesin spindle protein (KSP) allosteric inhibitor currently under evaluation in multiple clinical trials against various cancers [47–50]. The accuracy of 6 basis sets and 16 different XC functionals, as well as the role of solvation using a continuum model, and the influence of the molecular geometry and conformation on the calculated ^{1}H and ^{13}C NMR chemical shifts are reported.

2. Computational details

All calculations were conducted using density functional theory (DFT) as implemented in the Gaussian09 Revision B.01 suite of *ab initio* quantum chemistry programs [39]. Normal self-consistent field (SCF) and geometry convergence criteria were employed throughout, and structures were optimised in either gas or solution phase without the use of symmetry constraints. Harmonic frequency analysis based on analytical second derivative was used to characterise optimised gas phase structures as local minima on the potential energy surface.

Geometry optimisations, vibrational frequencies and NMR shielding tensors were calculated using a variety of exchange–correlation (XC) functionals and basis sets. NMR calculations employed the Gauge-Independent Atomic Orbital (GIAO) method with Gaussian09 default parameters. The calculations performed were designed to facilitate a full evaluation of the effects of XC functional, basis set, geometry, and solvation on the DFT calculated NMR shielding constants. In all cases ^{1}H and ^{13}C isotropic shielding constants (σ_{iso}) for the test molecules were corrected to chemical shift (δ_{iso}/ppm) relative to computed values for tetramethylsilane (TMS; C_1-symmetry) calculated at the same level of theory (i.e. using the same methodology including XC functional [XC], basis set [β] and geometry optimisation [G_n] in gas phase [s = gas] or solution [s = sol] for each independent calculation) in accordance with Eq. (1). Note that in all calculations, the ^{1}H and ^{13}C NMR isotropic shielding constants σ_{iso} for TMS were calculated as the mean for the 12 proton atoms and 4 carbon atoms.

$$\delta_{iso} = \sigma_{iso}(\mathrm{TMS;\ XC;}\ \beta;\ G_s) - \sigma_{iso}(\mathrm{Test;\ XC;}\ \beta;\ G_s) \quad (1)$$

The 16 XC functionals investigated can be divided into four categories: (i) Pure Generalised gradient approximation (GGA) functionals (where $X = 0$) included the first-generation BLYP [51,52], BP86 [52–54], BPW91 [52,55,56], and second-generation functional BPBE [52,57]. (ii) Hybrid-GGA functionals (where $X > 0$) included the first-generation B1LYP [51,52,58], B3LYP [51,52,59], B3P86 [52–54], B3PW91 [52,55,60], B98 [61,62], O3LYP [51,63,64], X3LYP [51,65] PBE1PBE (also known as PBE0) [31,57,66], and THCTHH [67]. (iii) Pure meta-GGA functionals included BTPSS [52,68] and VSXC [69]. (iv) The hybrid meta-GGA TPSS1KCIS [68,70,71] (TPSSKCIS and Iop(3/76 = 0870001300)). The origins, nomenclature and characteristics of the XC functionals studied have been described elsewhere [72–74].

Basis sets evaluated included all-electron double-ζ DGDZVP [75,76], and Pople 6-31++G(d,p) basis sets, as well the triple-ζ Pople basis sets augmented with additional polarisation functions 6-311++G(d,p), 6-311++G(2d,2p), 6-311++G(2df,2pd) and 6-311++G (3df,3pd) [77,78].

The effects of solvation were incorporated [79,80] iteratively by performing self-consistent reaction field (SCRF) calculations using the integral equation formalism polarisable continuum model (IEFPCM) initially developed by Tomasi and co-workers [81]. The solute–solvent boundary was defined by using a solvent excluding surface (SES) [82]. The molecular solute surface was defined by using the United Atom Topological model (UAHF) for the radii of the solute atoms [58]. Where specified, we used a chloroform solvent continuum model with dielectric constant, $\varepsilon = 4.7113$, and solvent sphere radius, $R_{solv} = 2.4800$ Å. Calculation of solvation free energies, ΔG_{sol}/kJ mol^{-1}, employed default settings [39]. The choice of solvation model reflects the solvent used in the experimental NMR spectra of the compound studied ($CDCl_3$) [47,83]. Optimised structures were analysed by using Chemcraft (version 1.7, build 365). ChemDraw Pro 12.0 (CambridgeSoft, Waltham, MA) was used to "predict" the ^{1}H and ^{13}C NMR chemical shifts of ispinesib for comparison with the DFT calculations.

The influence of molecular conformation on the calculated ^{1}H and ^{13}C NMR chemical shifts was also evaluated. Conformers of ispinesib starting from 15 different initial input geometries were optimised at the O3LYP/DGDZVP level of theory in chloroform. The calculated solvated free energies (ΔG_{sol}/kJ mol^{-1}) and calculated values of the root-mean squared displacement (RMSD/Å) between the optimised geometries were compared and we identify 12 unique conformational minima. Single point NMR calculations (also using O3LYP/DGDZVP in chloroform) were then performed. The calculated NMR spectra for each conformer were compared to the experimental spectra using a variety of statistical analysis tools (*vide infra*). In addition, the Boltzmann-weighted chemical shifts were calculated and used to generate the equilibrium NMR spectra (at 298 K) taking into account the relative contribution from the 12 unique conformers identified [46]. The Boltzmann-weighted chemical shifts were calculated by using Eq. (2), where in a system at thermal equilibrium at temperature T, P_i is the probability (also known as the Boltzmann factor) of occupying state i with energy E_i. The summation in the denominator is the canonical mole partition function summed over n states (here denoting different molecular conformers). The energy difference between conformers,

ΔE_i/kJ mol^{-1}, was defined using solvation free energies relative to the lowest energy conformer (Conformer 7; see Supporting Information Table S12).

$$P_i = \frac{e^{-\Delta E_i/k_B T}}{\sum_{i=1}^{n} e^{-\Delta E_i/k_B T}} \quad (2)$$

The Boltzmann-weighted NMR chemical shifts, δ_j (Boltz)/ppm, for resonance j, accounting for different conformational contributions at thermal equilibrium (T = 298.15 K) were calculated in accordance with Eq. (3). Similar methods were employed by Willoughby et al. [46]

$$\delta_j\ (\text{Boltz}) = \sum_{i=1}^{n} P_i \delta_i \quad (3)$$

Statistical analyses were performed using GraphPad Prism 5.01 (GraphPad Software, Inc., La Jolla, CA) and Microsoft Excel spreadsheets. The quality or "goodness of fit" between the calculated and experimental NMR spectra was analysed by using a variety of statistical methods. Statistical analyses used included the mean absolute error (MAE/ppm; Eq. (4)), where j is the number of unique chemical shifts used in the comparison; the mean squared error (MSE/ppm^2; Eq. (5)); the root-mean squared error (RMSE/ppm; Eq. (6)); and Chi-squared (χ^2/ppm; Eq. (7)).

$$\text{MAE} = \frac{1}{j}\sum_{i=1}^{j}\left|\delta_i^{calcd} - \delta_i^{expt}\right| \quad (4)$$

$$\text{MSE} = \frac{1}{j}\sum_{i=1}^{j}\left(\delta_i^{calcd} - \delta_i^{expt}\right)^2 \quad (5)$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{j}\sum_{i=1}^{j}\left(\delta_i^{calcd} - \delta_i^{expt}\right)^2} \quad (6)$$

$$\chi^2 = \sum_{(i=1)}^{j}\left(\left(\delta_i^{calcd} - \delta_i^{expt}\right)^2 \Big/ \left(\delta_i^{calcd}\right)\right) \quad (7)$$

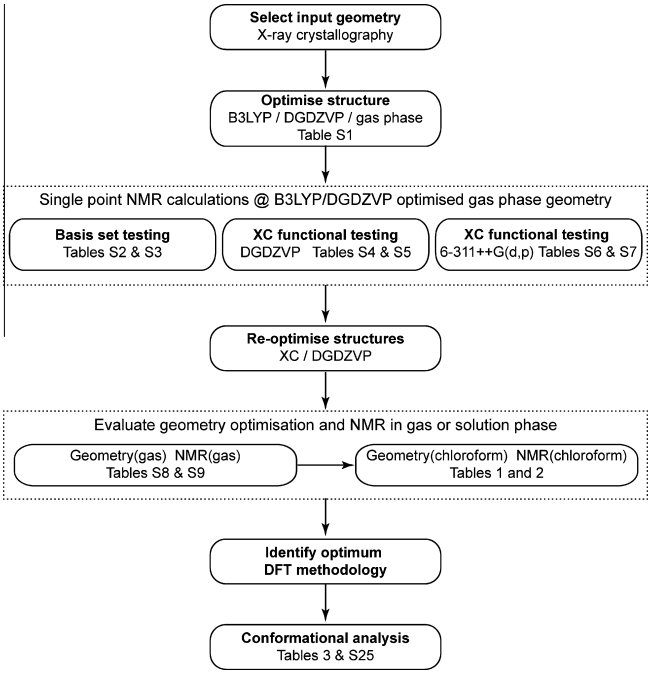
In addition to the MAE, MSE, RMSE and χ^2 analyses, least squares linear regression analysis was used to compare the calculated *versus* experimental ^{1}H and ^{13}C chemical shifts. Outliers were identified using the Robust regression and OUTlier removal procedure (ROUT), and where identified, were removed from the linear regression and statistical analyses [84]. The ROUT method is adapted from the false discovery rate approach. The recommended default Q-value of 1% with equal weighting for each individual chemical shift, j, was used throughout. Note that this choice of Q-value implies that the outliers are identified with a false discovery rate of less than 1% [84].

3. Results and discussion

To facilitate a clear understanding of the different calculations performed in this work, an overview of the data sets, with reference to their associated tables, is presented in Scheme 1.

3.1. Compound and geometry selection

To evaluate the potential of using NMR calculations to predict the ^{1}H and ^{13}C spectra of drug molecules we selected (*R*)-ispinesib, a quinazolinone-based anti-mitotic drug, as a model compound (Fig. 1). Ispinesib is a potent and highly selective allosteric inhibitor of kinesin spindle protein (KSP) [47]. KSP is expressed during mitosis and is the motor protein responsible for the separation of the centrosomes – a vital step in mitotic spindle formation.



Scheme 1. Schematic overview of the calculations used to evaluate the accuracy and contribution of different XC functionals, basis sets, geometry optimisation in gas and solvation phase, and molecular conformation on the calculated ^{1}H and ^{13}C NMR chemical shifts of ispinesib.

Inhibition of KSP in cells prevents centrosome separation leading to mitotic phase arrest and induced apoptosis *via* monopolar (monoaster) spindle formation [47]. Molecules targeting components of cellular mitosis are an important branch of anti-proliferative and anti-cancer drug development. Thus, understanding the molecular properties and interactions of drugs like ispinesib using experimental and computational NMR methods may facilitate the development of a new generation of anti-mitotic agents with improved efficacy.

Ispinesib ($C_{30}N_4H_{33}O_2Cl$; MW = 517.1 g mol^{-1}) represents an interesting and complex model compound for evaluating the accuracy of ^{1}H and ^{13}C NMR calculations. The molecule contains 16 chemically distinct protons (labelled a–p; Fig. 1A) with 19 magnetically distinct chemical shifts including aliphatic methyl, methylene and methane groups, as well as homocyclic and heterocyclic aromatic rings. In addition, there are 30 carbon atoms with 26 chemically and magnetically distinct resonance environments (labelled 1* and 1–29; Fig. 1B). Ispinesib contains one stereocentre (atom C3; Fig. 1B) and only the (*R*)-enantiomer is biologically active as a KSP inhibitor [47]. In the ^{1}H experimental NMR spectrum (recorded at 298 K in $CDCl_3$; Fig. 2) the presence of this stereocentre causes the methylene protons of the benzyl group (g and g*), as well as those of the central methylene group (n and n*) to be diastereotopic. The two methylene protons at position m are also diastereotopic, however, the resonance peaks partially overlap which makes identifying the centre of the individual proton resonances difficult. For this reason, calculated ^{1}H NMR chemical shifts for the two protons at position m were taken as the average. Furthermore, the calculated ^{1}H and ^{13}C resonances for each of the *ortho* and *meta* protons on the benzyl and toluoyl groups were taken as the average of the two positions (Fig. 1A and B). The experimental ^{13}C{^{1}H} NMR spectrum of (*R*)-ispinesib is shown in Supporting Information Fig. S1 [47].

Prior to full conformational analysis (see Section 3.6 *vide infra*) we used the experimental X-ray crystallographic geometry of (*R*)-ispine-

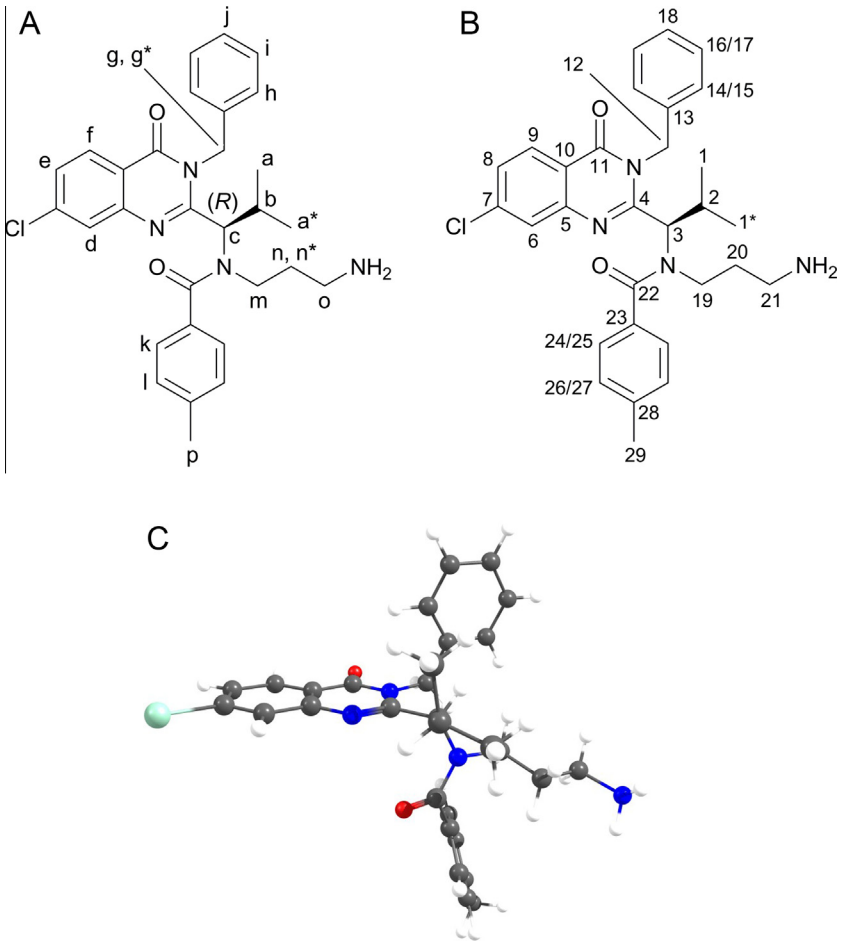


Fig. 1. Chemical structure of (*R*)-ispinesib used as a test compound for evaluating the accuracy ^{1}H and ^{13}C NMR calculated chemical shifts using DFT methods. (A) Labelling scheme for the ^{1}H proton resonances. (B) Labelling scheme for the ^{13}C carbon resonance. (C) Picture of the experimental X-ray crystallographic structure of ispinesib (in the conformation that shows optimal binding to human KSP) used as the initial input geometry for testing different XC functionals, basis sets and geometry optimisation in gas phase and solvated in chloroform. Note the orientation is aligned parallel to the C2–C3 bond.

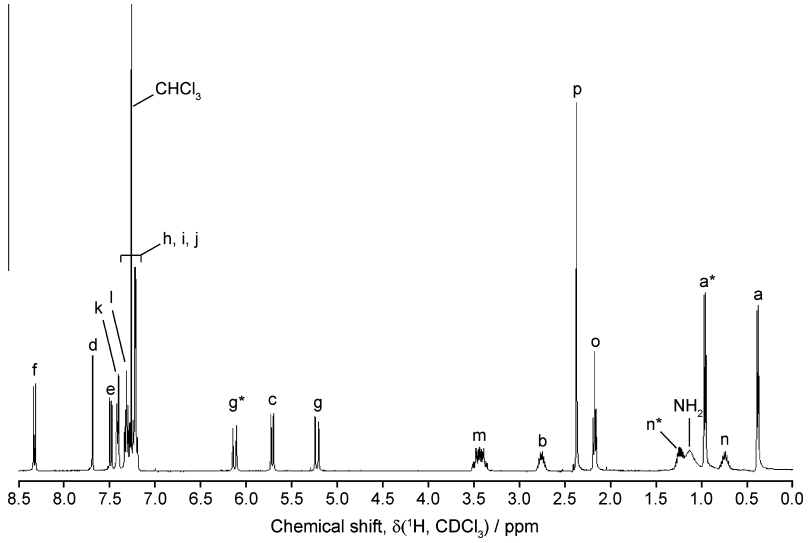


Fig. 2. Experimental ^{1}H NMR spectrum of (*R*)-ispinesib recorded on at 400 MHz in $CDCl_3$ at 298 K [47]. Resonance labels correspond to the protons shown in Fig. 1A.

sib (Fig. 1C and Table S1), taken from a crystal structure [85] of the drug bound to human KSP, as initial input coordinates for testing the accuracy of different XC functionals, basis sets, geometry optimisation in both the gas phase and a chloroform continuum solvation model.

3.2. Basis set dependence

Starting from the X-ray crystal structure Cartesian coordinates of (*R*)-ispinesib, we first optimised the geometry in gas phase using B3LYP/DGDZVP followed by single-point calculations of the ^{1}H and ^{13}C NMR chemical shifts (Tables S2 and S3, respectively) using B3LYP with increasing the basis set size. In total, six basis sets were evaluated with the number of Gaussian basis functions ranging from 588 (DGDZVP) to 2045 (6-311++G(3df,3pd)). ^{1}H and ^{13}C line spectra are shown in Figs. 3 and S2, respectively, with associated plots of the statistical errors *versus* the number of basis functions presented in Figs. S3 and S4. Overall, the calculations provided a qualitatively accurate description of the ^{1}H NMR spectrum of (*R*)-ispinesib. The ^{1}H chemical shifts of all methyl, methylene and aromatic protons were found to lie within 0.20 ppm of the experimental shifts. Exceptions were noted for the methane protons b and c, and the diastereotopic protons g and g*. Resonance b (δ_H(expt) = 2.75 ppm) was consistently underestimated with differences in calculated *versus* experimental chemical shifts ($\Delta\delta_H$) in the range 1.42–1.6 ppm. Resonance c (δ_H(expt) = 5.71 ppm) was found to be dramatically underestimated with $\Delta\delta_H$ in the range −2.27 to −2.63 ppm. The chemical shift difference between resonances g and g* was also overestimated by the DFT calculations: experimental $\Delta\delta_H$ (g*–g; expt) was 0.89 ppm but calculated difference for B3LYP/DGDZVP was 2.19 ppm. Evidently, the strong performance of the ^{1}H NMR calculations in the majority of the aliphatic and aromatic proton environments is negated for resonances b, c g and g* in ispinesib by the presence of the stereocentre at atom C3. A *Q*-test analysis of the calculated ^{1}H NMR spectra using the ROUT method revealed that resonances b and c were outliers. As such, resonances b and c were removed from the statistical analysis throughout. It is striking that for the overall ^{1}H NMR calculated spectra, the smallest basis set (DGDZVP) produced the most accurate result (MAE = 0.243 ppm; MSE = 0.094 ppm^2; χ^2 = 0.513 ppm; Table S2). No improvement was observed on moving from double-ζ to triple-ζ basis sets.

For the calculated ^{13}C NMR spectra, carbon atoms C3 and C19 were found to be outliers using the ROUT method. These resonances were subsequently removed from further statistical analysis. Overall, the calculated ^{13}C NMR spectra are in excellent quantitative and qualitative agreement with the experimental spectrum (Fig. S2 and Table S3). Due to the larger chemical shift range for ^{13}C NMR, the calculated errors MAE, MSE, RMSE and χ^2 for the ^{13}C spectra (Fig. S4) are considerably larger than those calculated for the ^{1}H spectra (Fig. S3). In contrast to the ^{1}H NMR data, DGDZVP was found to be the least accurate basis set for calculating the ^{13}C NMR spectrum (MAE = 11.28 ppm; MSE = 151.3 ppm^2; χ^2 = 12.3 ppm), with 6-31++G(d,p) performing the best (MAE = 4.35 ppm; MSE = 39.04 ppm^2; χ^2 = 6.25 ppm; Table S3). Only a slight difference was found between DGDZVP and the triple-ζ 6-311++G-type basis sets in the calculated ^{13}C chemical shifts. Taken together, the statistical analysis demonstrates that the smaller basis sets DGDZVP and 6-31++G(d,p) outperform the larger triple-ζ basis sets for NMR calculations on (*R*)-ispinesib.

3.3. Evaluation of XC functionals

Sixteen different XC functionals were evaluated for their accuracy in calculating ^{1}H and ^{13}C chemical shifts. As with the basis set evaluation (*vide supra*), single point NMR calculations were first performed using the B3LYP/DGDZVP optimised gas phase geometry of (*R*)-ispinesib. In addition, the performance of all XC functionals was evaluated using both DGDZVP and 6-311++G(d,p) basis sets. An overview of the statistical analysis for each data set is shown in Fig. 4 (white bars = DGDZVP; red bars = 6-311++G(d,p)) and complete ^{1}H and ^{13}C data sets are presented as Supporting Information (Tables S4–S7). The first point to note is that calculations using different XC functionals with the DGDZVP basis set outperform those equivalent calculations using the 6-311++G(d,p) basis set in all cases. These data provide further confirmation of the observations made and conclusions drawn from the basis set data (*vide supra*) in that expanding the number of basis functions used actually decreases the accuracy of the ^{1}H chemical shift calculations, irrespective of the XC functionals employed. For this reason the following discussion concentrates on the DGDZVP data sets.

Across all 16 XC functionals tested, the MAE for the calculated ^{1}H NMR spectra ranged from 0.20 ppm (X3LYP) to 0.32 ppm (TPSS1KCIS; Table S4; Fig. 4[left]). Interestingly, the pure DFT functionals, BLYP (ranked 2nd based on MAE), BPW91 (4th), BPBE (5th), BTPSS (6th) and BP86 (7th) generally outperformed the hybrid functionals with the exception of X3LYP and O3LYP (ranked 1st and 3rd respectively, with MAE values of 0.200 ppm and 0.222 ppm).

For the calculated ^{13}C NMR spectra using the DGDZVP basis set, MAE values ranged from 6.107 ppm to 11.461 ppm. Again, pure DFT methods performed well, generally providing more accurate results than many of the first and second generation hydrid DFT methods like B3LYP (Table S5; Fig. 4[right]). One notable exception was BLYP (^{13}C MAE = 8.245 ppm) which slipped to 8th place in rank order. For the ^{13}C data, O3LYP (^{13}C MAE = 6.107 ppm) emerged as the most accurate XC functional ahead of BPBE and X3LYP (MAE values of 7.249 ppm and 7.305 ppm, respectively). Interestingly, unlike in the ^{1}H data, use of the larger triple-ζ basis

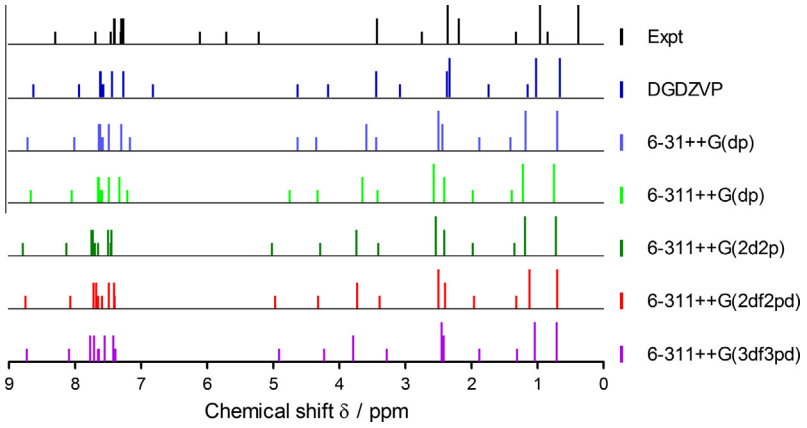


Fig. 3. Line spectra showing the basis set dependence in the calculated ^{1}H NMR chemical shifts. Note: the plot shows the delta functions corresponding to the isotropic chemical shifts (δ_{iso}/ppm) for single point NMR calculations using the gas phase B3LYP/DGDZVP optimised geometry of ispinesib.

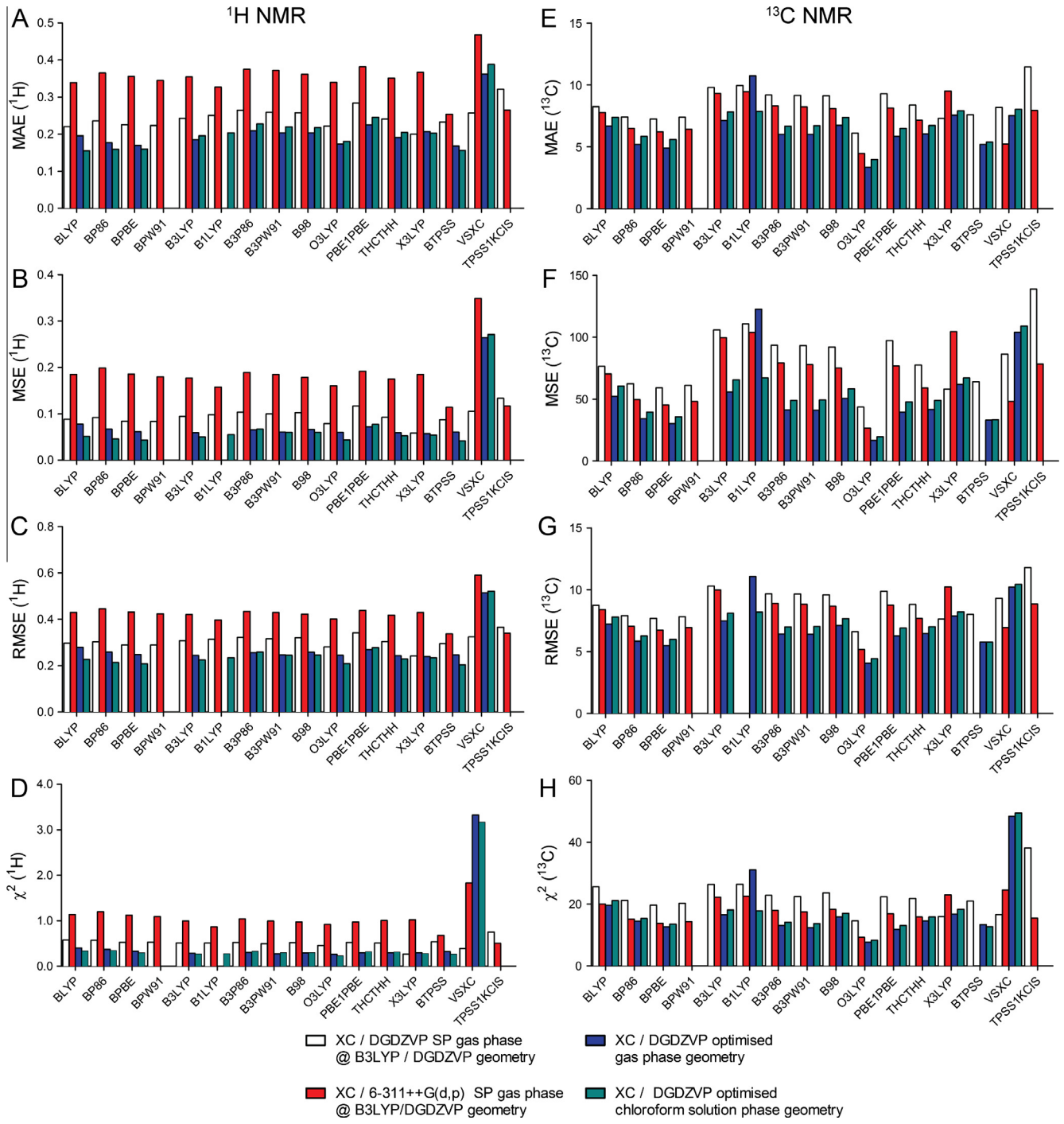


Fig. 4. Bar charts summarising the statistical analyses in the calculated ^{1}H (left) and ^{13}C (right) NMR spectra of (*R*)-ispinesib using different XC methodologies, basis sets, geometry optimisation and gas *versus* solution phase data sets. Charts (A) and (E) compare the MAE values; (B) and (F) compare the MSE values; (C) and (G) compare the RMSE values; and (D) and (H) compare the calculated χ^2 values. (For interpretation to colors in this figure, the reader is referred to the web version of this paper.)

set generally gave a very minor improvement in the calculated ^{13}C chemical shifts. However, we note that this increase in overall accuracy did not change the qualitative interpretation of the calculated ^{13}C NMR spectra.

Collectively, evaluation of different XC functionals in the gas phase at a fixed geometry revealed that across both ^{1}H and ^{13}C data sets, O3LYP provided the highest degree of quantitative accuracy in the calculated chemical shifts of (*R*)-ispinesib. In general, the smaller DGDZVP basis set provided a suitable balance between chemical accuracy and computational expedience.

3.4. Geometry optimisation

Next, we evaluated the contribution of full geometry relaxation in the gas phase to the accuracy of the calculated ^{1}H and ^{13}C NMR chemical shifts using the different XC/DGDZVP methodologies (Fig. 4; blue bars; Tables S8 and S9). Gas phase geometry optimisation led to a significant improvement in NMR spectra for all XC functionals tested though larger improvements were observed in the ^{13}C data compared to the ^{1}H data. For example, for the O3LYP/DGDZVP geometry optimised gas phase calculations, the

MAE values decreased from 0.222 ppm to 0.173 ppm for the calculated ^{1}H NMR chemical shifts (compare the white *versus* dark grey bars in Fig. 4), and from 6.107 ppm to 3.342 ppm in the calculated ^{13}C NMR chemical shifts.

In addition, we also performed calculations in which the bond lengths of selected ^{1}H atoms were manually adjusted to identify the influence of molecular bond length on calculated ^{1}H chemical shifts. Starting from the optimised O3LYP/DGDZVP geometry in the presence of a chloroform solvation field, the carbon–hydrogen bond lengths for the ^{1}H atoms of resonances a, b, c, d, e, f, j and o (Fig. 1A) were manually altered by moving the hydrogen atoms to ±15% of the optimised bond length in incremental steps of 1.5%. Note that where two or more H atoms contribute to the resonance (as in the methyl group of resonance a, and the methylene group of resonance o), the bond length for each individual hydrogen atom was adjusted by the required amount in the same calculation. The data are presented in Supporting Information Fig. S5. The fully optimised O3LYP/DGDZVP in chloroform minimum structure is given at 0% change in bond length. It is striking that with the exception of the calculated average ^{1}H resonance a, a pronounce minimum was observed for resonances b, c, d, e, f, j and o precisely at the optimised geometry (0% change in bond length). In general, when the carbon–hydrogen bonds were either compressed (negative values for the percentage variation) or elongated (positive values), the calculated chemical shift increased in magnitude indicating that the protons became less shield. In addition, there was a general trend for all atoms in which bond elongation resulted in decreased shielding and a consequent increase in calculated ^{1}H chemical shift. These data are consistent with the expected chemical behaviour of a carbon–hydrogen bond during vibration. These data indicate that full geometry optimisation using the test XC/basis set methodology of choice is essential to achieving the most accurate results.

3.5. Solvation effects

In the final step of methodology evaluation, we introduced solvation effects using the integral equation formalism polarisable continuum model (IEF-PCM). The structures were first re-optimised at the different XC/DGDZVP methodologies in the presence of a chloroform IEF-PCM employing Gaussian09 default parameters for the cavity formation, solvent–solute molecular and surface definitions, and dielectric constant. The isotropic shielding constants for each atom were then recalculated in the presence of the chloroform solvation model. Data for the ^{1}H and ^{13}C NMR chemical shifts are presented in Tables 1 and 2 (with full data presented in the Supporting Information Tables S10 and S11) and an overview of the statistical analysis for each methodology is shown in Fig. 4 (green bars). Comparisons of the experimental, ChemDraw Pro 12.0 "predicted", and DFT (O3LYP/DGDZVP/chloroform) calculated ^{1}H and ^{13}C chemical shifts are presented in Fig. 5A and B. Correlation scatter plots between the calculated *versus* experimental ^{1}H and ^{13}C chemical shifts are presented in Fig. 6.

Overall, incorporation of solvation effects led to a small improvement in the calculated ^{1}H chemical shifts but a minor decrease in the accuracy of the calculated ^{13}C chemical shifts for the majority of the XC/DGDZVP methodologies tested. Notably for O3LYP, the proton MAE increased slightly on solvation to 0.180 ppm compared to 0.174 ppm in the gas phase calculations (Tables 1 and S10). However, on close inspection it was evident that the overall ^{1}H chemical shift MSE (0.044 ppm^2) and χ^2 (0.232 ppm) values improved in the presence of the solvent field compared to the gas phase values (O3LYP gas phase: MSE = 0.060 ppm^2; χ^2 = 0.266 ppm). In general, inclusion of the solvent led to an improvement in the accuracy of the individual calculated ^{1}H chemical shifts for proton resonances b, c, g and g* – the same proton environments that were found to give the least accurate calculated shifts in the gas phase studies compared with the experiment

Table 1
Summary of the calculated ^{1}H NMR chemical shifts and statistical parameters for the solution phase geometry optimised XC/DGDZVP methodologies. Full data including calculated $\Delta\delta$ values between DFT and experiment are presented in Supporting Information Table S10.

Resonance[a]	Expt. δ (ppm)	XC functional, calcd. δ_{iso} (ppm)													
		BLYP	BP86	BPBE	B3LYP	B1LYP	B3P86	B3PW91	B98	O3LYP	PBE1PBE	THCTHH	X3LYP	BTPSS	VSXC
TMS δ_{iso} (^{1}H) (ppm)	N/A	31.91	31.80	31.88	32.21	32.28	32.19	32.18	32.21	32.22	32.16	32.17	32.21	31.85	32.22
a*	0.96	0.94	0.97	0.96	0.95	0.96	0.97	0.97	0.96	0.96	0.95	0.96	0.95	0.94	0.96
a	0.38	0.53	0.60	0.55	0.58	0.58	0.61	0.60	0.38	0.51	0.59	0.61	0.57	0.53	0.31
b	2.75	3.83	3.90	3.91	3.81	3.79	3.89	3.88	2.75	3.80	3.87	3.84	3.81	3.88	3.73
c	5.71	3.21	3.19	3.24	3.03	3.00	3.03	3.06	5.71	3.18	2.97	3.08	3.01	3.22	2.75
d	7.69	7.65	7.74	7.75	7.92	7.96	8.02	8.00	7.69	7.85	8.06	7.93	7.94	7.73	7.82
e	7.46	7.44	7.49	7.52	7.64	7.67	7.72	7.71	7.46	7.64	7.77	7.64	7.66	7.50	7.91
f	8.30	8.22	8.30	8.31	8.50	8.54	8.59	8.57	8.30	8.43	8.65	8.50	8.53	8.29	8.48
g*	6.11	6.38	6.51	6.46	6.36	6.33	6.51	6.43	6.11	6.27	6.49	6.44	6.38	6.44	6.92
g	5.22	4.56	4.74	4.70	4.71	4.71	4.83	4.81	5.22	4.67	4.82	4.78	4.71	4.67	4.92
h	7.27	7.10	7.13	7.14	7.25	7.27	7.30	7.29	7.27	7.21	7.33	7.24	7.26	7.12	6.77
i	7.40	7.46	7.52	7.54	7.66	7.69	7.73	7.72	7.40	7.64	7.78	7.67	7.68	7.51	7.56
j	7.31	7.43	7.46	7.47	7.60	7.62	7.66	7.64	7.31	7.56	7.70	7.60	7.62	7.45	7.50
k	7.41	7.27	7.36	7.34	7.51	7.55	7.60	7.57	7.41	7.41	7.64	7.51	7.54	7.31	7.36
l	7.30	7.31	7.38	7.39	7.52	7.54	7.58	7.57	7.30	7.48	7.64	7.54	7.53	7.37	7.75
m	3.43	3.39	3.43	3.40	3.27	3.25	3.34	3.29	3.43	3.21	3.28	3.30	3.28	3.36	3.60
n*	1.33	1.68	1.72	1.66	1.63	1.62	1.65	1.62	1.33	1.56	1.61	1.64	1.63	1.63	2.34
n	0.85	1.18	1.17	1.15	1.02	1.00	1.03	1.02	0.85	1.06	0.99	1.08	1.02	1.11	0.69
o	2.19	2.18	2.13	2.12	2.09	2.09	2.07	2.05	2.19	2.08	2.03	2.07	2.10	2.09	0.80
p	2.36	2.22	2.27	2.25	2.31	2.32	2.35	2.33	2.36	2.29	2.35	2.32	2.32	2.23	2.60
Statistical analysis and robust least squares linear regression analysis															
MAE		0.156	0.160	0.160	0.196	0.204	0.228	0.220	0.218	0.180	0.246	0.205	0.203	0.156	0.388
MSE		0.051	0.046	0.043	0.050	0.055	0.067	0.060	0.060	0.044	0.077	0.053	0.055	0.042	0.271
RMSE		0.227	0.214	0.208	0.224	0.234	0.259	0.245	0.245	0.209	0.278	0.229	0.234	0.204	0.521
χ^2		0.341	0.349	0.299	0.274	0.279	0.332	0.302	0.304	0.232	0.323	0.313	0.282	0.270	3.170
M		0.975	0.982	0.989	1.015	1.020	1.024	1.024	1.023	1.010	1.036	1.012	1.017	0.990	1.028
C		0.119	0.141	0.100	0.021	0.005	0.035	0.013	0.011	0.002	0.017	0.050	0.019	0.070	0.059
R^2		0.994	0.995	0.995	0.995	0.995	0.995	0.996	0.995	0.995	0.995	0.995	0.995	0.995	0.971

[a] Resonances b and c were identified as outliers using ROUT analysis and these data points were removed from the statistical analysis.

Table 2
Summary of the calculated ^{13}C NMR chemical shifts and statistical parameters for the solution phase geometry optimised XC/DGDZVP methodologies. Full data including calculated $\Delta\delta$ values between DFT and experiment are presented in Supporting Information Table S11.

Resonance[a]	Expt. δ (ppm)	XC functional, calcd. δ_{iso} (ppm)													
		BLYP	BP86	BPBE	B3LYP	B1LYP	B3P86	B3PW91	B98	O3LYP	PBE1PBE	THCTHH	X3LYP	BTPSS	VSXC
TMS δ_{iso} (^{13}C) (ppm)	N/A	178.84	182.15	182.62	184.17	184.81	187.18	186.84	185.86	185.03	188.39	185.78	184.56	182.34	188.26
1*	19.33	26.90	26.59	25.72	26.21	26.01	25.98	25.39	26.10	24.15	25.41	26.22	26.28	25.44	38.46
1	18.46	26.24	26.16	25.60	25.79	25.61	25.70	25.34	25.99	24.46	25.32	26.16	25.79	25.37	38.30
2	29.03	37.57	33.98	33.46	33.46	33.08	30.38	30.63	33.22	32.57	29.08	33.18	33.08	33.33	40.79
3	59.79	88.62	85.44	84.55	83.70	82.93	81.14	81.01	82.78	80.41	79.28	83.15	83.30	84.44	89.58
4	156.05	164.43	161.58	161.16	165.18	165.51	163.25	163.27	163.93	158.92	162.86	162.77	165.27	160.96	163.03
5	140.89	153.51	151.53	151.13	154.92	155.29	153.39	153.47	154.37	149.22	153.37	153.30	155.09	150.91	147.60
6	127.01	134.82	133.67	133.82	135.98	135.93	135.19	135.37	135.28	132.15	135.29	134.63	136.14	133.64	129.55
7	119.50	123.16	121.45	121.56	124.11	124.24	122.81	123.21	123.90	119.81	122.69	123.01	124.10	121.38	116.90
8	128.20	133.57	132.53	132.67	134.71	134.69	133.99	134.19	134.16	131.02	134.17	133.42	134.88	132.49	127.99
9	129.06	135.45	134.48	134.56	137.46	137.64	136.84	136.97	137.00	133.18	137.12	136.15	137.69	134.34	131.26
10	147.80	158.51	155.49	155.17	157.99	158.08	155.80	156.01	157.85	152.36	155.23	156.84	157.98	155.02	152.93
11	162.36	168.51	165.73	165.45	170.87	171.47	168.90	168.88	168.67	164.06	168.70	167.44	171.11	165.29	164.83
12	45.68	54.54	52.39	51.96	51.51	51.02	49.85	49.82	50.96	49.52	48.56	51.29	51.23	51.73	60.66
13	136.99	141.95	140.41	140.29	143.34	143.62	142.19	142.34	143.37	138.66	142.37	142.26	143.52	140.07	143.95
14/15	127.42	131.29	130.11	130.35	132.70	132.82	131.92	132.11	132.35	128.90	132.24	131.46	132.94	130.14	128.33
16/17	129.37	134.80	134.05	134.01	136.74	136.92	136.29	136.36	136.59	132.50	136.77	135.63	137.04	133.80	133.00
18	127.81	132.84	132.06	132.22	134.68	134.84	134.24	134.42	134.49	130.62	134.66	133.60	134.92	132.03	129.81
19	42.27	66.55	64.66	63.86	63.10	62.54	61.53	61.26	62.44	60.68	60.22	62.84	62.89	63.63	74.32
20	33.37	46.44	45.28	44.17	44.00	43.64	43.08	42.51	44.05	41.87	42.32	44.32	44.08	43.96	60.81
21	39.31	50.74	48.74	48.22	47.54	47.12	45.87	45.83	47.26	46.90	45.21	47.39	47.39	48.02	53.57
22	173.19	180.50	177.87	177.69	183.31	183.93	181.41	181.49	180.96	176.60	181.31	179.60	183.55	177.57	178.79
23	133.96	142.11	140.02	140.09	142.89	143.04	141.44	141.83	142.79	138.73	141.45	141.70	142.92	139.80	138.48
24/25	126.15	133.61	132.37	132.26	135.70	135.95	134.71	134.82	135.11	131.20	135.01	134.08	135.94	132.04	132.88
26/27	128.88	134.66	133.96	134.05	136.32	136.38	135.76	135.99	136.03	132.58	136.12	135.21	136.52	133.95	132.35
28	139.68	146.71	145.68	145.38	148.97	149.38	148.21	148.30	148.96	143.38	148.42	147.84	149.23	145.05	147.26
29	21.53	27.49	26.83	26.54	25.99	25.72	25.64	25.58	26.26	25.30	25.19	26.44	25.93	26.15	31.90
Statistical analysis and robust least squares linear regression analysis															
MAE		7.384	5.840	5.587	7.813	7.865	6.658	6.710	7.369	3.972	6.495	6.717	7.898	5.383	8.026
MSE		60.688	39.485	35.704	65.678	67.163	49.042	49.510	58.548	19.703	47.791	49.159	67.348	33.400	109.077
RMSE		7.790	6.284	5.975	8.104	8.195	7.003	7.036	7.652	4.439	6.913	7.011	8.207	5.779	10.444
χ^2		21.108	15.340	13.527	18.105	17.779	14.124	13.710	16.997	8.335	13.028	15.801	18.266	12.699	49.403
M		0.986	0.981	0.985	1.017	1.023	1.015	1.018	1.011	0.984	1.022	1.000	1.020	0.986	0.892
C		8.977	7.929	7.258	6.048	5.553	5.150	4.890	6.301	5.778	4.300	6.770	5.895	7.000	19.035
R^2		0.998	0.998	0.999	0.999	0.998	0.998	0.999	0.998	0.999	0.998	0.998	0.998	0.999	0.991

[a] Resonances 3 and 19 were identified as outliers using ROUT analysis and these data points were removed from the statistical analysis.

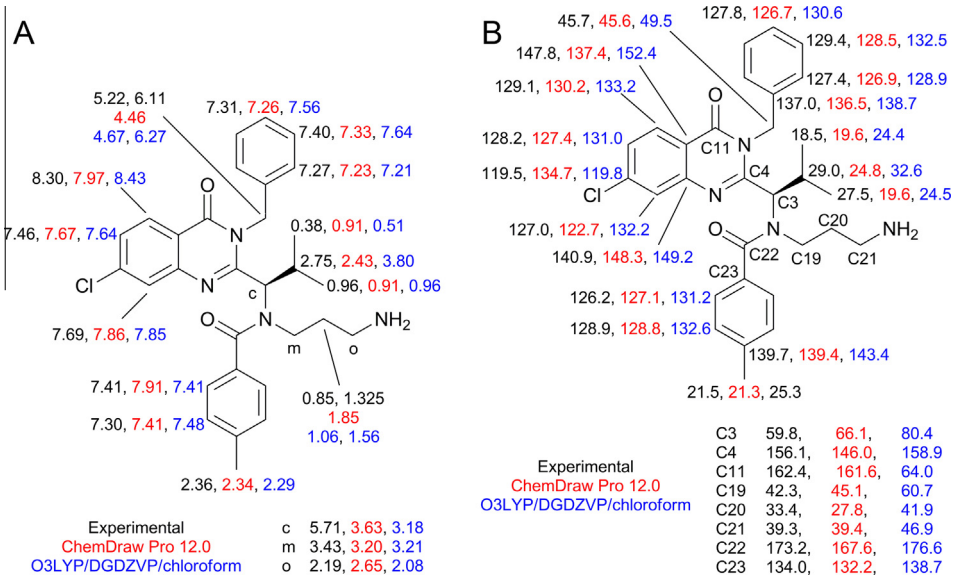


Fig. 5. Comparison of the experimental, ChemDraw Pro 12.0 "predicted" (italics) O3LYP/DGDZVP/chloroform calculated (underscored) and chemical shifts for (*R*)-ispinesib. (A) ^{1}H NMR, and (B) ^{13}C NMR data.

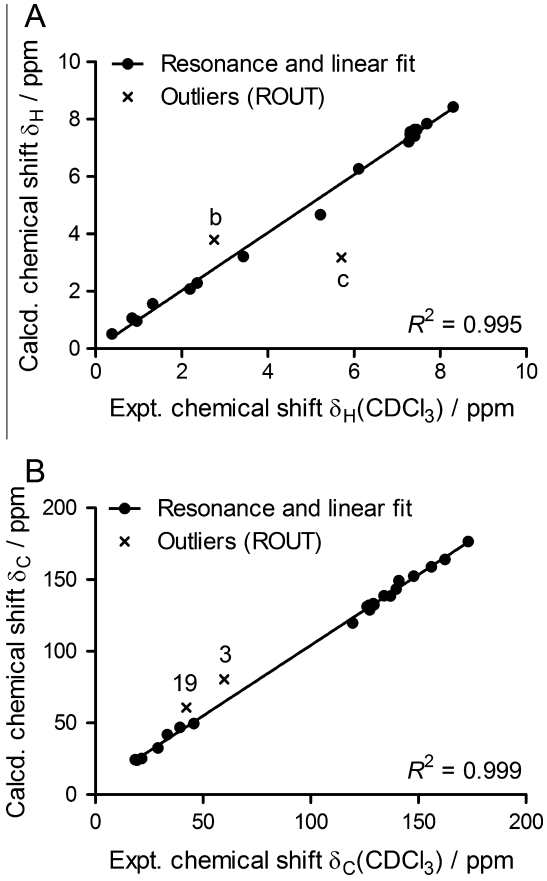


Fig. 6. Scatter plots and linear regression analyses showing the correlation between the calculated *versus* experimental chemical shifts for (A) the ^{1}H NMR and (B) the ^{13}C NMR spectra. Data presented are from the geometry optimised chloroform solvation model with the O3LYP/DGDZVP methodology.

(*vide supra*). This provides a strong indication that protons b and c, and the diastereotopic protons g and g* are influenced in solution by specific solvation effects (i.e. interactions beyond mere electrostatics) that are not explicitly included in the polarisable continuum model.

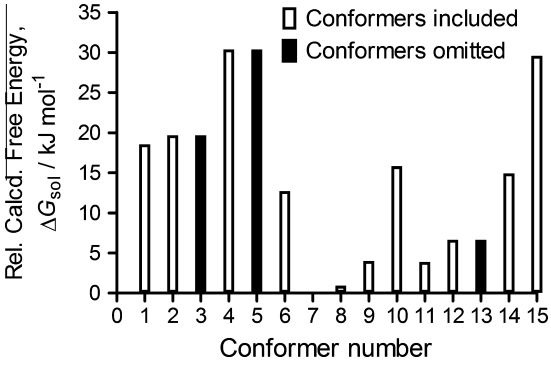


Fig. 7. Plot showing the relative calculated free energies (ΔG_{sol}/kJ mol^{-1}) for the different conformers of (*R*)-ispinesib optimised in chloroform using the O3LYP/DGDZVP methodology. Data sets for conformers shown in white bars were included and data sets in black (conformers 3, 5 and 13) were excluded from the in the Boltzmann analysis.

For the ^{13}C chemical shifts, the O3LYP methodology yielded the most accurate results of the different XC functionals (MAE = 3.972 ppm; MSE = 19.703 ppm^2; χ^2 = 8.335 ppm). Again, linear regression analysis and outlier detection (by the ROUT *Q*-test) demonstrated the excellent overall performance of DFT using O3LYP/DGDZVP in chloroform for calculating ^{13}C chemical shifts with a linear gradient of m = 1.010, a very small constant C = 0.002 ppm, and an excellent R^2 correlation coefficient of 0.999 (Fig. 6B).

Based on the collective assessment of the different basis sets, XC functionals, geometry optimisation and gas *versus* solution phase data, we conclude that for calculating the ^{1}H and ^{13}C NMR chemical shifts of complex drug molecules like (*R*)-ispinesib, which contain many different types of chemical resonance environments, full geometry optimisation in the presence of a solvent model combined with the O3LYP/DGDZVP methodology provides the most accurate results compared to our experimental data.

3.6. Conformational analysis: Boltzmann weighted equilibrium chemical shifts

It has been purported that calculated NMR properties are particularly sensitive to the molecular geometry employed [3,46].

Table 3
Calculated ^{1}H NMR chemical shifts of 12 different conformers using the optimised chloroform O3LYP/DGDZVP methodology.

		Conformer													
		1	2	4	6	7	8	9	10	11	12	14	15	Boltzmann-weighted	
Resonance	Expt. δ (ppm)	Calcd. δ_{iso} (ppm)												Calcd. δ_{iso} (ppm)	$\Delta\delta_{iso}$ (ppm)
a*	0.96	0.96	0.92	1.09	1.32	1.12	0.90	0.92	0.67	−0.14	0.80	0.73	1.06	0.90	−0.06
a	0.38	0.51	1.46	1.59	0.50	−0.07	−0.16	−0.13	0.77	0.92	1.12	1.51	1.02	0.04	−0.34
b	2.75	3.80	2.13	2.37	2.84	2.56	2.71	2.75	0.99	2.75	3.19	2.25	3.80	2.66	−0.09
c	5.71	3.18	3.59	3.90	4.35	4.34	4.05	4.06	4.40	4.06	4.20	4.06	3.56	4.19	−1.52
d	7.69	7.85	7.78	7.82	7.78	7.79	7.78	7.76	7.77	7.76	7.83	7.86	7.87	7.78	0.09
e	7.46	7.64	7.60	7.52	7.55	7.67	7.69	7.68	7.69	7.68	7.62	7.68	7.58	7.68	0.22
f	8.30	8.43	8.44	8.22	8.14	8.50	8.51	8.50	8.51	8.50	8.29	8.33	8.35	8.49	0.19
g*	6.11	6.27	6.17	4.88	3.87	5.62	5.76	5.76	5.61	3.74	3.93	5.40	5.74	5.44	−0.67
g	5.22	4.67	4.73	5.03	4.50	3.31	3.66	3.76	3.29	5.76	4.92	4.76	4.98	3.75	−1.47
h	7.27	7.21	7.23	7.33	7.09	6.62	6.54	6.51	6.92	6.51	6.88	7.14	7.15	6.58	−0.69
i	7.40	7.64	7.64	7.53	7.56	7.41	7.36	7.35	7.50	7.35	7.41	7.55	7.60	7.38	−0.02
j	7.31	7.56	7.57	7.54	7.58	7.41	7.42	7.41	7.47	7.41	7.42	7.46	7.56	7.41	0.10
k	7.41	7.41	7.45	6.93	7.16	7.64	7.46	7.47	7.36	7.47	7.44	7.27	7.27	7.54	0.13
l	7.30	7.48	7.56	7.31	7.23	7.67	7.72	7.72	7.56	7.72	7.50	7.31	7.54	7.69	0.39
m	3.43	3.21	2.98	2.96	3.81	3.54	3.48	3.64	3.75	3.65	3.53	3.55	2.86	3.54	0.11
n*	1.325	1.56	1.36	1.41	1.89	1.77	0.46	1.59	1.88	1.58	1.66	1.61	1.21	1.32	0.00
n	0.845	1.06	1.15	1.03	1.50	1.37	1.72	0.52	1.70	0.52	0.33	0.65	0.89	1.29	0.44
o	2.19	2.08	2.06	2.03	2.64	2.55	2.43	2.48	2.84	2.48	2.40	2.30	2.14	2.50	0.31
p	2.36	2.29	2.31	2.23	2.23	2.47	2.45	2.45	2.39	2.45	2.26	2.12	2.28	2.45	0.09
Boltzmann distribution															
Rel. ΔG_{sol} (kJ mol^{-1})	–	18.56	19.68	30.38	12.74	0.00	0.95	4.00	15.82	3.88	6.67	14.93	29.58	–	–
Boltzmann factor	–	5.581e^{-4}	3.554e^{-4}	4.745e^{-6}	5.856e^{-3}	1.000	6.829e^{-1}	1.994e^{-1}	1.690e^{-3}	2.089e^{-1}	6.782e^{-2}	2.421e^{-3}	6.562e^{-6}	–	–
Probability P_i	–	2.572e^{-4}	1.638e^{-4}	2.187e^{-6}	2.699e^{-3}	4.608e^{-1}	3.147e^{-1}	9.189e^{-2}	7.789e^{-4}	9.628e^{-2}	3.126e^{-2}	1.115e^{-3}	3.024e^{-6}	1.000	–

For this reason, in addition to studying the role of geometry optimisation in gas and solvent phases, we also performed a detailed analysis of the contribution of differential conformational species to the calculated ^{1}H and ^{13}C NMR spectra. The geometries of 15 different conformers were first optimised using O3LYP/DGDZVP in chloroform (Fig. 7; Structures are presented in Supporting Information Table S12; Cartesian coordinates are given in Tables S13–S24). In spite of the fact that geometry optimisation of the 15 conformers used different initial input coordinates, on comparison of the root-mean-square displacement values (RMSD/Å), visual inspection, and based on solvation free energies (ΔG_{sol}/kJ mol^{-1}) between the optimised structures, it was found that conformers 3, 5 and 13 converged to the same minima as conformers 2, 4 and 12, respectively. Thus, conformers 3, 5 and 13 were excluded from subsequent analysis to avoid erroneous duplication of the data sets (equivalent to unequal weighting). Conformer 7 was found to have the lowest calculated free energy (ΔG_{sol}/kJ mol^{-1}; at 298 K in chloroform). After identifying 12 unique conformers and calculating their individual ^{1}H (Table 3) and ^{13}C NMR chemical shifts (Table S25), we then calculated the Boltzmann-weighted chemical shifts using all ^{1}H and ^{13}C resonances assuming thermal equilibrium at 298 K (Fig. 8A and B).

Boltzmann-weighting improved the accuracy of the calculated ^{1}H resonances b and c (previously identified as outliers by the ROUT method). However, it was surprising to find that the overall accuracy of the Boltzmann-weighted ^{1}H chemical shifts was inferior (MAE = 0.365 ppm; MSE = 0.321 ppm^2; χ^2 = 4.836 ppm; m = 0.983; C = −0.064 ppm; and R^2 = 0.961) when compared to accuracy of the chemical shifts derived from the individual conformers (particular conformer 1 described in the previous sections [*vide supra*]). Surprisingly, no correlation was found between the overall calculated error analysis (MSE/ppm^2) for ^{1}H resonances and the relative conformer energy free energy ΔG_{sol}/kJ mol^{-1} (Fig. S6). These data indicate that explicit solvation effects are likely to play a varied and important role in stabilising some conformers over others at thermal equilibrium.

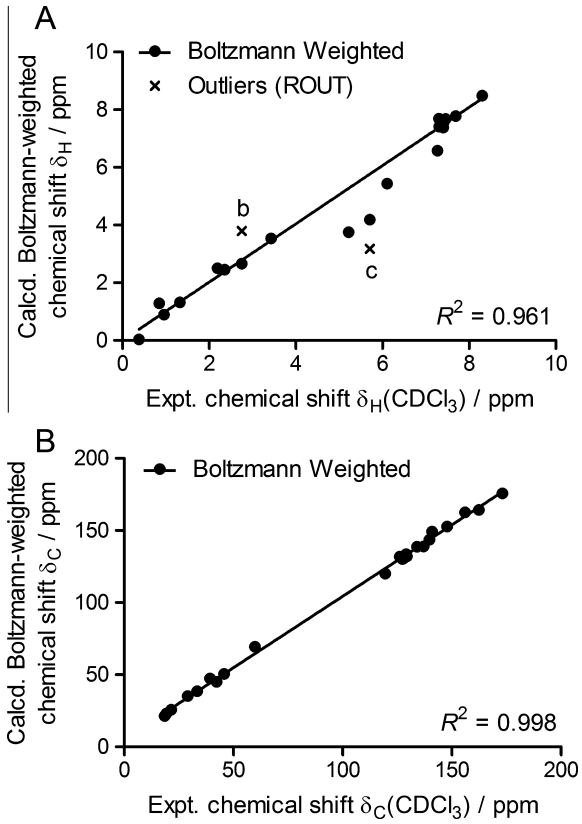


Fig. 8. Scatter plots and linear regression analyses showing the correlation between the Boltzmann-weighted calculated *versus* experimental chemical shifts for (A) the ^{1}H NMR and (B) the ^{13}C NMR spectra.

In contrast to the ^{1}H data, Boltzmann-weighted ^{13}C chemical shifts showed an improvement in accuracy (MAE = 4.367 ppm;

MSE = 23.385ppm^2; χ^2 = 9.027 ppm; m = 0.990; C = 5.351 ppm; and R^2 = 0.998) when compared to the calculations on individual isomers. In particular, calculated Boltzmann-weighted ^{13}C resonances for carbon atoms 3 and 19 displayed a dramatic improvement in accuracy, and were no longer considered outliers using the ROUT method. Further, we noted a positive correlation between the calculated MSE for ^{13}C resonances *versus* the calculated conformer energy (Fig. S7). In this system, lower energy conformers yield more accurate ^{13}C calculated chemical shifts. Reasons as to why the Boltzmann-weighted ^{1}H NMR chemical shifts display lower accuracy yet the ^{13}C data are improved are unknown. However, it seems likely that explicit solvation effects have a smaller influence on the (less accessible) carbon atoms than on solvent exposed protons.

4. Conclusion

Evaluation of the performance of different DFT methodologies for calculating the ^{1}H and ^{13}C NMR chemical shifts in a complex drug molecule revealed that the calculations provide highly accurate qualitative and quantitative data. Calculation of most aliphatic, as well as homo- and heteroaromatic groups typically have MAE values <0.2 ppm for ^{1}H resonances and <~6.0 ppm for ^{13}C resonances. Atoms spatially close to, or strongly influenced by, stereocentres show larger deviations. However, these deviations are most likely due to specific solvent–solute interactions that are not incorporated using continuum solvation models in DFT. Our investigation demonstrated that small basis sets like DGDZVP and 6-31++G(d,p) are sufficient, and indeed typically superior to triple-ζ basis sets, for calculating ^{1}H and ^{13}C chemical shifts in molecules up to ~500 Da. O3LYP was found to be the most accurate DFT methodology tested. Furthermore, we conclude that full geometry optimisation using the XC method of choice in either gas phase or solution phase is crucial to attaining the most accurate calculated NMR spectra. In addition, we found that the accuracy of all calculated ^{13}C chemical shifts improved by inclusion of contributions from different conformers using a Boltzmann weighted analysis. Given the high degree of quantitative and qualitative accuracy achieved by DFT in calculating ^{1}H and ^{13}C spectra, it is anticipated that NMR calculations are set to play an increasingly important role in drug discovery and lead optimisation.

Acknowledgements

We thank Dr. Thomas Betzel for technical assistance. We thank Dr. Eliane Fischer and Dr. T. Lee Collier for support with statistical analyses. We are grateful to the staff at the Massachusetts Green High Performance Computing Center (MGHPCC, Holyoke, MA, USA) and the Northeastern University Information Technology Service and Research Computing (Boston, USA; http://www.northeastern.edu/rc) for providing technical support and access to the Discovery Cluster. We are grateful to Dr. Rein U. Kirss (Northeastern University) for supporting access to the MGHPCC (DEH). We thank Dr. Ivan Greguric and the Australian National Science and Technology Organisation (ANSTO, Sydney) for providing financial support (JPH).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.comptc.2014.11.007.

References

[1] M. Kaupp, M. Bühl, V.G. Malkin, Calculation of NMR and EPR parameters. Theory and applications, Wiley-VCH Verlag GmbH & Co, 2004.
[2] M.A. Watson, P. Sałek, P. Macak, M. Jaszuński, T. Helgaker, The calculation of indirect nuclear spin-spin coupling constants in large molecules, chemistry, Eur. J. 10 (2004) 4627–4639.
[3] W. Koch, M.C. Holthausen, A Chemist's Guide to Density Functional Theory, second ed., 2001.
[4] J.C. Facelli, Chemical shift tensors: theory and application to molecular structural problems, Prog. Nucl. Magn. Reson. Spectrosc. 58 (2011) 176–201.
[5] M. Bühl, T. van Mourik, NMR spectroscopy: quantum-chemical calculations, Wiley Interdiscipl. Rev.: Comput. Mol. Sci. 1 (2011) 634–647.
[6] G. Schreckenbach, T. Ziegler, Density functional calculations of NMR chemical shifts and ESR g-tensors, Theor. Chem. Acc. 99 (1998) 71–82.
[7] T. Helgaker, M. Jaszuński, K. Ruud, Ab initio methods for the calculation of NMR shielding and indirect spin–spin coupling constants, Chem. Rev. 99 (1998) 293–352.
[8] H. Fukui, Theory and calculation of nuclear spin–spin coupling constants, Prog. Nucl. Magn. Reson. Spectrosc. 35 (1999) 267–294.
[9] J.R. Cheeseman, G.W. Trucks, T.A. Keith, M.J. Frisch, A comparison of models for calculating nuclear magnetic resonance shielding tensors, J. Chem. Phys. 104 (1996) 5497–5509.
[10] J.C. Facelli, Chemical shift tensors: theory and application to molecular structural problems, Prog. Nucl. Magn. Reson. Spectrosc. 58 (2011) 176–201.
[11] D. Santos, R. Suardiaz, L.A. Montero, C. Pérez, DFT analysis of rotational barriers, ^{1}H and ^{13}C NMR chemical shifts in neutral and protonated furfurylidenanilines, J. Mol. Struct. (Thoechem) 852 (2008) 78–82.
[12] K.J. Ooms, S.E. Bolte, J.J. Smee, B. Baruah, D.C. Crans, T. Polenova, Investigating the vanadium environments in hydroxylamido V(V) dipicolinate complexes using 51V NMR spectroscopy and density functional theory, Inorg. Chem. 46 (2007) 9285–9293.
[13] S.E. Bolte, K.J. Ooms, T. Polenova, B. Baruah, D.C. Crans, J.J. Smee, V51 solid-state NMR and density functional theory studies of vanadium environments in V(V)O2 dipicolinic acid complexes, J. Chem. Phys. 128 (2008).
[14] H. Cybulski, M. Pecul, T. Helgaker, M. Jaszuński, Theoretical studies of nuclear magnetic resonance parameters for the proton-exchange pathways in porphyrin and porphycene, J. Phys. Chem. A 109 (2005) 4162–4171.
[15] M. Pietrzak, C. Benedict, H. Gehring, E. Daltrozzo, H.-H. Limbach, NMR studies and DFT calculations of the symmetric intramolecular NHN-hydrogen bond of bis-(2-pyridyl)-acetonitrile: isotope labeling strategy for the indirect 13C-detection of 15N15N couplings, J. Mol. Struct. 844–845 (2007) 222–231.
[16] I. del Rosal, L. Maron, R. Poteau, F. Jolibois, DFT calculations of ^{1}H and ^{13}C NMR chemical shifts in transition metal hydrides, Dalton Trans. (2008) 3959–3970.
[17] O.B. Lutnaes, T.A. Ruden, T. Helgaker, The performance of hybrid density functional theory for the calculation of indirect nuclear spin–spin coupling constants in substituted hydrocarbons, Magn. Reson. Chem.: MRC (2004) S117–S127. 42 Spec no.
[18] J.E. Peralta, V. Barone, G.E. Scuseria, R.H. Contreras, Density functional theory calculation of indirect nuclear magnetic resonance spin–spin coupling constants in C(70), J. Am. Chem. Soc. 126 (2004) 7428–7429.
[19] L.M. Khalilov, A.R. Tulyabaev, V.M. Yanybin, A.R. Tuktarov, ^{1}H and ^{13}C NMR chemical shift assignments of spiro-cycloalkylidenehomo- and methanofullerenes by the DFT–GIAO method, Magn. Reson. Chem. 49 (2011) 378–384.
[20] M.D. Esrafili, H. Ahmadin, DFT study of 17O, ^{1}H and ^{13}C NMR chemical shifts in two forms of native cellulose, Iα and Iβ, Carbohydr. Res. 347 (2012) 99–106.
[21] M.A. Muñoz, P. Joseph-Nathan, DFT-GIAO ^{1}H and ^{13}C NMR prediction of chemical shifts for the configurational assignment of 6β-hydroxyhyoscyamine diastereoisomers, Magn. Reson. Chem. 47 (2009) 578–584.
[22] D.J. Marell, S.J. Emond, A. Kulshrestha, T.R. Hoye, Analysis of seven-membered lactones by computational NMR methods: proton NMR chemical shift data are more discriminating than carbon, J. Org. Chem. 79 (2013) 752–758.
[23] K.W. Wiitala, C.J. Cramer, T.R. Hoye, Comparison of various density functional methods for distinguishing stereoisomers based on computed ^{1}H or ^{13}C NMR chemical shifts using diastereomeric penam β-lactams as a test set, Magn. Reson. Chem. 45 (2007) 819–829.
[24] A. Bagno, F. Rastrelli, G. Saielli, Toward the complete prediction of the ^{1}H and ^{13}C NMR spectra of complex organic molecules by DFT methods: application to natural substances, chemistry, Eur. J. 12 (2006) 5514–5525.
[25] P. Bouř, Calculation of NMR chemical shifts for taxol and α-pinene within the generalized gradient approximation, Int. J. Quantum Chem. 91 (2003) 277–283.
[26] W. Kutzelnigg, U. Fleischer, M. Schindler, The IGLO-Method: Ab-initio Calculation and Interpretation of NMR Chemical Shifts and Magnetic Susceptibilities, Deuterium and Shift Calculation, Springer, Berlin, Heidelberg, 1991. pp. 165–262.
[27] V.G. Malkin, O.L. Malkina, D.R. Salahub, Calculations of NMR shielding constants beyond uncoupled density functional theory. IGLO approach, Chem. Phys. Lett. 204 (1993) 87–95.
[28] V.G. Malkin, O.L. Malkina, D.R. Salahub, Calculation of spin—spin coupling constants using density functional theory, Chem. Phys. Lett. 221 (1994) 91–99.
[29] O.L. Malkina, D.R. Salahub, V.G. Malkin, Nuclear magnetic resonance spin–spin coupling constants from density functional theory: problems and results, J. Chem. Phys. 105 (1996) 8793–8800.

[30] C. Adamo, V. Barone, Toward chemical accuracy in the computation of NMR shieldings: the PBE0 model, Chem. Phys. Lett. 298 (1998) 113–119.
[31] C. Adamo, M. Cossi, V. Barone, An accurate density functional method for the study of magnetic properties: the PBE0 model, THEOCHEM 493 (1999) 145–157.
[32] P.R. Rablen, S.A. Pearlman, J. Finkbiner, A comparison of density functional methods for the estimation of proton chemical shifts with chemical accuracy, J. Phys. Chem. A 103 (1999) 7357–7363.
[33] R. Laskowski, P. Blaha, F. Tran, Assessment of DFT functionals with NMR chemical shifts, Phys. Rev. B 87 (2013) 195130.
[34] G.K. Pierens, ^{1}H and ^{13}C NMR scaling factors for the calculation of chemical shifts in commonly used solvents using density functional theory, J. Comput. Chem. 35 (2014) 1388–1394.
[35] R. Jain, T. Bally, P.R. Rablen, Calculating accurate proton chemical shifts of organic molecules with density functional methods and modest basis sets, J. Org. Chem. 74 (2009) 4017–4023.
[36] F. London, Théorie quantique des courants interatomiques dans les combinaisons aromatiques, J. Phys. Radium 8 (1937) 397–409.
[37] K. Wolinski, J.F. Hinton, P. Pulay, Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations, J. Am. Chem. Soc. 112 (1990) 8251–8260.
[38] G. Schreckenbach, T. Ziegler, Calculation of NMR shielding tensors using gauge-including atomic orbitals and modern density functional theory, J. Phys. Chem. 99 (1995) 606–611.
[39] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.J.A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09 Revision B.01, Gaussian 09 Revision B.01., Gaussian Inc, Wallingford, CT, 2010.
[40] M. Bühl, M. Kaupp, O.L. Malkina, V.G. Malkin, The DFT route to NMR chemical shifts, J. Comput. Chem. 20 (1999) 91–105.
[41] T. Kupka, M. Stachów, J. Kaminsky, S.P.A. Sauer, Estimation of isotropic nuclear magnetic shieldings in the CCSD(T) and MP2 complete basis set limit using affordable correlation calculations, Magn. Reson. Chem. 51 (2013) 482–489.
[42] K. Aidas, A. Møgelhøj, C.B. Nielsen, K.V. Mikkelsen, K. Ruud, O. Christiansen, J. Kongsted, Solvent effects on NMR isotropic shielding constants. A comparison between explicit polarizable discrete and continuum approaches, J. Phys. Chem. A 111 (2007) 4199–4210.
[43] F. Jensen, Basis set convergence of nuclear magnetic shielding constants calculated by density functional methods, J. Chem. Theory Comput. 4 (2008) 719–727.
[44] M.W. Lodewyk, M.R. Siebert, D.J. Tantillo, Computational prediction of ^{1}H and ^{13}C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry, Chem. Rev. 112 (2011) 1839–1862.
[45] F.A.A. Mulder, M. Filatov, NMR chemical shift data and ab initio shielding calculations: emerging tools for protein structure determination, Chem. Soc. Rev. 39 (2010) 578–590.
[46] P.H. Willoughby, M.J. Jansma, T.R. Hoye, A guide to small-molecule structure assignment through computation of (^{1}H and ^{13}C) NMR chemical shifts, Nat. Protocols 9 (2014) 643–660.
[47] J.P. Holland, M.W. Jones, S. Cohrs, R. Schibli, E. Fischer, Fluorinated quinazolinones as potential radiotracers for imaging kinesin spindle protein expression, Bioorg. Med. Chem. 21 (2013) 496–507.
[48] T.A. Yap, L.R. Molife, S.P. Blagden, S. de Bono, Targeting cell cycle kinases and kinesins in anticancer drug development, Expert Opin. Drug Discovery 2 (2007) 539–560.
[49] C.A. Parrish, N.D. Adams, K.R. Auger, J.L. Burgess, J.D. Carson, A.M. Chaudhari, R.A. Copeland, M.A. Diamond, C.A. Donatelli, K.J. Duffy, L.F. Faucette, J.T. Finer, W.F. Huffman, E.D. Hugger, J.R. Jackson, S.D. Knight, L. Luo, M.L. Moore, K.A. Newlander, L.H. Ridgers, R. Sakowicz, A.N. Shaw, C.-M.M. Sung, D. Sutton, K.W. Wood, S.-Y. Zhang, M.N. Zimmerman, D. Dhanak, Novel ATP-competitive kinesin spindle protein inhibitors, J. Med. Chem. 50 (2007) 4939–4952.
[50] J.W. Purcell, J. Davis, M. Reddy, S. Martin, K. Samayoa, H. Vo, K. Thomsen, P. Bean, W.L. Kuo, S. Ziyad, J. Billig, H.S. Feiler, J.W. Gray, K.W. Wood, S. Cases, Activity of the kinesin spindle protein inhibitor Ispinesib (SB-715992) in models of breast cancer, Clin. Cancer Res. 16 (2010) 566–576.
[51] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B: Condens. Matter Mater. Phys. 37 (1988) 785–789.
[52] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A: At., Mol., Opt. Phys. 38 (1988) 3098–3100.
[53] J.P. Perdew, Density functional approximation for the correlation energy of the inhomogeneous electron gas, Phys. Rev. B: Condens. Matter Mater. Phys. 33 (1986) 8800.
[54] J.P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas: erratum, Phys. Rev. B: Condens. Matter 33 (1986) 8822.
[55] J.P. Perdew, Electronic structure of solids '91, Electronic structure of solids '91, 1991, 11.
[56] K. Burke, J.P. Perdew, Y. Wang, Electron density functional theory, Electron density functional theory, 1998, 81.
[57] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868.
[58] C. Adamo, V. Barone, Toward reliable adiabatic connection models free from adjustable parameters, Chem. Phys. Lett. 274 (1997) 242–250.
[59] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields, J. Phys. Chem. 98 (1994) 11623–11627.
[60] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.
[61] A.D. Becke, Density-functional thermochemistry. V. Systematic optimization of exchange–correlation functionals, J. Chem. Phys. 107 (1997) 8554–8560.
[62] H.L. Schmider, A.D. Becke, Optimized density functionals from the extended G2 test set, J. Chem. Phys. 108 (1998) 9624–9631.
[63] N.C. Handy, A.J. Cohen, Left-right correlation energy, Mol. Phys. 99 (2001) 403–412.
[64] W.M. Hoe, A.J. Cohen, N.C. Handy, Assessment of a new local exchange functional OPTX, Chem. Phys. Lett. 341 (2001) 319–328.
[65] X. Xu, A. Goddard William 3rd, From the cover: The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties, Proc. Natl. Acad. Sci. USA 101 (2004) 2673–2677.
[66] J.P. Perdew, K. Burke, Y. Wang, Generalized gradient approximation for the exchange–correlation hole of a many-electron system, Phys. Rev. B: Condens. Matter 54 (1996) 16533–16539.
[67] A.D. Boese, N.C. Handy, New exchange–correlation density functionals: the role of the kinetic-energy density, J. Chem. Phys. 116 (2002) 9559–9569.
[68] J. Tao, J.P. Perdew, V.N. Staroverov, G.E. Scuseria, Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids, Phys. Rev. Lett. 91 (2003). 146401/146401-146401/146404.
[69] T. Van Voorhis, G.E. Scuseria, A novel form for the exchange–correlation energy functional, J. Chem. Phys. 109 (1998) 400–410.
[70] J.B. Krieger, J. Chen, G.J. Iafrate, Electron correlations and materials properties, Electron correlations and materials properties, 1999, 463.
[71] Y. Zhao, D.G. Truhlar, How well can new-generation density functional methods describe stacking interactions in biological systems?, Phys Chem. Chem. Phys. 7 (2005) 2701–2705.
[72] M.M. Quintal, A. Karton, M.A. Iron, A.D. Boese, J.M.L. Martin, Benchmark study of DFT functionals for late-transition-metal reactions, J. Phys. Chem. A 110 (2006) 709–716.
[73] Y. Zhao, J. Pu, B.J. Lynch, D.G. Truhlar, Tests of second-generation and third-generation density functionals for thermochemical kinetics, Phys. Chem. Chem. Phys. 6 (2004) 673–676.
[74] J.P. Holland, J.C. Green, Evaluation of exchange–correlation functionals for time-dependent density functional theory calculations on metal complexes, J. Comput. Chem. 31 (2010) 1008–1014.
[75] N. Godbout, D.R. Salahub, J. Andzelm, E. Wimmer, Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation, Can. J. Chem. 70 (1992) 560–571.
[76] C. Sosa, J. Andzelm, B.C. Elkin, E. Wimmer, K.D. Dobbs, D.A. Dixon, A local density functional study of the structure and vibrational frequencies of molecular transition-metal compounds, J. Phys. Chem. 96 (1992) 6630–6636.
[77] V.A. Rassolov, J.A. Pople, M.A. Ratner, T.L. Windus, 6-31G* basis set for atoms K through Zn, J. Chem. Phys. 109 (1998) 1223–1229.
[78] V.A. Rassolov, M.A. Ratner, J.A. Pople, P.C. Redfern, L.A. Curtiss, 6-31G* basis set for third-row atoms, J. Comput. Chem. 22 (2001) 976–984.
[79] R.J. Hall, M.M. Davidson, N.A. Burton, I.H. Hillier, Combined density functional, self-consistent reaction field model of solvation, J. Phys. Chem. 99 (1995) 921–924.
[80] J.L. Chen, L. Noodleman, D.A. Case, D. Bashford, Incorporating solvation effects into density functional electronic structure calculations, J. Phys. Chem. 98 (1994) 11059–11068.
[81] J. Tomasi, B. Mennucci, R. Cammi, Quantum mechanical continuum solvation models, Chem. Rev. (Washington, DC, U.S.) 105 (8) (2005) 2999–3094.
[82] M.L. Connolly, Solvent-accessible surfaces of proteins and nucleic acids, Science (Washington, DC, U.S.) 221 (1983) 709–713.
[83] K.W. Wiitala, T.R. Hoye, C.J. Cramer, Hybrid density functional methods empirically optimized for the computation of ^{13}C and ^{1}H chemical shifts in chloroform solution, J. Chem. Theory Comput. 2 (2006) 1085–1092.
[84] H. Motulsky, R. Brown, Detecting outliers when fitting data with nonlinear regression – a new method based on robust nonlinear regression and the false discovery rate, BMC Bioinformatics 7 (2006) 123.
[85] B. Zhang, J.-F. Liu, Y. Xu, S.-C. Ng, Crystal structure of HsEg5 in complex with clinical candidate CK0238273 provides insight into inhibitory mechanism, potency, and specificity, Biochem. Biophys. Res. Commun. 372 (2008) 565–570.