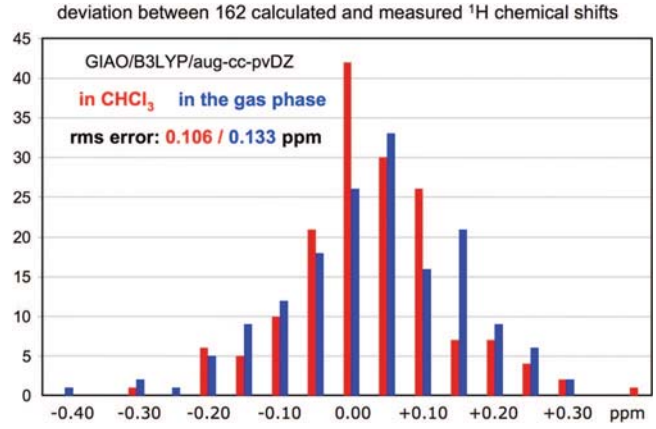
Published in "The Journal of Organic Chemistry 74(11): 4017–4023, 2009"
which should be cited to refer to this work.

Calculating Accurate Proton Chemical Shifts of Organic Molecules with Density Functional Methods and Modest Basis Sets

Rupal Jain,[†,‡] Thomas Bally,[†,*] and Paul R. Rablen[§,*]

Department of Chemistry, University of Fribourg, CH-1700 Fribourg, Switzerland, and Department of Chemistry and Biochemistry, Swarthmore College, Swarthmore, Pennsylvania 19081-1397

thomas.bally@unifr.ch; prablen1@swarthmore.edu



The purpose of this paper is to convince practitioners of ^{1}H NMR spectroscopy to consider simple quantum chemical calculations as a viable option to aid them in the assignment of their spectra. To this end, it is demonstrated, on a test set of 80 conformationally stable molecules of various kinds carrying different functional groups, that, in contrast to what is claimed in the literature, large basis sets are not needed to obtain rather accurate predictions of ^{1}H NMR chemical shifts by quantum chemical calculations. On the other hand, modeling the solvent by an SCRF-type calculation may improve certain predictions significantly. The best accuracy/cost ratio is provided by GIAO calculations in chloroform as a solvent with the specially parametrized WP04 functional of Cramer et al. using the cc-pVDZ or 6-31G** basis set, closely followed by similar calculations with the ubiquitious B3LYP functional (both predict ^{1}H chemical shifts with an average deviation of ca. 0.12 ppm, if the results are scaled linearly). A slightly higher accuracy can be attained by adding diffuse functions to the basis set, but going to the triple-ζ basis sets which have invariably been used hitherto in calculations of chemical shifts does not lead to any improvement. The popular increment schemes such as those implemented in the ChemDraw or ACD programs do not do nearly as well and are often incapable of correctly distinguishing stereoisomers.

Introduction

NMR spectroscopy is arguably one of the most powerful analytic tools available to chemists for assigning the identities and, to some extent, the structures of molecules.[1] Therefore, much effort has gone into aiding this assignment task by developing procedures to predict NMR chemical shifts and coupling constants.[2] As it turns out, increment schemes which account for the effects of neighboring and more remote groups do quite well in predicting chemical shifts (and, where available, also coupling constants, i.e., entire NMR spectra), so such

[†] University of Fribourg.
[‡] Current address: IIT Bombay, India.
[§] Swarthmore College.

schemes, e.g., that which has been implemented in the widespread ChemDraw program, have become quite popular among organic chemists.

On the other hand, very few practitioners of NMR spectroscopy ever consider the option to use quantum chemical methods to predict NMR spectra, although such methods are nowadays well developed[3−6] and implemented in many popular quantum chemical program packages. One of the reasons is probably that the theoreticians who study the calculation of electronic shielding tensors by quantum chemical methods almost invariably use very large basis sets (triple-ζ or higher), probably in an effort to get as close as possible to the basis set limit of whatever method they employ for these calculations.[7,8] As such basis sets cannot be routinely applied to the large molecules that often stand at the focus of the work of practicing organic chemists, these usually do not consider quantum chemical calculations of NMR chemical shifts as something that is feasible in practice.

In addition, in spite of the fact that most of the millions of NMR spectra that are recorded every day by organic chemists are targeted at ^{1}H nuclei, most of the effort on the part of quantum chemists has gone into calculating chemical shifts of heavier nuclei. Apart from occasional studies of specific compound groups,[9,10] we note the 1997 studies of Chesnut[11] and Sauer et al.[12] who demonstrated that ^{1}H chemical shifts of small molecules can be obtained to within ca. ± 0.1 ppm, if very large basis sets are employed, and the comprehensive 1999 study of Rablen et al.[13] who computed ^{1}H chemical shifts of 80 compounds by different hybrid density functional methods, coupled with triple-ζ basis sets augmented by diffuse functions and by several sets of polarizaton functions. These authors concluded that the best cost/performance ratio was achieved by the GIAO/B3LYP/6-311++G** model which provided predictions with an rms deviation of 0.15 ppm on their test set of 80 molecules.

However, even that basis set, which would probably be considered as very modest by the experts in the field, features 22 contracted functions per atom of the second period and 7 per H-atom, compared to 14 and 5 basis functions, respectively, in typical double-ζ basis sets. As quantum chemical calculations typically scale with the third to the fifth power of the number of basis functions, these differences quickly translate into orders of magnitude increases in calculation time for larger molecules, and this may decide whether a calculation is feasible or not.

More recently, the Pulay group has undertaken some studies of proton chemical shifts in some substituted aromatic hydrocarbons[14] and cyclic peptide analogues[15] which actually showed quite an acceptable accuracy for the simple GIAO/B3LYP/6-31G** method, but the authors nevertheless eventually recommended the use of triple-ζ basis sets. At the other end of the theoretical spectrum, Patchkovskii and Thiel reparametrized the semiempirical MNDO method to optimally reproduce NMR chemical shifts,[16] but in the case of ^{1}H, agreement with experiment was inferior to what can be achieved with good increment methods (which are still faster than even a semiempirical quantum chemical calculation). Recently, the Merz group has improved the efficiency of that approach to the extent that they could apply it to rather large proteins,[17] but the performance of the method for ^{1}H chemical shifts in unsaturated compounds is still unsatisfactory, and it depends very much on the method used for geometry optimization (cf. Table II in ref 17).

We had occasionally done calculations of ^{1}H chemical shifts for our own purposes and we had obtained the impression that double-ζ basis sets were really all that is needed to get predictions that are of sufficient accuracy, e.g., to allow distinguishing stereoisomers (which is something that the increment schemes mentioned at the outset are not very good at). Thus we decided to address this issue in a coherent manner, and this paper reports on a systematic study of the effects that different basis sets, methods of calculating chemical shifts, inclusion of solvent effect, and different density functionals have on the accuracy of predicting ^{1}H chemical shifts.

Methods

Theory.[18] Chemical shifts are properties that depend on the interaction of static magnetic fields (the strong external field and the small internal fields of the nuclei) with the magnetic field created by the electron's *movement* inside a molecule. Hence, these static fields perturb the *kinetic* energy term of the Hamiltonian, and it turns out that, if finite basis sets are used to model the electron distribution, the results of the calculation with such a perturbed operator depend on the *origin* of the coordinate system (the "gauge"). Different schemes have been proposed to circumvent this problem, two of which are implemented in several quantum chemical programs: the GIAO (gauge invariant atomic orbitals) method which uses basis functions that have an explicit field dependence[19,20] and the CSGT (continuous set of gauge transformation) method which computes the current density induced by the magnetic fields by performing a gauge transformation at every point in space.[21]

In 1996, Cheeseman et al. performed a comparison of the two methods and concluded that, for the case of ^{13}C, ^{15}N, and ^{17}O chemical shifts, they lead to the same results for very large basis sets, but that the GIAO results converge more rapidly toward that limit than the CSGT results do.[22] This observation was confirmed for ^{1}H chemical shifts by Rablen et al.[13] but,

(1) Claridge, T. D. W. *High-Resolution NMR Techniques in Organic Chemistry*, 2nd ed.; Elsevier: New York, 2008; Vol. 27.
(2) Abraham, R. J.; Mobili, M. *Modelling 1H NMR Spectra of Organic Compounds: Theory and Applications*; John Wiley & Sons: Chichester, 2008.
(3) Facelli, J. C. *Concepts Magn. Reson., Part A* **2004**, *20A*, 42.
(4) Wilson, P. J. *Annu. Rep. NMR Spectrosc.* **2003**, *49*, 117.
(5) Jameson, C. J.; de Dios, A. C. *Nucl. Magn. Res.* **2001**, *30*, 46.
(6) For an exhaustive list of recent reviews on the topic, see: Casblanca, L. E.; de Dios, A. C. *J. Chem. Phys.* **2008**, *128*, 052202.
(7) Kupka, T.; Ruscic, B.; Botto, R. E. *J. Phys. Chem. A* **2002**, *106*, 10396.
(8) Magyarfalvi, G.; Pulay, P. *J. Chem. Phys.* **2003**, *119*, 1350.
(9) Perez, M.; Peakman, T. M.; Alex, A.; Higginson, P. D.; Mitchell, J. C.; Snowden, M. J.; Morao, I. *J. Org. Chem.* **2006**, *71*, 3103.
(10) d'Antuono, P.; Botek, E.; Champagne, B.; Wieme, J.; Reyniers, M.-F.; Marin, G. B.; Adriaensens, P. J.; Gelan, J. M. *Chem. Phys. Lett.* **2007**, *436*, 388.
(11) Chesnut, D. B. *Chem. Phys.* **1997**, *214*, 73.
(12) Sauer, S. P. A.; Spirko, V.; Paidarova, I.; Kraemer, W. P. *Chem. Phys.* **1997**, 214.
(13) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357.
(14) Wang, B.; Fleischer, U.; Hinton, J. F.; Pulay, P. *J. Comput. Chem.* **2001**, *22*, 1887.
(15) Wang, B.; Hinton, J. F.; Pulay, P. *J. Comput. Chem.* **2002**, *23*, 492.
(16) Patchkovskii, S.; Thiel, W. *J. Comput. Chem.* **1999**, *20*, 1220.
(17) Wang, B.; Brothers, E. N.; van der Vaart, A.; Merz, K. N. *J. Chem. Phys.* **2004**, *120*, 11392.
(18) See, e.g.: Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, 2004.
(19) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789.
(20) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251.
(21) Keith, T. A.; Bader, R. W. F. *Chem. Phys. Lett.* **1993**, *210*, 223 and earlier references cited therein.
(22) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497.

again, only for triple-ζ basis sets. Unfortunately, the GIAO method has not been coded yet for the newer generation of meta-GGA density functionals (which model the *kinetic energy* of the fully interacting system of electrons explicitly), so we had to include both the GIAO and the CSGT method (which does work with meta-GGA functionals) in our investigations.[23]

Functionals. Exploratory calculations showed clearly that hybrid functionals are more apt to yield accurate predictions of ^{1}H chemical shifts than pure functionals do, but also confirmed the earlier result of Rablen et al. that it makes little difference which hybrid GGA functional one uses. We decided to focus on two of them, the popular B3LYP functional[24,25] and the recently proposed WP04 functional, a version of the B3LYP functional that was reparametrized explicitly for calculating chemical shifts in chloroform.[26] Once we had determined what is the best basis set for such calculations, we repeated them with that basis set (*aug*-cc-pVDZ) using the VSXC,[27] TPSS,[28] and BMK[29] meta-GGA functionals, coupled with the CSGT method to evaluate chemical shieldings. All calculations were carried out with the Gaussian program package.[30]

Basis Sets. We systematically explored the double- and triple-ζ basis sets of the Pople (6-31[++]G(d,p),[31] 6-311[++]-G(d,p)[32]) and Dunning families ([aug]-cc-pVXZ, X = D,T),[33] where the characters in brackets denote diffuse functions that were optionally added in either case. The two families of basis sets sometimes gave surprisingly different results, especially with the CSGT method to calculate chemical shieldings (see the Results and Discussion).

Reference Value. In principle, the evaluation of chemical shifts requires separate calculations of the isotropic chemical shielding for all protons in the compound of interest and that of TMS which is usually used as a reference (the chemical shift is the difference of these two numbers). However, instead of plotting calculated against measured chemical shifts, one can also plot the experimental values against the calculated isotropic shieldings, and use the intercept of the resulting regression line instead of the calculated chemical shift of TMS as a constant that is subtracted from the calculated shieldings to convert them into chemical shifts (ideally the two values should be identical).[13] This approach has the advantages that it allows compensation for systematic errors of the method, and does not depend critically on the accurate calculation of one particular species (TMS), and we therefore use it in the present work.

Solvent Modeling. We found that, in some cases, the chemical shift is quite sensitive to the environment of a molecule, and that these cases are often the ones that constitute outliers in a regression involving chemical shieldings calculated in the gas phase. Thus, we systematically modeled solvent effects by representing chloroform (the solvent mostly used in practice) as a polarizable continuum, according to the method implemented in the PCM-SCRF (self-consistent reaction field) procedure[34] in the Gaussian program.[30]

Increment Methods. Finally, we wanted to compare the performance of the different quantum chemical methods we have tested to that of two popular increment schemes, i.e., the one that is implemented in the "Ultra" versions of the ChemDraw program[35] and the one implemented in the ^{1}H-predictor from ACD, Inc.[36] There we did not perform a linear regression, because these procedures were designed from the outset to yield correct predictions without scaling.

The Test Set

Exploring the performance of different methods requires a set of experimental data against which comparison can be made. For this purpose, the data set assembled earlier by Rablen et al. and documented extensively in reference[13] was selected. This test set (the structures of which are shown in the Supporting Information and on the cover of this issue) had been assembled according to the following criteria:

(a) If a molecule has more than one significantly populated conformer, then a prediction of chemical shifts requires a separate calculation for each conformer, followed by Boltzmann weighting, since magnetic shielding values frequently show a strong conformational dependence.[37] The Boltzmann weighting requires relative free energies of different conformations with an accuracy and a reliability that is difficult to achieve with affordable methods. In order to avoid these complications, and the concomitant sources of error, only compounds with a single, relatively well-defined conformation were included in the test set.

(b) The chemical shifts of protons bonded to heteroatoms are often strongly affected by the solvent and/or the presence of water with which proton exchange can occur. In addition, such protons engage in hydrogen bonding to lone pairs which can lead to aggregation, in which case chemical shifts become concentration-dependent. Therefore, only compounds containing no O−H or N−H bonds were considered (with the sole exception of methanol and indole).

(c) In order to achieve a reasonable degree of reliability and consistency in the way ^{1}H NMR spectra were measured, only a limited number of sources were used. With one exception, the spectra were all taken from three databases, as documented in reference.[13] The vast majority of the data are from spectra taken in either chloroform or carbon tetrachloride as solvent.

The geometries of all the molecules in the test set were optimized by the B3LYP/6-31G* method which has amply been proven to give very good ground-state geometries. All calculations were carried out with the Gaussian program.[30]

Results and Discussion

Table 1 sums up the results obtained with all combinations of the two methods to calculate magnetic shieldings (GIAO/

(23) In Gaussian, these calculations are simply invoked by the NMR=GIAO or NMR=CSGT keyword, respectively. The isotropic chemical shieldings in ppm appear in the output for each atom in the section labelled "Magnetic shielding (ppm)". They are followed by the complete shielding tensors and their eigenvalues which can, however, be ignored for the present purposes.

(24) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(25) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(26) Wiitala, K. W.; Hoye, T. R.; Cramer, C. J. *J. Chem. Theory. Comput.* **2006**, *2*, 1085. In Gaussian, the WP04 functional is invoked by specifying the B3LYP keyword and adding iop (3/76 = 1000001189, 3/77 = 0961409999, 3/78 = 0000109999) to the keyword line.
(27) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.
(28) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, 91.
(29) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.
(30) Frisch, M. J., et al. *Gaussian 03, Revision E.01*; Gaussian, Inc.: Wallingford, CT, 2003. (For the complete reference, see the Supporting Information).
(31) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(32) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.
(33) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
(34) Cossi, M.; Scalmani, G.; Rega, N.; Barone, J. *J. Chem. Phys.* **2002**, *117*, 43.
(35) ChemDraw Ultra, Versions 8 and later, CambridgeSoft Co., Cambridge, MA.
(36) ACD 1HNMR Predictor, Version 12.0, Advanced Chemistry Development, Ontario, Canada, 2009.
(37) Wiitala, K. W.; Al-Rashid, Z. F.; Dvornikovs, V.; Hoye, T. R.; Cramer, C. J. *J. Phys. Org. Chem.* **2007**, *20*, 345.

TABLE 1. Results of Calculations of ^{1}H NMR Chemical Shifts (Recommended Methods in Italic)

				rms error		linear regression[e]		
method[a]	functional[b]	basis set[c]	SCRF[d]	(unscaled)	(scaled)	slope	intercept	CPU time[f]
GIAO	*WP04*	*aug-cc-pVDZ*	*yes*	*0.281*	*0.103*	*1.0544*	*31.905*	*21.2*
GIAO	B3LYP	aug-cc-pVDZ	yes	0.419	0.106	1.0812	31.676	21.9
GIAO	WP04	aug-cc-pVTZ	yes	0.239	0.112	1.0505	31.867	363.2
GIAO	*WP04*	*aug-cc-pVDZ*	*no*	*0.150*	*0.112*	*1.0277*	*31.951*	*20.8*
GIAO	WP04	cc-pVTZ	yes	0.189	0.113	1.0440	31.887	54.1
GIAO	WP04	cc-pVDZ	yes	0.145	0.115	1.0205	31.844	3.6
GIAO	B3LYP	aug-cc-pVTZ	yes	0.374	0.117	1.0794	31.626	377.4
GIAO	B3LYP	cc-pVTZ	yes	0.303	0.118	1.0716	31.666	54.5
GIAO	WP04	6-31G(d,p)	yes	0.149	0.119	1.0332	32.018	2.5
GIAO	*WP04*	*6-31G(d)*	*yes*	*0.480*	*0.120*	*0.9927*	*32.433*	*1.8*
CSGT	WP04	aug-cc-pVTZ	yes	0.235	0.120	1.0480	31.824	227.9
GIAO	WP04	6-31++G(d,p)	yes	0.196	0.121	1.0424	31.934	5.0
GIAO	WP04	aug-cc-pVTZ	no	0.140	0.122	1.0241	31.912	360.4
CSGT	B3LYP	aug-cc-pVTZ	yes	0.377	0.123	1.0781	31.588	240.8
GIAO	WP04	cc-pVTZ	no	0.143	0.124	1.0176	31.931	51.7
CSGT	B3LYP	aug-cc-pVDZ	yes	0.355	0.127	1.0816	31.552	15.2
CSGT	WP04	aug-cc-pVTZ	no	0.141	0.127	1.0218	31.870	226.5
GIAO	B3LYP	6-31G(d,p)	yes	0.218	0.129	1.0552	31.840	3.8
GIAO	WP04	6-31++G(d,p)	no	0.149	0.129	1.0140	31.988	4.2
CSGT	WP04	aug-cc-pVDZ	no	0.155	0.130	1.0272	31.828	13.3
GIAO	B3LYP	6-31++G(d,p)	yes	0.322	0.132	1.0690	31.705	5.3
GIAO	B3LYP	cc-pVDZ	yes	0.177	0.132	1.0442	31.685	3.7
CSGT	TPSS	aug-cc-pVDZ	yes	0.188	0.132	1.0334	31.550	6.2
GIAO	B3LYP	aug-cc-pVDZ	no	0.280	0.133	1.0554	31.719	15.2
CSGT	WP04	aug-cc-pVDZ	yes	0.226	0.135	1.0537	31.782	14.4
GIAO	BLYP	aug-cc-pVDZ	yes	0.220	0.137	1.0662	31.590	11.0
CSGT	B3LYP	aug-cc-pVDZ	no	0.230	0.139	1.0559	31.595	13.8
CSGT	WP04	cc-pVTZ	yes	0.149	0.139	1.0138	31.756	32.7
GIAO	B3LYP	aug-cc-pVTZ	no	0.246	0.140	1.0540	31.671	374.0
GIAO	*WP04*	*6-31G(d,p)*	*no*	*0.196*	*0.140*	*1.0068*	*32.065*	*1.8*
GIAO	B3LYP	cc-pVTZ	no	0.199	0.143	1.0460	31.708	52.9
CSGT	B3LYP	aug-cc-pVTZ	no	0.249	0.144	1.0527	31.631	238.1
GIAO	WP04	cc-pVDZ	no	0.257	0.146	0.9961	31.889	3.6
GIAO	B3LYP	6-31++G(d,p)	no	0.200	0.153	1.0407	31.754	7.0
GIAO	BLYP	6-31++G(d,p)	yes	0.204	0.153	1.0505	31.614	3.3
GIAO	B3LYP	6-311++G(d,p)	no	0.206	0.153	1.0405	31.984	8.6
CSGT	VSXC	aug-cc-pVDZ	yes	0.215	0.154	1.0172	31.318	6.2
CSGT	TPSS	aug-cc-pVDZ	no	0.167	0.158	1.0085	31.591	5.4
GIAO	WP04	6-31G(d)	no	0.308	0.159	0.9664	32.476	1.4
CSGT	WP04	cc-pVTZ	no	0.240	0.159	0.9885	31.799	30.9
GIAO	BLYP	cc-pVDZ	yes	0.175	0.162	1.0228	31.572	2.9
GIAO	BLYP	6-31G(d,p)	yes	0.188	0.162	1.0338	31.765	2.1
CSGT	BMK	aug-cc-pVDZ	no	0.404	0.164	1.1195	31.697	13.8
GIAO	B3LYP	6-31G(d,p)	no	0.191	0.165	1.0301	31.883	1.8
CSGT	BMK	aug-cc-pVDZ	yes	0.551	0.166	1.0427	31.631	14.6
GIAO	BLYP	aug-cc-pVDZ	no	0.299	0.172	1.0332	31.507	11.1
CSGT	B3LYP	cc-pVTZ	yes	0.200	0.173	1.0210	31.725	34.7
GIAO	B3LYP	cc-pVDZ	no	0.221	0.173	−0.9977	0.067	2.8
		ACD	no	0.185	0.176	0.9926	31.361	0.0
CSGT	VSXC	aug-cc-pVDZ	no	0.176	0.177	0.9958	32.296	6.0
GIAO	B3LYP	6-31G(d)	no	0.230	0.188	1.0088	31.545	1.2
CSGT	B3LYP	cc-pVTZ	no	0.226	0.200	1.0088	31.545	32.5
		ChemDraw	no	0.329	0.311	0.8918	30.928	0.0
CSGT	WP04	cc-pVDZ	yes	0.961	0.385	0.8278	30.215	2.7
CSGT	WP04	6-31++G(d,p)	yes	1.304	0.422	0.8675	30.965	4.2
CSGT	WP04	cc-pVDZ	no	1.111	0.435	0.8967	30.664	2.0
CSGT	B3LYP	cc-pVDZ	yes	1.001	0.452	0.8429	29.902	2.8
CSGT	B3LYP	6-31++G(d,p)	yes	1.260	0.473	0.8001	30.255	4.2
CSGT	WP04	6-31++G(d,p)	no	1.474	0.483	0.8739	30.697	3.5
CSGT	B3LYP	cc-pVDZ	no	1.139	0.504	0.7940	29.763	2.0
CSGT	VSXC	6-31++G(d,p)	yes	1.334	0.506	0.9126	31.056	5.8
CSGT	B3LYP	6-311G(d,p)	no	0.959	0.526	0.8159	29.938	5.0
CSGT	B3LYP	6-31++G(d,p)	no	1.423	0.533	0.7676	29.800	3.6
CSGT	VSXC	6-31++G(d,p)	no	1.495	0.570	0.6582	29.680	5.7

[a] Method for calculating chemical shieldings (see the Methods). [b] BLYP is a "pure" functional, B3LYP and WP04 are hybrid functionals, and BMK, TPSS and VSXC are "meta-GGA"-functionals. [c] The basis sets are denoted by their corresponding keywords in Gaussian; pVDZ and 6-31G denote double-ζ basis sets; pVTZ and 6-311G denote triple-ζ basis sets. [d] Here, "yes" means that chloroform was included as a solvent in a PCM-SCRF calculation; "no" refers to gas-phase calculations. [e] Scaled chemical shift δ = (intercept − isotropic magnetic shielding)/slope. [f] CPU time in hours for calculating the entire test set of 80 molecules on a typical Linux workstation.

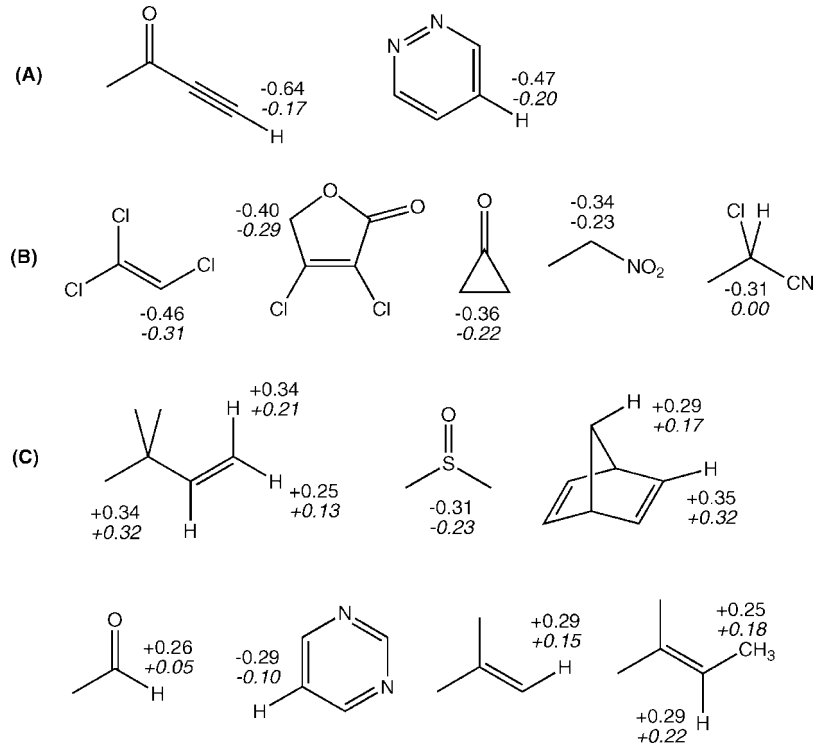


FIGURE 1. Changes of calculated ^{1}H chemical shifts on solvation. The numbers are differences between experimental values and those calculated in the gas phase (normal font) and in chloroform (italic) by the GIAO/B3LYP/cc-pVDZ method. The two compounds under (A) decrease the rms difference over the whole test set from 0.173 to 0.16 ppm, the five under (B) decrease it by another 0.01 ppm, as do the seven compounds under (C). The last bit of improvement, to the final value of 0.132 ppm, is due to the remaining 66 species.

CSGT), different functionals, and different basis sets that we tried. The results are sorted according to the rms error over the entire data set, after scaling with the parameters from the linear regression (the same table is available, along with the full set of data, as a spreadsheet in the Supporting Information, which allows for a sorting of the results according to different criteria).

In order to assess the cost at which the predictions listed in Table 1 come, we have added in the last column a measure of the CPU time for a calculation of the entire data set of 80 molecules (calculations were run on different computers, but scaled appropriately, by running some identical calculations on each machine, such that the numbers become comparable). A gas-phase calculation of all 80 molecules with the GIAO/B3LYP/6-31G* method took approximately 1 CPU hour on the computer used as a reference (a typical Linux workstation) and may thus be regarded as an internal cost-standard.

Inspection of Table 1 reveals a number of interesting observations:

(a) The accuracy of the predictions shows no correlation at all with the cost of the calculations. In fact one of the least costly methods (WP04/cc-pVDZ/SCRF) and one which is a hundred times more expensive (WP04/aug-cc-pVTZ/SCRF) have almost the same rms errors (0.115 and 0.112 ppm, respectively).

(b) Inclusion of solvent invariably leads to an improvement of the results, although to very different degrees in different methods. An analysis of these changes shows that the effect is largely due to dramatic improvements for a small number of otherwise problematic cases. For example, for the GIAO/B3LYP/cc-pVDZ method, where the rms error decreases from 0.173 to 0.132 ppm on inclusion of chloroform, almost all of that decrease is due to hydrogen atoms where the error in the gas-phase calculation is 0.25 ppm or more (see Figure 1). We were unable to detect common features of these molecules which would lead to rules of when accounting for solvent effects is particularly warranted.

(c) WP04 is the functional capable of making the most accurate predictions of ^{1}H chemical shifts over our whole test set by means of a gas-phase calculation, even with a relatively modest basis set (aug-cc-pVDZ). This may appear surprising because the parameters in this functional were optimized for it to reproduce experimental chemical shifts by SCRF calculations in chloroform. However, for WP04 the results improve much less on account of solvation than for B3LYP (or other hybrid functionals) which may serve to explain why WP04 is better able to predict ^{1}H chemical shifts in solution by gas-phase calculations. In contrast to B3LYP, the WP04 functional works almost equally well with Pople's as with Dunning's basis sets.

(d) Our investigations do not confirm the previous finding that the CSGT method requires larger basis sets than the GIAO method to achieve results of similar accuracy: Quite good results are obtained even with double-ζ basis sets, provided they are of the right type and contain diffuse functions: surprisingly, the Pople basis sets turn out to be singularly unsuitable for CSGT calculations, whereas the aug-cc-pVDZ basis set of Dunning allows to reach an accuracy that is comparable to the best GIAO calculations, with the B3LYP or the WP04 functionals. Increasing that basis set to triple-ζ quality (which results in a 20-fold increase in CPU time for our test set!) does not lead to any improvement.

(e) Also in the case of GIAO calculations, adding diffuse functions improves the results considerably in gas-phase calculations, but here the effect is strongly attenuated in chloroform.

(f) When it comes to general purpose functionals, pure functionals (we checked only BLYP) do not do quite as well as hybrid ones but, if solvent effects are included, the results

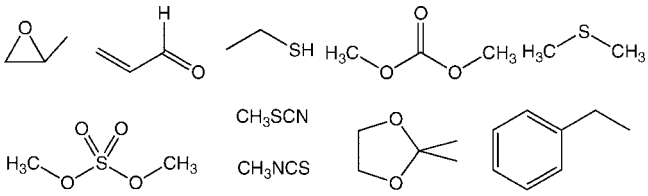


FIGURE 2. Molecules in the "probe" set.

can be quite good, and the savings in CPU times may warrant using a pure rather than a hybrid functional, provided a good basis set is used (e.g., with aug-cc-pVDZ, the rms error for BLYP/GIAO is 0.137 ppm).

(g) The new generation meta-GGA functionals do not do better than simple hybrid ones, and they have the disadvantage that they can only be used in conjunction with the CSGT method which, even with good basis sets, gives slightly worse results than GIAO (see above).

(h) Scaling by linear regression parameters significantly improves results in some, but not in all cases. Obviously the increment methods profit very little from this, because the parameters are already scaled on experimental quantities. The same applies to the WP04 functional, but not uniformly: thus, GIAO calculations with the aug-cc-pVDZ basis set in chloroform which give the best of all results after scaling only do that after scaling (which results in a decrease of the rms error from 0.281 to 0.103 ppm!).

(i) Regarding the performance of the increment methods, ACD (rms: 0.185) is clearly superior to the procedure in ChemDraw (rms: 0.329), but neither method can compete even with the most economical quantum chemical calculations (e.g., with GIAO/WP04/6-31G(d,p)/gas phase, rms: 0.140 after scaling, 1.8 CPU hours to calculate the entire test set). In addition, the increment methods are not usually suited for distinguishing stereoisomers which is often possible with quantum chemical methods.

In order to test to what extent the methods which offer the best cost/accuracy ratio are bound to the test set that was used to derive the scaling parameters, we carried out calculations on a "probe" set of 10 additional molecules shown in Figure 2 (the conformationally rigid molecules of the test set that had been used to optimize the parameters for the WP04 functional[26]), which show features not found in any molecules of the test set. The results of this exercise, which are documented in detail in the Supporting Information, are very encouraging in that the rms errors for the probe set were invariably equal or smaller than for the test set. The most notable exceptions to this rule were all calculations with the cc-pVTZ basis set where, for reasons that we do not understand, the rms error for the probe set was always 5−20% larger than for the test set (and mostly larger than for comparable calculations with the cc-pVDZ basis set!). Surprisingly, the best method for the "probe" set turned out to be the very economical GIAO/WP04/6-31G* (rms 0.08 with SCRF, 0.09 without).

Thus, it looks as though the scaling parameters derived in this work for the different methods are quite generally applicable.

Conclusions

We have tested combinations of various methods, functionals, and basis sets to calculate ^{1}H chemical shifts of a rather wide range of organic compounds. The most important conclusion of this study is that, contrary to what has been claimed in the literature, it is *not necessary* to employ very large basis sets in such calculations: double-ζ quality basis set give results that are just as good as those obtained with the triple-ζ or larger basis sets that are commonly used in chemical shift calculations, at a fraction of the cost in CPU time.

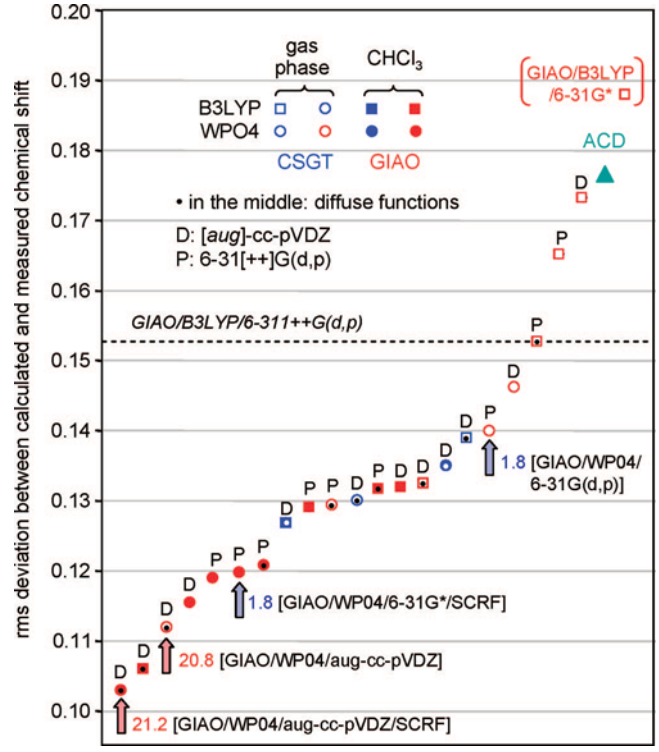


FIGURE 3. Graphical representation of some results from Table 1 that were obtained by calculations with double-ζ basis set (explanation of symbols, see inset). The most recommended methods (see Conclusions) are marked by arrows (with relative timings and description of the method). The dashed line corresponds to the level of theory that was recommended in ref 13.

Although this is not invariably true, it generally pays to model the solvent in which an NMR spectrum was taken as a polarizable continuum in an SCRF-type calculation.[38] The additional cost of this is insignificant (typically 10%), and one is then on the safe side with regard to cases that may constitute outliers in the gas phase. The price to pay for this is that, in some cases, SCRF calculations do not succeed on the first attempt, i.e., one may have to adjust the parameters of the cavitiy-defining algorithm. This was the case for 10 out of the 80 molecules of the test set, many of which contained *tert*-butyl groups, the cavities of which seem to be particularly problem-prone.[39]

When one has a choice, it is preferable to use the GIAO rather than the CSGT method, although we found, again contrary to what is reported in the literature, that the CSGT method does also not require very large basis sets to give quite good results, provided the right basis sets are used.

Regarding functionals, hybrid ones give good results even with the simple double-ζ basis sets, although the only pure functional we have tested is BLYP, which requires the aug-cc-

(38) In Gaussian, such calculations are invoked by the SCRF(solvent = $CHCl_3$) keyword.

(39) In all the cases where we encountered such problems, they could be remedied by adding "read" in the options for the SCRF keyword, and to add at the end of the input (after a blank line following the geometry), the word "NoSymmCav".

pVDZ basis set to give satisfactory results. Meta-GGA functionals, which can presently only be used in conjunction with the CSGT method, do not do better than simple hybrid ones. We found that the WP04 and the B3LYP functionals both give similarly good results after scaling with the regression parameters that we have determined with our test set.

Some of the results from calculations with double-ζ quality basis sets are represented schematically in Figure 3 which serves to illustrate the above conclusions. It shows in particular that a number of methods, including the very economical gas-phase GIAO/WP04/6-31G(d,p), do better than what was previously considered to represent the best cost/performance compromise, i.e., GIAO/B3LYP/6-311++G(d,p),[13] represented by the dashed horizontal line in Figure 3.

Recommendations. (a) The best cost/performance ratio is for GIAO calculations with the WP04 functional and the 6-31G* basis set plus account for solvation by an SCRF calculation (lowest blue arrow in Figure 3). With the WP04 functional (which requires an "iop" statement[26] to be invoked in Gaussian 03) scaling the results is not necessary, but it does improve the results slightly. The B3LYP functional does not fall far behind, but there scaling is necessary to achieve good results. (b) If one can afford to use the *aug*-cc-pVDZ basis set, either B3LYP or WP04 plus solvation give excellent results (average error close to 0.1 ppm, lowest red arrow in Figure 3), but these calculations use roughly 10 times more CPU time than those with the simple double-ζ basis sets. (c) If SCRF calculations present problems (which is sometimes the case), the best choice is gas-phase GIAO/WP04/*aug*-cc-pVDZ calculations (second lowest red arrow). If computational economy is a problem, the 6-31G(d,p) basis set gives also quite acceptable results (highest blue arrow in Figure 3).

A Caveat. The above results were obtained for molecules that have a single, well-defined conformation, or one that is significantly more stable than any others. In practice, the biggest difficulty in assigning experimental NMR spectra consists often in the fact that the observed spectra are averaged over several energetically proximate conformations that are in equilibrium at room temperature. Even a procedure that would give perfect predictions for single conformers would not be able to yield predictions in accord with experiment in such cases. What is needed there are accurate calculations of the relative free energies of the different conformers which translate into weighting factors for the chemical shifts of the individual conformers that allow to carry out a proper averaging. Molecular mechanics calculations which account for solvation by one or another method would in many cases be the method of choice for the calculation of relative free energies. Once these are available, chemical shifts can be calculated by increment methods or by the quantum chemical calculations listed above under "Recommendations". This approach has indeed been demonstrated to give quite good predictions in the case of penam β-lactams, where six stereoisomers could be clearly distinguished.[40]

In sum, we hope to have convinced the practitioners of ^{1}H NMR spectroscopy that quantum chemical calculations of chemical shifts are by no means something that should be left to specialists with access to humongous computing resources. Every modern desktop PC can nowadays carry out calculations of the type we recommend for molecules with up to, say, 50 heavy atoms and the associated hydrogens during times when such PCs usually do little else but heat rooms.

Acknowledgment. This work is part of project 200020-121747 of the Swiss National Science Foundation who made it in particular possible for R.J. to visit the Fribourg laboratories during the summer of 2008. P.R.R. thanks The Henry Dreyfus Teacher−Scholar Awards Program administered by The Camille and Henry Dreyfus Foundation for partial support of this research.

Supporting Information Available: (a) Complete ref 30; (b) structures of the 80 molecules in the test set; (c) table with the complete data from the calculations with the "probe" set depicted in Figure 2; (d) Excel spreadsheets containing all the raw data and the statistical workup for the test and for the "probe" set; (e) a complete Gaussian input to run a calculation on the entire test set or probe set of molecules ((d) and (e) are packed into a zip-archive).

(40) Wiitala, K. W.; Cramer, C. J.; Hoye, T. R. *Magn. Reson. Chem.* **2007**, *45*, 819.