A Comparison of Density Functional Methods for the Estimation of Proton Chemical Shifts with Chemical Accuracy

Paul R. Rablen,*,† Shoshannah A. Pearlman, and John Finkbiner

Department of Chemistry, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081-1397

Received: May 24, 1999; In Final Form: July 15, 1999

Fifteen procedures based on hybrid density functional theory were used to calculate magnetic properties for the carbon-bound hydrogen nuclei of 80 small to modest-sized organic molecules. The predicted isotropic shieldings derived from the various methods were compared with each other and also with solution experimental data. The computational methods investigated included the IGAIM and GIAO procedures, the 6-311++G-(d,p), 6-311++G(2df,p), and 6-311++G(3df,2p) basis sets, the B3LYP, B3P86, and B3PW91 hybrid density functionals, and molecular geometries optimized using both MP2 and B3LYP methods. Although agreement with experiment consistently improved as the basis set was enlarged, the improvement upon going from 6-311++G(2df,p) to 6-311++G(3df,2p) was extremely small, and even the difference between 6-311++G-(d,p) and 6-311++G(2df,p) was of a modest size. The GIAO and IGAIM procedures yielded very similar results in conjunction with the largest basis set, but GIAO suffered considerably less degradation than did IGAIM as the basis set size was decreased. The three functionals B3LYP, B3P86, and B3PW91 performed in an extremely similar fashion, although B3LYP proved marginally superior to the others. The method of geometry optimization also was found to make little difference. Of the computational methods investigated, the GIAO/B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) procedure probably represents the best compromise between accuracy and expense and yielded proton chemical shifts having a root-mean-square error of 0.15 ppm in comparison with solution experimental values after empirical linear scaling. The more expensive GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-31+G(d) method provided only a slightly lower root-mean-square error of 0.14 ppm.

Introduction

It goes without saying that nuclear magnetic resonance has proven an extraordinarily powerful spectroscopic technique over the past few decades. Although a tremendous wealth of chemical information can be derived from the interpretation of NMR spectra without the need for computational aids, the ability to predict chemical shifts accurately from ab initio molecular orbital calculations would further increase the utility of NMR spectroscopy. Much effort has been directed toward the goal of calculating accurate magnetic shielding tensors, and several excellent reviews of this topic are available in the recent literature.[1-6] For a number of magnetic nuclei, notably ^{13}C, there is already a strong history of success, and calculations have been used extensively to help in the interpretation of experimental carbon spectra. Just to mention two typical examples, calculated ^{13}C chemical shifts have frequently been used to help in the identification of reactive cationic species in solution,[7] and the calculated environmental dependence of the chemical shifts of amino acid carbon atoms can be used to help elucidate the three-dimensional structures of proteins.[8]

However, because proton chemical shifts span a much smaller range than do the chemical shifts of other nuclei, the conventional wisdom has long held that ab initio calculations cannot generally be used to predict proton chemical shifts with useful accuracy. For instance, since the typical proton chemical shift range is 0–12 ppm, while that of carbon is roughly 0–220 ppm, calculations for proton shifts would need to be at least an order

of magnitude more accurate in an absolute sense in order to be equally useful. Nonetheless, there has been progress in the calculation of proton magnetic shieldings. Chesnut and others have shown that proton shielding values for very small molecules can be obtained with ±0.1 ppm accuracy if very large basis sets are used and if rovibrational and correlation corrections are included,[9,10] and even the effect of temperature on the chemical shift of liquid water has been modeled.[11]

However, it is impractical to carry out such elaborate calculations for the larger molecules more likely to be of interest to experimental organic chemists. Here we have explored the possibility that useful results can be achieved using simpler calculations that are feasible with a minimum of user effort and a modest amount of CPU time. Chesnut has already noted that much of the error in proton magnetic shielding calculations is systematic in nature and thus amenable to correction by a linear scaling procedure. Following this logic, we have primarily focused on the ability of ab initio methods to give isotropic shielding values that are linearly related to the experimental values, even if the absolute agreement is only fair. Furthermore, we have used the chemical shift measured relative to tetramethylsilane (TMS), rather than the absolute shielding, as the target experimental quantity.

It is known that large basis sets are required to calculate accurate magnetic shielding values and also that a correction for electron correlation is desirable. Density functional theory (DFT) has emerged in the past decade as an approximate but computationally inexpensive method of treating electron correlation that often gives performance comparable to second-

† E-mail: prablen1@swarthmore.edu.

TABLE 1: Correlations between Different Computational Procedures Using the Same Geometry[a,b]

	1	2	3	4	5	6	7	8	9	10
1		0.99776	0.98306	0.99980	0.99972	0.99942	0.99793	0.99794	0.99954	0.99950
2	0.950		0.99062	0.99769	0.99777	0.99823	0.99963	0.99960	0.99756	0.99755
3	0.879	0.928		0.98277	0.98275	0.98496	0.99109	0.99100	0.98296	0.98304
4	1.003	1.054	1.122		0.99986	0.99954	0.99792	0.99795	0.99973	0.99970
5	1.008	1.059	1.128	1.005		0.99966	0.99803	0.99806	0.99989	0.99986
6	0.993	1.044	1.112	0.990	0.985		0.99854	0.99856	0.99962	0.99960
7	0.959	1.010	1.078	0.956	0.951	0.966		0.99999	0.99805	0.99806
8	0.961	1.012	1.081	0.958	0.953	0.968	1.002		0.99811	0.99812
9	1.016	1.067	1.137	1.013	1.008	1.023	1.057	1.055		0.99999
10	1.017	1.069	1.138	1.014	1.009	1.024	1.059	1.056	1.001	

[a] Numbers above the diagonal are correlation coefficients (r^2). Numbers below the diagonal are slopes. Nonzero intercepts were allowed in the fits, although the values are not shown. [b] Key: 1=IGAIM/B3LYP/6-311++G(3df,2p) 2=IGAIM/B3LYP/6-311++G(2df,p) 3=IGAIM/B3LYP/6-311++G(d,p) 4=GIAO/B3LYP/6-311++G(3df,2p) 5=GIAO/B3LYP/6-311++G(2df,p) 6=GIAO/B3LYP/6-311++G(d,p) 7=IGAIM/B3P86/6-311++G(2df,p) 8=IGAIM/B3PW91/6-311++G(2df,p) 9=GIAO/B3P86/6-311++G(2df,p) 10=GIAO/B3PW91/6-311++G(2df,p). All the methods 1−10 use the MP2/6-31+G(d) optimized molecular geometries.

order Møller−Plesset theory (MP2) and has already yielded promising results for the calculation of NMR shieldings.[12] We have thus chosen to use DFT methodology, in the hopes of achieving satisfactory agreement with experiment even with modest-sized molecules for which MP2 calculations using large basis sets are impractical.

Magnetic shieldings can depend strongly on the conformation of a molecule. To avoid ambiguities and complications having to do with multiple or highly flexible conformations, we have focused here on organic compounds that have a single clearly preferred conformer and little conformational flexibility. Hydrogen bonding strongly influences proton chemical shifts and thus can lead to severe complications in comparing gas-phase calculations to experimental solution data. Consequently, in this study we have also chosen to avoid protons that can form hydrogen bonds, and so have included only carbon-bound hydrogen nuclei.

Results and Discussion

Computational Procedures. To explore more thoroughly how well DFT calculations might be able to reproduce experimental proton magnetic shieldings, several different procedures, rather than just a single method, were tested. The three Pople-type basis sets 6-311++G(d,p), 6-311++G(2df,p), and 6-311++G(3df,2p) were selected to provide a reasonable sampling of basis set size. Only fairly large basis sets were included since it has already been demonstrated that basis sets at least as large as 6-311++G(d,p) provide considerably better agreement with experiment, at least for very small molecules, than is obtained with smaller basis sets.[13] Although the focus here is on proton chemical shifts, it is important to improve the treatment of molecules in their entirety, and not just of the hydrogen atoms, when expanding the basis set. The basis sets listed above were chosen, despite the fact that the first two do not differ in the number of basis functions at hydrogen, because they have been widely used and validated for a variety of purposes.[14]

For the density functional calculations, three popular hybrid methods using Becke's three-parameter functional[15] were selected, namely B3LYP,[16] B3P86,[17] and B3PW91.[18] For the calculation of the magnetic properties themselves, both the GIAO[19−22] and IGAIM,[13,23−25] procedures were used, the latter representing a derivative of the CSGT method.[13,23,24]

The MP2/6-31G(d) and very similar MP2/6-31+G(d) methods are known to provide molecular geometries with high accuracy and reliability,[26] and so MP2/6-31+G(d)-optimized geometries were used for the majority of the magnetic property

calculations. However, B3LYP optimizations were also carried out, with the three different basis sets 6-31+G(d), 6-311++G-(d,p), and 6-311++G(2df,p), to explore the effect of molecular geometry on the calculation of magnetic shielding. The complete set of calculated proton isotropic shielding values is provided in Tables S1a, S1b, S2a, and S2b in the Supporting Information. The names of the 80 compounds for which our calculations were compared with experiment appear in Table 4.

Internal Comparison of Computational Procedures. Before examining the degree of agreement between theory and experiment, it is of interest to consider how well the different computational procedures agree with each other. Such comparisons should reveal how sensitive the predictions are to differences between the computational methods. Sensitivity to such differences can be both a blessing and a curse. A large degree of sensitivity would mean that selection of the optimal procedure is important, and that even modest changes to an established protocol are potentially perilous. However, at least in principle, variation of the computational procedure would provide an obvious avenue for improving the quality of results. On the other hand, a relative lack of sensitivity, particularly with the larger basis sets, would provide confidence in robustness of the predictions. Insensitivity would also suggest that the procedures examined are at or near the limits of what can be expected with density functional methods and without explicit inclusion of rovibrational and solvent effects. However, this limit having been achieved, few simple options would remain for further improving agreement with experiment.

Table 1 shows the correlation coefficients and slopes for the best fit lines that relate the various computational procedures in a pairwise manner. These comparisons include the 80 compounds listed in Table 4 (and in Table S1a and S1b) for which experimental data are available, and also another 25 molecules, listed in Tables S2a and S2b, for which solution experimental data could not be located but for which calculations were nonetheless performed. Perhaps not surprisingly, the proton magnetic shieldings derived from the various computational procedures are very highly correlated, with correlation coefficients ranging from a low of 0.98275 to a high of 0.99999.

Values of the correlation coefficients close to 1 do not, of course, necessarily indicate close *absolute* agreement. In fact, the slopes of the best fit lines, also given in Table 1, vary by as much as ±7% from the ideal value of 1. This variation indicates that the absolute level of agreement between the various computational procedures is considerably less than the degree of correlation between them. In other words, the different computational procedures do yield significantly different shield-

DFT Methods to Estimate Proton Chemical Shifts

J. Phys. Chem. A, Vol. 103, No. 36, 1999 **7359**

TABLE 2. Correlations between Calculations Using Different Geometries.[a,b]

	5	11	13	15	6	12	14
5		0.99941	0.99912	0.99910			
11	0.987		0.99993	0.99992			
13	0.991	1.004		0.99998			
15	0.989	1.003	0.999				
6						0.99938	0.99906
12					1.001		0.99992
14					0.990	1.004	

[a] Numbers above the diagonal are correlation coefficients (r^2). Numbers below the diagonal are slopes. Note that only comparisons between calculations at the same level of theory, where *only* the geometry differs, are shown. [b] Key: 5 = GIAO/B3LYP/6-311++G-(2df,p)//MP2/6-31+G(d) 6 = GIAO/B3LYP/6-311++G(d,p)//MP2/6-31+G(d) 11 = GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-31+G(d) 12 = GIAO/B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) 13 = GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-311++G(d,p) 14 = GIAO/B3LYP/6-311++G(d,p)//B3LYP/6-311++G(d,p) 15 = GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-311++G(2df,p)

ing values, but most of the divergence can be eliminated by linear scaling.

Table 1 indicates that GIAO shielding calculations are less sensitive to basis set size than are IGAIM calculations. For instance, the GIAO isotropic shieldings obtained using the largest basis set (6-311++G(3df,2p)) are related to those with the two smaller basis sets (6-311++G(2df,p) and 6-311++G-(d,p)) with correlation coefficients of 0.99986 and 0.99954, respectively, whereas the same relationships between the IGAIM shieldings are characterized by correlation coefficients of only 0.99776 and 0.98306. The IGAIM procedure suffers a major loss of consistency on going to 6-311++G(d,p) from 6-311++G-(2df,p), whereas the GIAO procedure is apparently quite insensitive to this change.

It is also encouraging that the IGAIM and GIAO values agree very closely with each other when the largest basis set is used, even if there is some divergence with the smaller basis sets. The GIAO and IGAIM shieldings are related by a correlation coefficient of 0.99980 at 6-311++G(3df,2p), although the agreement drops to 0.99777 and 0.98496 for the two somewhat smaller basis sets. Thus GIAO and IGAIM do converge to the same limit as the basis set is enlarged.

It is also readily apparent that, at least for the calculation of GIAO proton magnetic shieldings for these 105 molecules, the three density functionals B3LYP, B3P86, and B3PW91 offer essentially identical performance. With the intermediate 6-311++G(2df,p) basis set, the GIAO shielding values for B3P86, B3PW91, and B3LYP are related by correlation coefficients of 0.99989, 0.99986, and 0.99999. The IGAIM calculations demonstrate almost as much internal consistency, yielding correlation coefficients of 0.99963, 0.99960, and 0.99999.

Table 2 addresses the influence of molecular geometry on the calculated shieldings. Correlations are shown between GIAO/B3LYP calculations carried out on molecular geometries optimized at four different levels of theory: MP2/6-31+G(d), B3LYP/6-31+G(d), B3LYP/6-311++G(d,p), and B3LYP/6-311++G(2df,p). Both the 6-311++G(d,p) and 6-311++G-(2df,p) basis sets were used for the actual computation of the magnetic properties in these comparisons. The correlation coefficients in Table 2 demonstrate that there is very little difference in the quality of results obtained using the variously optimized molecular geometries. Importantly, it seems that using geometries optimized with significantly smaller basis sets, or even using a different treatment of electron correlation (MP2

as opposed to DFT), does not degrade the quality of the DFT calculations of magnetic properties.

Comparisons with Experiment. Our primary interest in the present study was to investigate the possibility of using DFT calculations to predict with useful accuracy proton chemical shifts determined under typical experimental conditions. Since most measurements are in fact made in solution, frequently with $CDCl_3$ as the solvent, we have sought to model the solution experimental values directly. Experimental spectra determined under a fairly consistent set of conditions were deemed desirable in order to minimize the noise resulting from variation in conditions, and so with one exception[27] the experimental numbers were drawn from only three sources: the Sadtler index,[28] the Properties of Organic Compounds database,[29] and the spectroscopy text by Silverstein, Bassler, and Morrill.[30] The vast majority of the experimental measurements were carried out with either $CDCl_3$ or CCl_4 as the solvent. Comparisons of the calculated shielding values with solution experimental chemical shifts appear in Table 3.

It is readily apparent that theory and experiment are highly correlated, although not so highly as the various theoretical levels are with each other. The correlation coefficients between calculated absolute shielding and experimental chemical shift range from a low of 0.97237 for IGAIM/B3LYP/6-31++G-(d,p) to a high of 0.99685 for GIAO/B3LYP/6-311++G(3df,-2p). The slopes are close to -1, since absolute shielding and chemical shift are related by the equation $\delta = \sigma_0 - \sigma$, where δ is the chemical shift, σ is the absolute shielding, and σ_0 is the absolute shielding of the standard (in this case, TMS). That the absolute value of the slope is close to 1, rather than significantly greater or less than 1, is also encouraging, as it suggests reasonable absolute agreement between theory and experiment, in addition to close correlation. With the exception of IGAIM/B3LYP/6-311++G(d,p), which shows the worst correlation with experiment of any of the procedures, the slopes all have an absolute value somewhat less than 1. This finding is consistent with Chesnut's observation of a scaling factor of 0.9428 that relates GIAO/HF/6-311++G(d,p) calculated proton shieldings to the corresponding experimental values for a smaller set of molecules.[1]

It is highly encouraging to note that correlation with experiment consistently improves as the basis set is enlarged, as expected for a well-behaved theoretical method. However, with the GIAO procedure, increasing the basis set from 6-311++G-(d) to 6-311++G(2df,p) yields only marginal improvement, raising the correlation with experiment from 0.99592 to 0.99666, while further improvement of the basis set to 6-311++G(3df,-2p) results in an even smaller increment, to a correlation coefficient of 0.99685. The figures quoted above refer to the calculations carried out using the MP2/6-31+G(d) optimized molecular geometries, but use of the DFT optimized geometries yields equally good agreement with experiment. The best compromise between quality and expense is represented by the GIAO/B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) procedure. Figure 1 shows the comparison with experiment for this method, to emphasize in a visual manner the quality of fit. The additional expense of the GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-31+G(d) procedure yields a very small but nonetheless significant improvement in the correlation with experiment and might be justified for some applications.

Linear Scaling. Correlation coefficients close to 1 do not necessarily indicate good absolute agreement between theory and experiment. However, the sort of disagreement that remains is easily eliminated via linear scaling. For example, linear scaling

TABLE 3: Correlation between Calculated and Experimental Proton Magnetic Shielding Values

method	functional	basis[a]	geom[b]	r^2	slope[c]	int[d]	RMSE[e]
IGAIM	B3LYP	3df,2p	MP2	0.99679	−0.9370	29.70	0.14
IGAIM	B3LYP	2df,p	MP2	0.99344	−0.9819	31.09	0.20
IGAIM	B3LYP	d,p	MP2	0.97237	−1.0359	32.34	0.41
GIAO	B3LYP	3df,2p	MP2	0.99685	−0.9352	29.74	0.14
GIAO	B3LYP	2df,p	MP2	0.99666	−0.9294	29.76	0.14
GIAO	B3LYP	d,p	MP2	0.99592	−0.9430	30.23	0.16
IGAIM	B3P86	2df,p	MP2	0.99334	−0.9719	30.70	0.20
IGAIM	B3PW91	2df,p	MP2	0.99328	−0.9694	30.65	0.20
GIAO	B3P86	2df,p	MP2	0.99651	−0.9220	29.44	0.15
GIAO	B3PW91	2df,p	MP2	0.99644	−0.9204	29.41	0.15
GIAO	B3LYP	2df,p	B3/S	0.99679	−0.9422	30.10	0.14
GIAO	B3LYP	d,p	B3/S	0.99616	−0.9567	30.60	0.15
GIAO	B3LYP	2df,p	B3/M	0.99657	−0.9380	30.07	0.15
GIAO	B3LYP	d,p	B3/M	0.99591	−0.9526	30.58	0.16
GIAO	B3LYP	2df,p	B3/L	0.99662	−0.9392	30.15	0.14

[a] Basis set: all basis sets are 6-311++G, with the polarization functions indicated in the table. [b] Molecular geometry optimization: MP2 ⟹ MP2/6-31+G(d); B3/S ⟹ B3LYP/6-31+G(d); B3/M ⟹ B3LYP/6-311++G(d,p); B3/L ⟹ B3LYP/6-311++G(2df,p). [c] Slope of best fit line. [d] Intercept of best fit line in ppm units. [e] Root-mean-square error between calculation and experiment in ppm units.
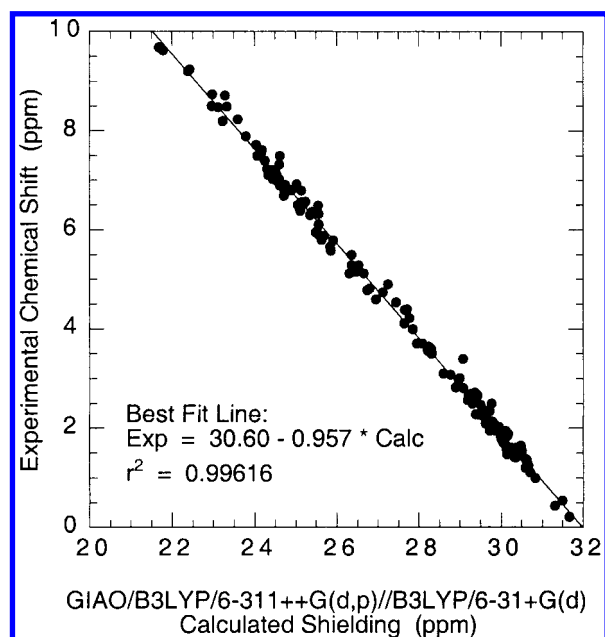
Figure 1. Comparison of experimental chemical shifts with proton magnetic shieldings calculated using the B3LYP/GIAO/6-311++G-(d,p)//B3LYP/6-31+G(d) level of theory (model B). The best fit line has the equation exp = 30.60 − 0.957 × calc and the correlation coefficient r^2 is 0.99616.

is already widely and reliably used for vibrational frequencies that are calculated via ab initio theory.[31] It was our hope that a major portion of the error in our DFT-based proton magnetic shielding calculations could similarly be removed. Indeed, Chesnut has pointed out previously that, particularly in the case of protons, the linear scaling approach is a promising and quite reasonable means by which to improve agreement with experimental chemical shifts.[1]

The calculations reported here correspond to the gas phase, while all of the experimental chemical shifts used for comparison were determined in solution. It is known that proton chemical shifts depend to some extent on environmental factors such as solvent and temperature.[32] It is thus entirely possible that the calculations have greater intrinsic accuracy than suggested by the comparisons in Table 3. We hoped that in nonpolar solvents such as CDCl$_3$ a major portion of any solvent effect would be accounted for in the linear scaling procedure. Comparison with gas-phase proton chemical shifts would also be of interest, but relatively little such data are available in the literature,[33] since most NMR experiments are carried out in solution.

The equation for the best fit line relating the calculated shieldings to the experimental data can be used to convert the DFT results into scaled chemical shift predictions. This operation has been carried out for each of the fifteen computational procedures examined, and the root-mean-square errors (RMSEs) relating the scaled chemical shift predictions with experiment appear in the rightmost column of Table 3. While perhaps not quite as small as one might like, the 0.14 ppm RMSE for the best procedures nonetheless indicates that DFT calculations of proton chemical shifts, followed by linear scaling, do in fact provide a chemically useful degree of accuracy and reliability. The procedure that we recommend as the best compromise between accuracy and economy, GIAO/B3LYP/6-311++G-(d,p)//B3LYP/6-31+G(d) ("model B"), yields only a very marginally greater RMSE of 0.15 ppm than does the more expensive GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-31+G-(d) ("model A").

The DFT methods that we have investigated might not be the best ones conceivable, and it is quite possible that a still more economical method could be found that would give similar or even better accuracy. However, the relative insensitivity of the correlation coefficients in Table 3 to variations in the computational procedure might indicate that relatively little improvement can be expected without explicit treatment of rovibrational corrections and/or solvent effects. The RMSE of ±0.14 ppm reported here compares favorably with the standard errors of ±0.21 ppm and ±0.11 ppm reported previously by Chesnut for smaller series of compounds.[1,9]

Discussion of Individual Cases. Table 4 lists the 80 individual compounds used to make comparisons between theory and experiment and gives the RMSE between the calculated and experimental chemical shifts for each, using both model A and model B. The molecules are listed in order of decreasing quality of agreement between experiment and model A. It is encouraging to note that very few cases show large deviations from experiment. There are only three or four molecules out of 80 in which the RMSE is greater than 0.3 ppm, only one molecule where the RMSE is greater than 0.4 ppm, and none where the RMSE is above 0.5 ppm. The calculations thus not only reproduce the experimental results accurately in aggregate, but also are quite consistent in the degree of accuracy which they provide for individual molecules.

Table 5 lists the individual hydrogen atoms in the most problematic compounds. Every molecule in which any proton shift is predicted by either model A or model B with an absolute error greater than 0.25 ppm is included. Even by this more

DFT Methods to Estimate Proton Chemical Shifts

J. Phys. Chem. A, Vol. 103, No. 36, 1999 **7361**

TABLE 4: Deviation between Calculated and Experimental Chemical Shifts (ppm)

compound[a]	source[b]	# Res[c]	model A[e]	model B[f]	compound[a]	source[b]	# Res[c]	model A[e]	model B[f]
trans−1,2-dichloroethylene	S-6742	1	0.00	0.06	cyclohexane	S-1196	1	0.11	0.12
acetone	S-9288	1	0.00	0.00	cyclopropane	M-214	1	0.11	0.09
2-chloro-2-methylpropane (tBuCl)	S-6768	1	0.00	0.03	thiophene	S-10285	2	0.11	0.12
cyclopentanone	S-1195	1	0.00	0.00	propionitrile	S-1877	2	0.12	0.12
chlorodimethylsilane	S-53563	1	0.01	0.07	3-chloropropene	S-7833	4	0.12	0.14
tetrahydropyran	S-7566	2	0.01	0.03	2-butyne	S-8170	1	0.12	0.12
anisole	S-9115	4	0.03	0.02	*tert*-butylacetylene	S-13789	2	0.12	0.14
1,4-pyrazine	S-553	1	0.03	0.00	2-cyclohexen-1-one	S-9880	3	0.12	0.12
chlorobenzene	S-714	1	0.03	0.03	2-chloropropionitrile	S-45921	2	0.13	0.16
oxirane	M-214	1	0.03	0.02	2-methylpropanenitrile	P-78−82−0	2	0.13	0.15
cyclohexanone	S-10208	2	0.04	0.04	*tert*-butylethylene	S-5318	4	0.13	0.20
2-chloropropene	S-61833	2	0.04	0.03	2-methylpropene (isobutylene)	S-31633	2	0.14	0.18
thiirane	M-214	1	0.04	0.06	furfural	S-10203	4	0.14	0.14
2-methyl-2-nitropropane (tBuNO$_2$)	S-19967	1	0.04	0.04	bicyclobutane	J-5272[g]	3	0.14	0.12
chloroethane	P-75-00-3	2	0.04	0.07	cyclopentane	S-3435	1	0.14	0.16
methanol (not incl. OH proton)	P-67-56-1	1	0.04	0.02	acetaldehyde	S-10493	2	0.14	0.11
furan	S-16937	2	0.05	0.03	acetyl chloride	S-8780	1	0.15	0.21
2-chloropropane	S-6866	2	0.05	0.03	1,1-dichloroethylene	S-6385	1	0.15	0.13
cyclobutanone	S-5006	2	0.06	0.06	acetonitrile	S-9154	1	0.15	0.16
benzene	S-3429	1	0.06	0.11	indole (not including NH)	S-473	4	0.15	0.14
1-chloro-2-methylpropene	S-53505	3	0.07	0.06	nitromethane	S-9146	1	0.15	0.19
tetrahydrofuran	S-14667	2	0.07	0.10	cyclopentene	S-5162	3	0.16	0.17
pyridine	S-10200	3	0.07	0.07	nitrobenzene	S-4	2	0.16	0.14
N-methylpyrrolidine	S-5091	3	0.07	0.07	cyclohexene	S-3409	3	0.16	0.20
2-cyclopenten-1-one	S-14728	4	0.07	0.07	butyrolactone	M-214	3	0.16	0.15
pivalonitrile (tBuCN)	S-21352	1	0.08	0.06	cyclopropanone	M-214	1	0.17	0.21
methyl acetate (Z conformer)	S-10261	2	0.08	0.04	2,5-norbornadiene	S-16943	3	0.18	0.19
cyclobutane	M-214	1	0.08	0.08	nitroethane	S-2	2	0.19	0.24
N-methylpyrrole	S-21175	3	0.08	0.08	isoxazole	S-12784	3	0.19	0.20
1,1,1-trichloroethane	S-9171	1	0.08	0.10	pyrimidine	S-33550	3	0.20	0.17
oxetane	M-214	2	0.08	0.16	*p*-benzoquinone	S-10391	1	0.20	0.24
N,N-dimethylaniline	S-1	4	0.08	0.08	2-methyl-2-butene	S-3411	2	0.20	0.26
methyl *tert*-butyl ether	S-19010	2	0.08	0.12	trichloroethylene	S-9266	1	0.20	0.33
N,N-dimethylacetamide	S-8875	3	0.09	0.11	cyclobutene	M-217	2	0.21	0.20
fluorobenzene	S-8774	1	0.09	0.12	methylthiazole	S-49793	3	0.22	0.28
N-methylpiperidine	S-2849	3	0.09	0.11	chloroform	S-10513	1	0.23	0.05
N,N-dimethylformamide	S-9537	3	0.10	0.06	dimethyl sulfoxide	P-67-68-5	1	0.28	0.37
toluene	S-10216	2	0.10	0.14	1,2-pyrazine	M-219	2	0.34	0.34
dichloromethane	S-6401	1	0.10	0.06	3,4-dichloro-2(5H)-furanone	S-53884	1	0.38	0.37
1,1,1-trichloroacetone	S-61585	1	0.11	0.13	3-butyn-2-one	S-13388	2	0.43	0.44

[a] Compounds are in order from best agreement with experiment to worst. [b] Source of experimental spectrum; S = Sadtler Index; POC = Properties of Organic Compounds database; M = Silverstein, Bassler, and Morrill; number is compound number for S, CAS registry number for P, and page number for M. [c] Number of distinct resonances in experimental ^{1}H NMR spectrum. [d] RMSE deviation of calculated and experimental chemical shifts. [e] Predicted chemical shift derived from linear scaling model A: GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-31+G(d). Predicted shift = 30.10−0.9422 × calculated shielding. [f] Predicted chemical shift derived from linear scaling model B: GIAO/B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d). Predicted shift = 30.60−0.9567 × calculated shielding. [g] Wiberg K. B.; Lavanish, J. M. *J. Am. Chem. Soc.* **1966**, *88*, 5272−5275.

stringent test, very few chemical shifts are predicted with poor accuracy. The worst case by far is the alkynyl proton of 3-butyne-2-one, for which experiment and theory disagree by 0.61 ppm. The second worst case is 1,2-pyrazine, for which the hydrogens more distant from the nitrogens are predicted with an error of 0.47 ppm. The protons of 3,4-dichloro-2(5H)-furanone are predicted with an error of 0.38 ppm, and also one of the ring protons of 4-methylthiazole is predicted with an error of 0.31 ppm. There is no other hydrogen atom for any compound where the error is greater than 0.30 ppm.

It is interesting to note that every one of the four protons for which the calculated shielding deviates from experiment by more than 0.30 ppm occurs in a molecule with a polar conjugated π system. The solvent might be expected to interact with these polar and highly polarizable molecules especially strongly, which could be the source of the unusually large deviations.

Unsuccessful Exploration of Other Approaches for Improving the Computational Methodology. We also explored the possibility of using basis set additivity to improve the economy of calculation. The G-1, G-2, and G-3 procedures employ this approach, using a series of smaller calculations to obtain an energy that is supposed to be equivalent to QCISD-

(T)/6-311+G(3df,2p), a level of theory which it would be prohibitively expensive actually to calculate for most molecules.[31] We found that the contributions of additional diffuse and polarization functions on going from 6-311G(d,p) to 6-311++G(3df,2p) were in fact highly additive. However, the savings in computer time associated with dividing a single large calculation into a series of smaller ones was negligible. The economy of this approach in ab initio calculations lies mostly in avoiding high-level correlation treatments with large basis sets, by using cheaper correlation treatments with the largest basis sets and smaller basis sets with the most computationally expensive electron correlation treatments. However, since the calculations performed here used DFT to include the effects of electron correlation, there was little advantage to the divide-and-conquer approach.

We also explored the possibility that a more elaborate empirical correction scheme might be used to further improve agreement with experiment. To this end, we computed the average bond angle; range of bond angles (maximum minus minimum); coordination number; and number of bonded hydrogen atoms at each of the carbon atoms for which proton chemical shift values were calculated. In addition, the Mulliken

TABLE 5: Details for All Molecules for which a Deviation of >0.25 ppm Occurs.

compound	atom description	exp	model A	model B
3-butyn-2-one	methyl	2.38	2.31	2.32
	alkyne	3.40	2.79	2.78
1,2-pyrazine	meta to N	7.50	7.03	7.04
	ortho to N	9.24	9.13	9.14
3,4-dichloro-2(5H)-furanone	unique	4.90	4.52	4.53
4-methylthiazole	methyl	2.50	2.48	2.57
	alpha to S	6.92	6.68	6.65
	between N and S	8.72	8.41	8.32
cyclobutene	alkane	2.57	2.64	2.67
	alkene	5.95	6.24	6.21
pyrimidine	meta to N	7.32	7.03	7.06
	ortho to N	8.74	8.57	8.61
	between N's	9.21	9.15	9.18
dimethyl sulfoxide	unique	2.50	2.22	2.13
butyrolactone	aliphatic	2.08	2.08	2.04
	alpha to C=O	2.31	2.27	2.22
	alpha to ester O	4.38	4.10	4.14
indole[a]	position 3	6.38	6.51	6.57
	position 2	6.76	7.02	6.93
	positions 5, 6, & 7	7.03	7.10	7.13
	position 4	7.50	7.50	7.56
2-methyl-2-butene	methyl (lone)	1.48	1.73	1.76
	geminal methyls (av)	1.58	1.75	1.77
	alkene	5.12	5.35	5.43
2,5-norbornadiene	methylene	1.96	2.20	2.16
	bridgehead (methine)	3.52	3.54	3.51
	alkene	6.69	6.89	6.95
nitroethane	methyl	1.55	1.43	1.41
	methylene	4.40	4.17	4.09
cyclohexene	aliphatic methylene	1.62	1.71	1.71
	allylic methylene	1.97	2.10	2.14
	alkene	5.58	5.80	5.86
trichloroethylene	unique	6.49	6.29	6.16
tert-butylethylene	methyl	1.00	1.07	1.11
	alkene (cis to tBu)	4.78	4.94	5.01
	alkene (trans to tBu)	4.82	4.91	4.96
	alkene (geminal to tBu)	5.80	5.98	6.07

[a] The position numbers for indole given in this table follow the standard numbering scheme for the indole ring system.

atomic charges of the protons themselves were computed. However, only with the latter was any correlation found with the residual error in the scaled chemical shift predictions. The residual error from model A correlated with the Mulliken charges with $r^2 = 0.13$, and a best-fit equation gave residual error $= 0.275-1.566 \times$ Mulliken charge. Subtraction of a correction factor derived from this equation reduced the RMSE for the 80 compounds tested from 0.14 to 0.13 ppm for model A. While apparently significant, the improvement in quality does not seem great enough to warrant the concomitant computational complications.

Finally, we also explored the possibility of using MP2 calculations to derive magnetic shieldings, either on their own or as a means to provide corrections to those obtained from the DFT calculations. However, even with the 6-31+G(d) basis set, the CPU time and disk space requirements of the MP2 calculations proved impractical for molecules of the size of those investigated here.

Summary

Calculations of the magnetic shielding of the carbon-bound hydrogen nuclei in a series of 80 small to medium-sized organic molecules were carried out using fifteen procedures based on hybrid density functional theory. The different methods studied encompassed two procedures for computing magnetic properties (GIAO and IGAIM) as well as a variety of basis sets, density functionals, and molecular geometries. Comparisons among the different theoretical methods and between theoretical predictions

and solution experimental data yielded the following major conclusions and observations.

(1) Although the calculated shieldings and the experimental chemical shifts do not agree very well in an absolute numerical sense, they are in fact very highly correlated in a linear fashion. Consequently, linear scaling of the calculated magnetic properties yields greatly improved accuracy. After application of such scaling, GIAO/B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) yields an RMSE of 0.15 ppm and GIAO/B3LYP/6-311++G(2df,p)//B3LYP/6-31+G(d) yields an RMSE of 0.14 ppm. Over the entire set of 80 molecules, there are only four protons for which the scaled predictions of chemical shift are in error by more than 0.30 ppm.

(2) Although the GIAO and IGAIM procedures converge to the same limit as the basis set size is increased, GIAO suffers less degradation in quality with smaller basis sets than does IGAIM.

(3) With the GIAO method, there is little improvement in the correlation with experiment when the basis set is enlarged beyond 6-311++G(d,p), and virtually no improvement at all when the basis set is enlarged past 6-311++G(2df,p).

(4) Whether the molecular geometry is optimized using MP2/6-31+G(d) or using B3LYP with any of several different basis sets has almost no effect on the predicted magnetic properties.

(5) Whether the B3LYP, B3P86, or B3PW91 density functional is used makes very little difference for the calculated magnetic properties, although the correlation with experiment is very slightly superior for B3LYP.

DFT Methods to Estimate Proton Chemical Shifts

J. Phys. Chem. A, Vol. 103, No. 36, 1999 **7363**

(6) The largest errors, and in fact all of the errors that are in excess of 0.30 ppm, occurr in polar conjugated π systems.

Calculations

The *Gaussian 94*[34] and *Gaussian 98*[35] packages were used to carry out all ab initio calculations. Standard Pople-type basis sets were used.[26] Initial geometry optimizations for all molecules were carried out at HF/6-31G*, followed by vibrational frequency calculations to verify the stationary point to be a minimum (no negative force constants). Subsequent MP2/6-31+G(d) optimizations were carried out with the same symmetry, and the resulting geometries used for the majority of the magnetic property calculations. B3LYP optimizations were also performed, with the three different basis sets 6-31+G(d), 6-311++G(d,p), and 6-311++G(2df,p), to explore the effect of geometry on the calculated shielding values. Magnetic property calculations were carried out using both the GIAO and IGAIM procedures coupled with the three Pople-type basis sets 6-311++G(d,p), 6-311++G(2df,p), and 6-311++G(3df,2p). Density functional calculations were carried out using the B3LYP, B3P86, and B3PW91 keywords.

Supporting Information Available: Complete tabulation of calculated proton isotropic shielding values; structural drawings with atomic numbering schemes; MP2/6-31+G(d) optimized geometries in z-matrix form. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Chesnut, D. B. *The Ab Initio Computation of Nuclear Magnetic Resonance Chemical Shielding*, in Lipkowitz, K. B.; Boyd, D. B., Eds.; *Rev. Comput. Chem.* **1996**, *8*, 245−297.

(2) de Dios, A. C. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 229−278.

(3) Chesnut, D. B. *The Ab Initio Computation of Nuclear Magnetic Resonance Chemical Shielding*. In *Annual Reports on NMR Spectroscopy*, Webb, G. A., Ed.; Academic Press: London, 1994; Vol. 29, pp 71−122.

(4) Kutzelnigg, W.; Fleischer, U.; Schindler, M. *The IGLO-Method: Ab-Initio Calculation and Interpretation of NMR Chemical Shifts and Magnetic Susceptibilities*. In *NMR, Basic Principles and Progress*; Diehl, P.; Fluck, E.; Günther, H.; Kosfield, R.; Seelig, J. Eds.; Springer-Verlag: Berlin, 1990; Vol. 23, pp 165−262.

(5) Jameson, C. J. In *Nuclear Magnetic Resonance*; Webb, G. A., Ed. (Specialist Periodical Reports); The Royal Society of Chemistry: London, 1995; Vol. 24.

(6) Helgakar, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293−352.

(7) Three of the many examples in the recent literature: (a) Olah, G. A.; Shamma, T.; Burrichter, A.; Rasul, G.; Prakash, G. K. S. *J. Am. Chem. Soc.* **1997**, *119*, 12923−12928. (b) Wiberg, K. B.; McMurdie, N. *J. Am. Chem. Soc.* **1994**, *116*, 11990−11998. (c) Haw, J. F.; Nicholas, J. B.; Xu, T.; Beck, L. W.; Ferguson, D. B. *Acc. Chem. Res.* **1996**, *29*, 259−267.

(8) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491−1496.

(9) Chesnut, D. B. *Chem. Phys.* **1997**, *214*, 73−79.

(10) Sauer, S. P. A.; Spirko, V.; Paidarova, I.; Kraemer, W. P. *Chem. Phys.* **1997**, *214*, 91−101.

(11) Nymand, T. M.; Astrand, P. O.; Mikkelsen, K. V. *J. Phys. Chem. B* **1997**, *101*, 4105−4110.

(12) Rauhut, G.; Puyear, S.; Wolinski, K.; Pulay, P. *J. Phys. Chem.* **1996**, *100*, 6310−6316.

(13) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497−5509.

(14) For instance, the G-2 procedure: see ref 31.

(15) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(16) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789. (b) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200−206.

(17) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(18) (a) Burke, K.; Perdew, J. P.; Wang, Y. In *Electronic Density Functional Theory: Recent Progress and New Directions*; Dobson, J. F., Vignale, G., Das, M. P., Eds.; Plenum: New York, 1998. (b) Perdew, J. P. In *Electronic Structure of Solids '91*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11. (c) Perdew, J. D.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671−6687. (d) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolohais, C. *Phys. Rev. B* **1993**, *48*, 4978 (erratum). (e) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533−16539.

(19) (a) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789−807. (b) Ditchfield, R. *J. Am. Chem. Soc.* **1971**, *93*, 5287−5288.

(20) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251−8260.

(21) Wolinski, K.; Sadlej, A. J. *Mol. Phys.* **1980**, *41*, 1419−1430.

(22) McWeeny, R. *Phys. Rev.* **1962**, *126*, 1028−1034.

(23) Keith, T. A.; Bader, R. F. W. *Chem. Phys. Lett.* **1992**, *194*, 1−8.

(24) Keith, T. A.; Bader, R. F. W. *Chem. Phys. Lett.* **1993**, *210*, 223−231.

(25) Keith, T. A.; Bader, R. F. W. *J. Chem. Phys.* **1993**, *99*, 3669−3682.

(26) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(27) Wiberg, K. B.; Lavanish, J. M. *J. Am. Chem. Soc.* **1966**, *88*, 5272−5275.

(28) Sadtler Research Laboratories, Division of Bio-Rad Laboratories, Researchers, Editors, and Publishers: 3316 Spring Garden St., Philadelphia, PA 19104; 1966−present.

(29) Properties of Organic Compounds database, Version 5.0, CRC Press: 1996.

(30) Silverstein, R. M.; Bassler, G. C.; Morrill, T. C. *Spectrometric Identification of Organic Compounds*, 5th ed; John Wiley and Sons: New York, 1991.

(31) (a) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622−5629. (b) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1990**, *93*, 2537−2545. (c) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221−7230. (d) *J. Chem. Phys.* **1998**, *109*, 7764−7776.

(32) (a) de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*, 11485−11488. (b) Gu, Z.; Zambrano, R.; McDermott, A. *J. Am. Chem. Soc.* **1994**, *116*, 6368−6372. (c) Radeglia, R.; Engelhardt, G.; Lippmaa, E.; Pehk, T.; Nolte, K.-D.; Dähne, S. *Org. Magn. Reson.* **1972**, *4*, 571−576. (d) Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed.; VCH Publishers: Weinheim, Germany, 1988; p 315.

(33) Raynes, W. T. *Nucl. Magn. Reson.* **1978**, *7*, 1−25.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94* (Revision C.2); Gaussian, Inc.: Pittsburgh, PA, 1995.

(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98* (Revision A.6); Gaussian, Inc.: Pittsburgh, PA, 1998.