REGULAR ARTICLE

# Quantum chemical prediction of the $^{13}$C NMR shifts in alkyl and chlorocorannulenes: correction of chlorine effects

**Kim K. Baldridge · Jay S. Siegel**

**Abstract** Prediction of the $^{13}$C NMR shifts of *sym*-pentachlorocorannulene and decachlorocorannulene provided impetus for the development of a correction scheme based on a regression of experimental and quantum chemical data. A training set of 15 compounds (18 carbon signals) comprising carbons atoms bearing 1–4 chlorine atoms leads to an estimated error per chlorine atom of about 10–12 ppm. Specifically, linear regression of the data obtained at B3LYP/cc-pVDZ leads to $y = -3.77 + 13.11x$, with $R = 0.982$. Ultimately, experiment and theory converge for *sym*-pentachlorocorannulene and decachlorocorannulene, the former by correction of the theory, the latter by collecting the proper experimental data.

**Keywords** $^{13}$C NMR · Corannulene · Density functional theory · Chlorine

## 1 Introduction

Prediction of the $^{13}$C NMR shift of organic compounds greatly aids the validation of structure elucidation in natural products [1]. The accuracy of the quantum chemical values for C, H, N, O compounds even allows one to identify experimental anomalies with great confidence [2]. As such, one should expect that computational methods appropriate for the prediction of the geometry of corannulene (**1**) would also be able to provide accurate data for the $^{13}$C NMR of **1** and simple alkyl derivatives of corannulene, such as *sym*-pentamethylcorannulene (**1** − Me$_5$) and decamethylcorannulene

(**1**–Me$_{10}$) [3,4]. In contrast, chlorinated compounds present a special problem for such computations as has been previously noted [5–7]. For example, most methods, including higher–order methods, have difficulty in the prediction of $^{13}$C NMR shifts for carbon tetrachloride, the calculated values deviating from experiment by almost 50 ppm [5]. Therefore, the challenge associated with the prediction of the $^{13}$C NMR of *sym*-pentachlorocorannulene (**1** − Cl$_5$) and decachlorocorannulene (**1** − Cl$_{10}$), two sparingly soluble corannulene derivatives [3,4,8–11],[1] offers an opportunity to test wavefunction type and basis set, correlation and relativistic effects, as well as NMR algorithms. In the absence of any truly suitable combination of these methods to produce results comparable to experiment, these results motivate the derivation of an empirical correction on the basis of a regression of experimental vs. quantum chemical data over a training set of chlorohydrocarbons [12,13] (Fig. 1)[2].

There have been many theoretical developments made specifically for the prediction of NMR phenomenon; they range from higher order electronic structure theory methods to specialized density functional theory methods [14–18]. Although the former methods together with large basis sets provide highly accurate results, they can quickly become computational demanding with increasing size and complexity of molecular system. Attempts to alleviate the need to do full QM computations have led to a series of extrapolation algorithms for prediction of $^{13}$C chemical shifts [19–22]. In addition, the development of empirical correction schemes to the quantum mechanical prediction of NMR shifts is well documented [23–27]. These schemes range from statistical

K. K. Baldridge (✉) · J. S. Siegel
Organic Chemistry Institute, University of Zurich,
190 Winterthurerstrasse, 8057, Zurich, Switzerland
e-mail: kimb@oci.unizh.ch

---

[1] The only NMR data reported on **1** − Cl$_{10}$ comes from work claiming to prepare **1** − Cl$_{10}$ by electrical discharge in liquid chloroform see [11].

[2] In general, empirical corrections to QM calculations has historical roots, see for example [12,13].
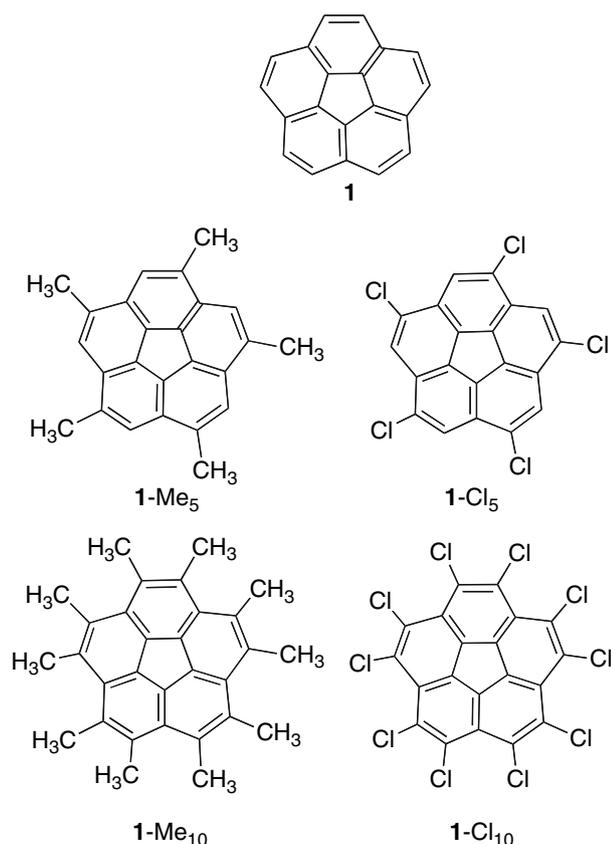
**Fig. 1** Set of penta- and deca-substituted corannulene derivatives for experimental vs theoretical structure and property analysis

enhancement (regression analysis) of data obtained from established quantum methods [23–26], to the development of specific functionals tailored toward increasing the accuracy of NMR predictions [27]. In cases where a systematic deviation of theory from experiment can be demonstrated across a set of reference "training" compounds, these regression schemes allow a simple way to predict NMR shifts accurately for novel compounds outside the training set. This situation often arises when appropriate levels of ab initio and/or density functional theory are prohibitively expensive, or reliable methodology is not available. Although often viewed as quick fixes for the experimentalist, the results obtained with regression analysis tend to compete well with those from the latest tailored functionals [27]. In addition, such corrections can serve as diagnostics to motivate the development of functionals and other methods that obviate the need for correction altogether.

Comparison of computed with experimental values is simplified by reference to a standard, usually tetramethylsilane (TMS) for ^{13}C NMR [14]. Although linear correlations typically exist between experimental and theoretical values predicted in this manner, the slope of the line can deviate more or less from 1, depending on the computational method and

choice of basis set [23,24,28]. Lack of unit slope means that simply subtracting the computed TMS value from the value of the signal in question introduces a variable error depending on the distance away from the reference. As such, linear correction schemes may be necessary if values range over the full 200 ppm spectrum. Another type of correction arises when the method systematically errs in assessing the contribution from specific atom types, for example halogens, specifically chlorine. As chlorohydrocarbons are the central theme here, we propose a further correction on the basis of the observation that the number of chlorine atoms attached to a given carbon atoms accounts for the primary cause of variance between experimental and predicted ^{13}C NMR values [5][3].

## 2 Computational methods

All calculations have been carried out using the GAMESS [29], and Gaussian03 [30] software packages. The computations employ a variety of levels of theory for comparative purposes. Wavefunction-based methods considered include Hartree–Fock, second-order Møller–Plesset perturbation theory (MP2) [31], and hybrid density functional theory (HDFT) [32], including the functionals, B3LYP [33,34], mPW1PW91 [35], B3PW91 [36], B1LYP [35,37], and LG1LYP [38]. Choice of optimal methodology to use was made by performing computations on a training set of compounds, using these wavefunction methods in combination with a variety of basis sets, including 6-3abG($n$d,$m$p) [39–41], dzv(2d,p) [42], cc-pVnZ and aug-cc-pVnZ, where $n = 0$–$2$,$m = 0$–$1$, $a = 1$ or $11$ and $b = $ null,+,++ [43–45]. For each optimized geometry, the Hessian (matrix of second derivatives) was calculated to determine local minima (positive definite) or $n$th-order saddle point (n negative eigenvalues). From these data, correlation lines were generated and evaluated for linear fit and slope.

A level of methodology, which has previously been shown to be highly chemically meaningful for all our studies involving corannulene, was made through extensive evaluation of structure and property variance with increasing sophistication of basis set and the addition of dynamic electron correlation [3,46–52]. The resulting optimal methodology is the MP2/cc-pVDZ//B3LYP/cc-pVDZ level of theory. While the B3LYP/cc-pVDZ provides excellent agreement of structural features of corannulene based on systems optimized with HDFT methods, energetics, barrier heights and several other properties require the MP2 level of theory for satisfactory agreement. In general, computational investigations involving molecules of this type have revealed the necessity for

---

[3]  In the course of this work, we have come to notice a paper that recently appeared having similar views on this subject, see [5].

dynamic correlation and at least double-$\zeta$ valence plus polarization basis sets, with some higher order properties needing more advanced basis sets. In particular, cases where large discrepancies (i.e., greater than 1 ppm) are observed in prediction of NMR shifts, it has been established that dynamical factors play an important role in perturbing the NMR spectrum. Relativistic effects were investigated using the Douglas–Kroll–Hess Hamiltonian correction [53–57].

Calculation of magnetic properties necessitates consideration of the gauge-invariance problem, resulting from the fact that the magnetic field appears in the Hamiltonian in a form of a vector potential whose origin (gauge origin) is not fixed [58–61]. A common gauge origin (CGO) may be chosen, such as the origin of the molecular coordinate system, but computations short of an infinite basis set yield magnetic properties dependent on the choice of the gauge origin. Thus, the truncated basis sets, commonly used in ab initio calculations, can result in significant error due to gauge variance. Distributed origin approaches, which introduce local gauge origins to define the vector potential, include the gauge factors explicitly in either the atomic orbitals or incorporated into the molecular or localized molecular orbitals. Density functionals, while widely used for NMR, may not include magnetic field dependence in some cases, and as such, will not provide systematically improved NMR predictions with increasing basis set, as do HF and MP2 theory. In this study, we employ the class I CGO method [60], class II methods CSGT [62] and IGAIM [63], and the class III method GIAO [64,65]. The latter method is most efficient in terms of convergence of chemical shift values as a function of basis set, and enables the most reliable predictions with relatively modest basis sets. All predictions are calibrated with respect to TMS.

## 3 Results and discussion

Corannulene and functionalized corannulene systems create computational challenges in and of themselves. There are notions in the literature indicating that simple HF level methods enable prediction of the structure of corannulene as well as, or better than, higher order methods [66]. This misconception comes from a limited analysis of the structure restricted to trivial prediction of bond lengths/angles, and/or a limited set of functionalized derivatives; however, previous work involving systematic analysis of the steric properties, including full geometry of the bowl and flat forms, bowl depth, curvature, and activation energy for bowl inversion, reveals a much different picture [47,67,68]. When one considers the physical properties of corannulenes as well, for example electron affinities and ionization potentials, a HF wavefunction with limited basis set functionality does not constitute an appropriate computational level, and can lead one to unreliable conclusions and oversimplified analysis of important phenomenon.

As mentioned in Sect. 2, the B3LYP/cc-pVDZ level of theory for nuclear geometry together with MP2/cc-pVDZ// B3LYP/cc-pVDZ single point energetics was deduced as the optimal level of theory for the prediction of a wide range of corannulene and substituted corannulene structure and properties [47]. Moreover, the B3LYP/cc-pVDZ methodology applied for the prediction of $^{13}$C NMR of **1** also works well compared with several other methods (e.g., Table 1), giving values of 134.5, 131.0, and 127.2 ppm, compared with the experimentally measured shifts of 135.8, 130.9, and 127.0 ppm. Expanding the comparison to two alkyl derivatives of corannulene, $1 - Me_5$ and $1 - Me_{10}$, for which no X-ray crystallographic data are available at present, this method still provides a reassuring prediction of the $^{13}$C NMR shifts with the largest deviation, computation vs. experiment, being 3.3 ppm and the average deviation being 1.2 ppm. Considering only the aromatic carbons the fit is 2.8 and 0.9 ppm, respectively (Table 1).

The same structural methodology applied to the more challenging $1 - Cl_5$ and $1 - Cl_{10}$ systems was first evaluated geometrically through comparison with the gas phase geometry of $1 - Cl_{10}$ [9]. Excellent agreement was found for carbon–carbon bond lengths, angles, and bowl depth (Table 2). Computed carbon-chorine bond lengths showed a characteristic difference from experiment (calcd = 1.743 vs. 1.732 Åexptl). Despite the general good geometrical comparison in $1 - Cl_{10}$ at this level, $^{13}$C NMR of $1 - Cl_5$ predicted at the same level of theory showed the value for the chlorine bearing carbon as a strong outlier (over 10 ppm from experiment); an even larger deviation was predicted for the chlorine bearing carbon in $1 - Cl_{10}$.

At present there is no experimental structural data for $1 - Cl_5$, so the assessment of the performance of the method on the basis of structure was not possible; however, the fact that similar $^{13}$C NMR shift discrepancies between computation and experiment were found for the chlorine bearing carbons in $1 - Cl_5$ and $1 - Cl_{10}$ focused attention on the general performance of computational methods for the prediction of $^{13}$C NMR shifts of chlorine bearing carbons.

The failure of methods to predict chlorocarbon NMR spectra is well known, [5–7,69–71] but a reliable correction scheme has not been presented. As such we selected a training set of 15 compounds (methane, ethylene, and their chlorocongeners as well as benzene, 1,3,5-trichlorobenzene and perchlorobenzene constituted this set; Fig. 2), which gave rise to 18 carbon signals that could be predicted and compared to experiment as a whole and in various structure classes.

The classic set of problems leading to poor predictions of NMR shifts includes: (1) geometrical errors; (2) insufficient basis set completeness; (3) neglect of correlation; rovibrational effects; and (4) relativistic effects. Initial suspicion

**Table 1** Calculated $^{13}$C isotropic shift values for corannulene and substituted corannulenes

| Molecule | Calculated $^{13}$C NMR in ppm[a] | | | | | | | | EXPT[b] (EXPT[c]) |
|---|---|---|---|---|---|---|---|---|---|
| | RHF/6-31+G(d,p)+G(d,p) | RHF/cc-pVDZ | B3LYP/6-31+G(d,p) | B3LYP/6-311 | MP2(full)/6-31+G(2d,p) | B3LYP/cc-pVTZ+G(d,p) | MP2/cc-pVDZ | B3LYP/cc-pVDZ | |
| **1** | 126.96 | 131.12 | 125.02 | 131.9 | 120.55 | 132.15 | 121.11 | 127.2 | 127.0 |
| | 132.43 | 136.55 | 129.43 | 136.1 | 122.23 | 136.65 | 126.00 | 130.9 | 130.9 |
| | 136.80 | 140.94 | 133.28 | 140.4 | 129.38 | 140.55 | 130.44 | 134.5 | 135.8 |
| **1 – Me$_5$** | **** | **** | 22.33 | 20.86 | **** | 21.65 | **** | 21.4 | 19.1 |
| | | | 121.0 | 127.27 | | 128.05 | | 123.4 | 122.8 |
| | | | 129.0 | 136.11 | | 135.95 | | 130.4 | 130.5 |
| | | | 132.8 | 139.54 | | 139.65 | | 133.8 | 134.4 |
| | | | 136.6 | 143.27 | | 143.68 | | 137.9 | 136.1 |
| **1 – Me$_{10}$** | **** | **** | 25.11 | 22.73 | **** | 23.19 | **** | 23.8 | 20.5 |
| | | | 128.68 | 134.60 | | 134.57 | | 129.6 | 129.6 |
| | | | 128.85 | 135.69 | | 135.29 | | 130.0 | 130.6 |
| | | | 134.31 | 140.89 | | 145.92 | | 136.5 | 133.7 |
| **1 – Cl$_5$** | 126.88 | 130.44 | 123.08 | 130.3 | **** | 130.35 | **** | 124.9 | 124.5 |
| | | | 128.43 | 135.5 | | 136.15 | | 130.2 | 130.1 |
| | | | 130.64 | 138.2 | | 138.65 | | 132.2 | 133.3 |
| | | | 141.59 | 148.6 | | 149.05 | | 145.0 | 133.6 |
| **1 – Cl$_{10}$** | 128.89 | 132.48 | 125.17 | 132.04 | **** | 132.26 | **** | 126.5 | 128.2 (126.7) |
| | 128.84 | 133.06 | 125.73 | 132.42 | | 132.61 | | 127.5 | 128.4 (128.9) |
| | 142.37 | 148.00 | 142.32 | 148.92 | | 149.25 | | 145.1 | 129.6 (133.3) |

[a] Geometries all calculated at the B3LYP/cc-pVDZ (see text for rational)

[b] Experimental data from [11]

[c] Experimental data from this work (Siegel lab)

**Table 2** Decachloro-corannulene geometry



*NR* not reported

[a] Experimental study from [9]. We report the values from the abstract and those reported as "preferred" in the text as "Model A"

[b] Deduced from [9], Table 2, as $5* < r(C-C) > - r(\text{hub}) - r(\text{spoke}) - 2*r(\text{flank}) = r(\text{rim})$
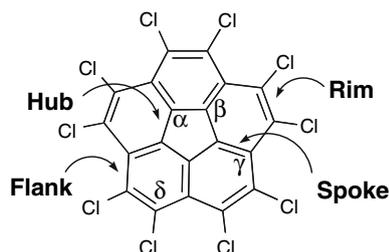
| Parameter | B3LYP/cc-pVDZ | B3LYP/6-311G*[a] | Exptl 2 (abstract)[a] | Exptl 2, Model A[a] |
|---|---|---|---|---|
| CC hub | 1.410 | 1.406 | 1.421(17) | 1.419 |
| CC spoke | 1.380 | 1.375 | 1.383(23) | 1.382 |
| CC flank | 1.460 | 1.457 | 1.472(18) | 1.469 |
| CC rim | 1.415 | 1.409 | 1.410(27) | 1.405[b] |
| CCl | 1.743 | 1.741 | 1.732(5) | 1.729 |
| Angle $\alpha$ (sym) | 108 | 108 | 108 | 108 |
| Angle $\beta$ | 124.97 | 124.9 | 124.3 | 124.3 |
| Angle $\gamma$ | 112.0 | 112.1 | 111.9 | 111.9 |
| Angle $\delta$ | 122.67 | 122.6 | 122.2 | 122.2 |
| Bowl depth | 0.485 | NR | NR | NR |

that the discrepancy between calculated and experimental predictions of $^{13}$C NMR shifts ($\delta$) arises due to the error in the prediction of the C–Cl bond length was not defensible. A basis set analysis (18 basis sets: cc–pVnZ, $n = 2 - 3$ with and without augmentation; 6–31 and 6–311 splits with up to 2df, pd and diffuse functionality; and a variety of correlation methods, including HDFT, MP2, and coupled cluster) of C–Cl bond length prediction in CCl$_4$ reveals at most 0.028 Å the difference in bond length from that of the experimental value, 1.767 Å, with average deviation across the entire set of 0.015 Å; no special correlation of structural accuracy and predicted NMR value was seen, and in particular, when the experimental geometry is used for prediction of chemical shift at several method/basis set combinations, similar discrepancy is noted.[4] A more general basis set study for the entire training set exceeds the focus of the present work, but has now become the subject for a future paper.

In parallel to testing for geometry, basis set convergence could be probed as well as various methods for including electron correlation. These too were unable to resolve the discrepancy. Attempts to estimate the effect of rovibrational and/or spin–orbit effects suggested these effects not to be



**Fig. 2** Training set (**2a-o**) for computational analysis of chlorine-containing systems

greater than about 1 ppm. Equally, we find that correction for relativistic effects did not seem to play a key role in agreement with other literature [6], although this possibility has been suggested for small saturated chlorohydrocarbons [5].

An alternative approach to arriving at reliable $^{13}$C NMR chemical shift predictions is to derive an empirical correc-

---

[4] Note that experimental C–Cl value for CCl$_4$ is reported to be 1.767; B3LYP/cc-pVDZ predicts 1.7947. Good agreement was provided by MP2(Full)/6-31+$G$(d,p), and therefore this method was also included in the analysis. Of all basis set/wavefunction combinations analyzed, including coupled cluster and density functional theory with extensive basis set functionality, bond length predictions of 1.77 and higher were found.

**Table 3** B3LYP/cc-pVDZ calculated $^{13}$C NMR compared to experiment

| Molecule | $^{13}$C$\delta$, ppm | | | | | | |
|---|---|---|---|---|---|---|---|
| | B3LYP/cc-pVDZ | Exptl[a] | Exptl[b] | $\Delta$ (Expt–Theory)(ppm) | | $\Delta$ (Expt–Theory) (ppm) including C–Cl correction | |
| | | | | Exptl[a] | Exptl[b] | Exptl[a] | Exptl[b] |
| **1** | 127.2 | 127.0 | – | −0.2 | – | −0.2 | – |
| | 130.9 | 130.9 | | 0.0 | | 0.0 | |
| | 134.5 | 135.8 | | 1.3 | | 1.3 | |
| **1 − Me$_5$** | 21.4 | 19.1 | – | −2.3 | – | −2.3 | – |
| | 123.4 | 122.8 | | −0.6 | | −0.6 | |
| | 130.4 | 130.5 | | 0.1 | | 0.1 | |
| | 133.8 | 134.4 | | 0.6 | | 0.6 | |
| | 137.9 | 136.1 | | −1.8 | | −1.8 | |
| **1 − Me$_{10}$** | 23.8 | 20.5 | – | −3.3 | – | −3.3 | – |
| | 129.6 | 129.6 | | 0.0 | | 0.0 | |
| | 130.0 | 130.6 | | 0.6 | | 0.6 | |
| | 136.5 | 133.7 | | −2.8 | | −2.8 | |
| **1 − Cl$_5$** | 124.9 | 124.7 | – | 0.02 | – | 0.02 | – |
| | 130.2 | 130.2 | | 0.0 | | 0.0 | |
| | 132.2 | 133.4 | | 1.2 | | 1.2 | |
| | 145.0 (135.7)[c] | 133.8 | | −11.2 | | −1.9 | |
| **1 − Cl$_{10}$** | 126.5 | 126.7 | 128.2 | 0.2 | 1.7 | 0.2 | 1.7 |
| | 127.5 | 128.9 | 128.4 | 1.4 | 0.9 | 1.8 | 0.9 |
| | 145.1 (135.7)[c] | 133.3 | 129.6 | −11.8 | −15.5 | −2.4 | −6.1 |

[a] Experimental data from this work (Siegel group)
[b] Experimental data from [11]
[c] Corrected C–Cl signal

tion, as discussed above. A series of wavefunction/basis set combinations, functionals (e.g., see Sect. 2) was analyzed to establish correlation trends, including full optimizations of geometries as well as single points at the B3LYP/cc-pVDZ geometry. The B3LYP/cc-pVDZ method, which provides excellent geometric agreement in the corannulene series well below the 0.015 Å limits, was observed to provide optimal results across the set of methods tested. Therefore, we focus on this method for development of the empirical correlation correction scheme.

One testable hypothesis is that the error in $^{13}$C NMR shifts correlates solely with the number of chlorines attached to the carbon, which would then result in the ability to establish a simple correction such as a computational equivalent of a linear correction to prediction made by Shoolery's rules [72]. Alternatively, the effect of chlorine on carbon shifts could be dependent on the hybridization at carbon and/or some nonlinear relation to the number of chlorines a carbon might bear and other subtle structural details. A quick glance at the data for chlorinated methane points to the importance of the chlorine count.

The set of chlorinated methanes proved to be challenging for theoretical predictions due to the inclusion the Cl/C = 3

and 4 species, and therefore a much larger subset of method/basis set combinations was analyzed for this set. Subsequently, a subset of these methods was used for the entire training set. Discrepancy between experiment and computation is the key point of our analysis, and therefore a plot of (experimental-calculated) $^{13}$C shifts vs number of chlorines born by a single carbon atom (C[Cl$_n$]) for the methane series reveals a rough linear fit ($R = 0.98$). Although a quadratic fit is almost perfect ($R = 0.999$) the simplicity of the linear form is preferred. The error of about 10 ppm per chlorine atom is consistent with the above-mentioned error of nearly 50 ppm for the shift prediction in carbon tetrachloride. This is the case whether one conducts the NMR computation including full geometry optimization at the respective level (Fig. 3b), or at the B3LYP/cc-pVDZ geometry (Fig. 3a). Additionally, if one considers evaluation of the chemical shift at the experimentally known geometry, or, using alternative NMR methods (e.g., IGAIM shown in Fig. 1b), the trend still remains with slight variation. MP2/cc-pVDZ lowers the error over that of B3LYP/cc-pVDZ, but as the trend is the same and a residual error correction is still needed, the extra computational cost associated with MP2 is not warranted for large structures.
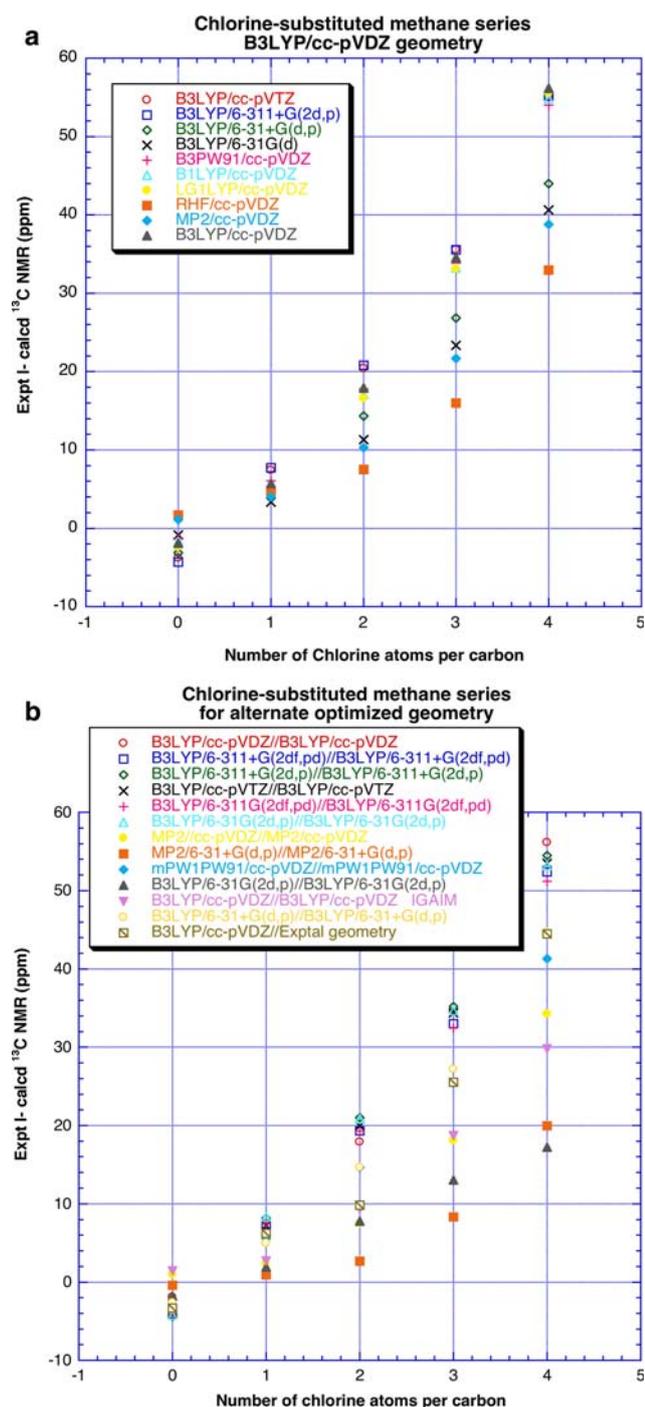
**Fig. 3** Plots of $^{13}$C NMR of methanes vs. number of chlorines per carbon



**Fig. 4** Plots of the error between experimental and computed $^{13}$C NMR shifts for all molecules in the training set vs. number of chlorines per carbon. *Left* the effect of basis set and correlation; *right* the best level showing the residual systematic error

Regression on a scatter plot of (experimental-calculated) $^{13}$C shifts as a function of number of chlorines per carbon atom over the entire training set yields similar trends to that seen for the methyl compounds (Fig. 4). Use of simple RHF treatment does a very poor job ($R = 0.877$) as does adding MP2 without increasing basis set level ($R = 0.7795$). The best systematic accounting for the error is obtained at B3LYP/cc-pVDZ ($R = 0.982$), a density functional method

for including correlation with a correlation consistent double-zeta basis set. At this level, linear correlation results in a correction equation of $y = -3.77 + 13.1x$, where $x$ is the number of chlorine atoms attached to a carbon and y is the $^{13}$C correction factor in ppm. For a single chlorine atom attached to carbon the correction factor would be about 9.3 ppm by this method.

An immediate word of caution is appropriate here. This scheme indicates a failing in the computational NMR method,

and within the confinements of time and resources it demonstrates how to improve the predictive quality of a specific set of data. Such an empirical correction is not intended to represent a general treatment or to suggest that one can simply replace proper quantum chemical methods by quick fixes. Further methodological refinements to account for chlorine and other heavy atom effects are still the ultimate objectives. Nonetheless, the effectiveness of this simple assumption is impressive. A direct comparison of experimental vs. predicted values without chlorine-scaled correction gives a fit with $R^2 = 0.866$, relative poor correlation. In contrast, after applying the chlorine-scaled correction, a similar plot yield a fit with $R^2 = 0.995$ (Fig. 5).

Consideration of the entire training set in classes defined by $C[Cl_n]$, shows that a much smaller data variance exists due to particular structural factors within a given class. Indeed for $C[Cl_n]$ values of 0, 1 and 2, where at least four $^{13}C$ NMR shift values each were predicted, the mean variance is never higher than about 3 ppm, a factor of 3–4 smaller than the effect due to chlorine count. For the purposes of this work, further correction due to hybridization was not sought. Equally small, is the variance due to comparison of the errors for *gem*, *cis*, and *trans* dichloroethylene, which do not differ substantially from those of other members of the training set. Thus, as far as the computational error is regarded, this study reveals at most a negligible additional error due to secondary geometric effects such as these.

One way to smooth out these non-specific errors would be to average all of the errors within a given class, say for $n = 1$, the class of interest. In this case, over nine carbon signals the average deviation is about 8 ppm, close to 9.34 ppm predicted by the linear regression analysis (7.728, via quadratic). This method is presented as just another way of showing consistency within the data, and all methods show that a chlorine attached to carbon leads to a computational error in the $^{13}C$ NMR shift of about 8–10 ppm. The diagnostic is clear albeit that the underlying causes are not, and they exceed the scope of this particular paper. Within the limited focus of predicting $^{13}C$ NMR of chlorinated corannulenes, this correction is helpful.

Returning to our objective of accurate prediction of the $^{13}C$ NMR shifts for $\mathbf{1} - Cl_5$ and $\mathbf{1} - Cl_{10}$, one finds a substantial improvement in the prediction of the $^{13}C$ NMR shift values for $\mathbf{1} - Cl_5$ when the linear regression correction equation is applied (Table 3). The previous outlier in the pentachlorocorannulene structure now falls within 2 ppm of the experimental observed value with mean overall deviation is less than 1 ppm. However, in the case of decachlorocorannulene, the previous outlier still falls 6.1 ppm away from the experimental value reported by Huang et al. on a sample of $\mathbf{1} - Cl_{10}$, claimed to be prepared by passing an electrical arc through chloroform [11].
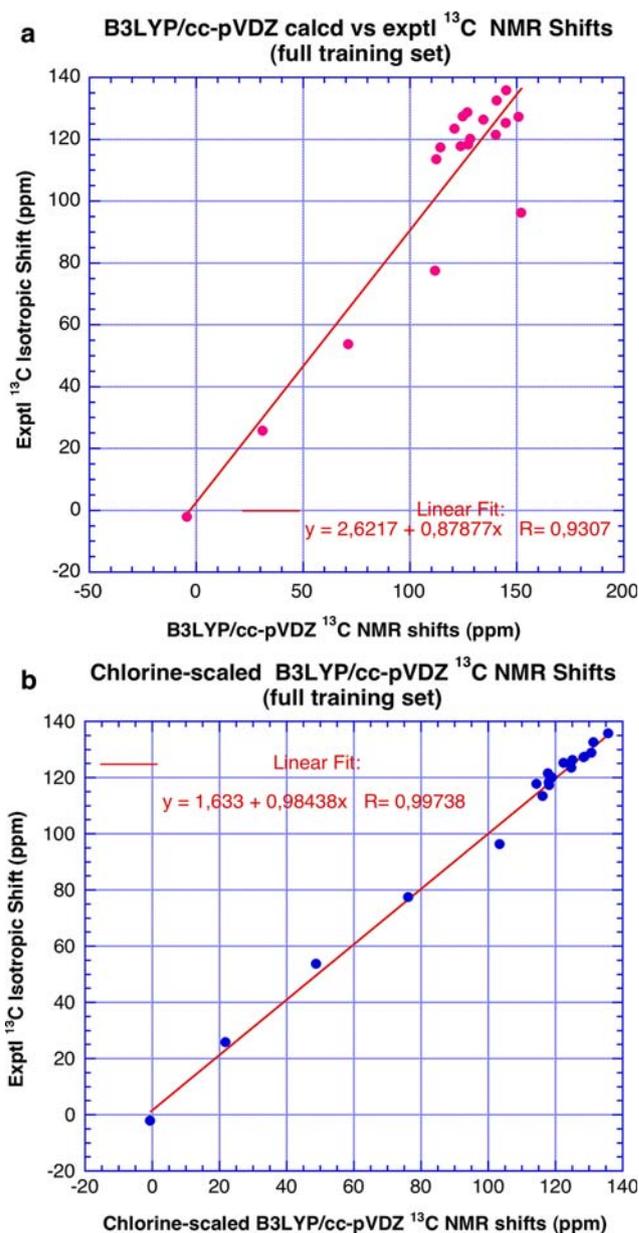


**Fig. 5** Calculated vs. experimental $^{13}C$ shifts for the full training set ($R^2 = 0.8662$ vs. 0.9948, respectively)

At the outset of this study our lab had yet to measure the $^{13}C$ NMR of $\mathbf{1} - Cl_{10}$ due to its limited solubility; however, the residual discrepancy between the predicted $^{13}C$ NMR values for $\mathbf{1} - Cl_{10}$ and the literature spectrum reported [11] lead us to consider whether the residual error lie with the experiment. Since then, we have been able to measure the $^{13}C$ NMR of $\mathbf{1} - Cl_{10}$ produced by the same method used to prepare the sample for which electron diffraction was accomplished; thus we feel confident in the structural assignment of our material. Collection of $^{13}C$ NMR data on a warm saturated solution of $\mathbf{1} - Cl_{10}$ in tetrachloroethylene yielded aromatic $^{13}C$ NMR shifts at 133.3, 128.9, and 126.7. These

**Table 4** ^{13}C NMR signal dependence
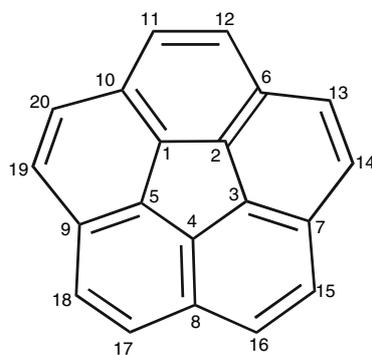


| Molecule (bowl depth) | Electrostatic map[a] | Carbon reference | B3LYP/cc-pVDZ ^{13}C NMR | | Calc-Exptl **1**[b] |
|---|---|---|---|---|---|
| | | | Bowl | Flat | |
| **1**(0.873) | | 11–20 | 127.2 | 128.7 | −0.2 |
| | | 6–10 | 130.9 | 127.6 | 0.0 |
| | | 1–5 | 134.5 | 131.6 | 1.3 |
| **1** − Cl₅ (0.843) | | 12,14,16,18,20 (C–H) | 124.9 | – | 0.02 |
| | | 6–10 | 130.2 | – | 0.0 |
| | | 1–5 | 132.2 | – | 1.2 |
| | | 11,13,15,17,19 (C–Cl) | 145.0 (135.7)[c] | – | 1.9 |
| **1** − Me₅(0.8498) | | Methyl groups | 21.43 | – | −2.3 |
| | | 12,14,16,18,20(C–H) | 123.4 | – | -0.6 |
| | | 6–10 | 130.4 | – | 0.1 |
| | | 1–5 | 133.8 | – | 0.6 |
| | | 11,13,15,17,19(C–CH₃) | 137.9 | – | −1.8 |
| **1** − Cl₁₀ (0.485) | | 1–5 | 126.5 | 126.3 | 0.2 |
| | | 6–10 | 127.5 | 127.6 | 1.4 |
| | | 11–20 (C–Cl) | 145.1 (135.7)[c] | 146.0 | −2.4 |

**Table 4** continued

| Molecule (bowl depth) | Electrostatic map[a] | Carbon reference | B3LYP/cc-pVDZ $^{13}$C NMR | | Calc-Exptl **1**[b] |
|---|---|---|---|---|---|
| | | | Bowl | Flat | |
| **1** $-$ Me$_{10}$ (0.552) |  | Methyl groups | 23.8 | 22.8 | −3.3 |
| | | 1–5 | 129.6 | 128.9 | 0.0 |
| | | 6–10 | 130.0 | 129.4 | 0.6 |
| | | 11–20 | 136.5 | 136.97 | −2.8 |
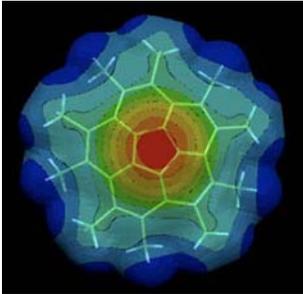
[a] Electrostatic surfaces (red negative blue positive) for the convex face
[b] This work, using the chlorine corrected value
[c] Signal corrected for chlorine according to regression analysis

values compare well with the 135.7, 127.5, and 126.5 predicted by computation. The largest difference is 2.4, and the mean difference is 1.3 ppm (Table 3). [Details of this experiment will appear in a technical report in collaboration with Bruker, Fallenden.]

Once one has a reliable set of NMR shift data in hand and the accompanying high-level ab initio computations, one can probe questions of substituent effects, the premise of this paper being how pervasive linear substituent effects have been in empirical NMR spectral analysis. Analysis of the change in chemical shift from **1** to **1**$-$Me$_5$ to **1**$-$Me$_{10}$ of hub and spoke carbons reveals values 134.53, 133.85, 129.65 ppm and 130.96, 130.42, 130.01, respectively—a poor line even for three points! The situation is similar for **1** to **1** $-$ Cl$_5$ to **1**$-$Cl$_{10}$; 134.53, 132.16, 126.46 and 130.96, 130.21, 127.47, respectively. What could cause such a deviation from linearity? Here the answer is perhaps obvious. Whereas the penta-derivatives are more or less similar in bowl-geometry to the parent (**1** to **1** $-$ Me$_5$ to **1** $-$ Cl$_5$ $-$ 0.873, 0.850, 0.843 Å, respectively) the deca-derivatives are much flatter (**1** $-$ Me$_{10}$ to **1** $-$ Cl$_{10}$ $-$ 0.552, 0.485 Å, respectively). This flattening of the bowl results in an additional upfield shift of the carbon signals, as evidenced by the NMR computation on the flat corannulene structure. Although the deca-substituents may still seem far from flat, previous work has shown that the barrier to inversion in such compounds in extremely low [47], and as such, time averaged over conformational space, the planer form will contribute significantly [73–75].

Another notion that often arises in the interpretation of NMR shifts is that the shift is related to the electron density at the carbon atom. Although Norton commented on the generality of this issue already some 20 years ago [76], it is often expected to hold for compounds with relatively similar "constitutions". Again, having the ab initio data and the NMR values together allows for a direct comparison (Table 4). The electrostatic maps in Table 4 predict electron

density at the hub carbons to follow the order **1** to **1** $-$ Cl$_5$ to **1** $-$ Me$_5$, **1** to **1** $-$ Cl$_{10}$ to **1** $-$ Me$_{10}$, which does not correlate even monotonically with the hub carbon shifts. The bowl depth is the culprit in this case also. The flatter bowls show an additional upfield shift that swamps out any other correlation. The electrostatic computations imply however that **1** $-$ Cl$_{10}$ is more electron rich at the hub than **1** $-$ Me$_5$, and should show interesting metal binding properties as a result [77].

## 4 Conclusions

Despite the great success in $^{13}$C NMR chemical shift prediction for hydrocarbon compounds by quantum chemical and density functional methods, the shift values for carbons bearing chlorine atoms are poorly predicted directly from quantum mechanical data. The error in such predictions for a training set of 15 compounds (18 carbon shifts) is shown to reside primarily with a linear effect dependent on the number of chlorine atoms born by the carbon, with a much smaller variance due to secondary geometry factors. A simple correction function can be derived by a linear regression of $\delta$ vs. C[Cl$_n$] over the training set. This function applied to the test compounds **1** $-$ Cl$_5$ and **1** $-$ Cl$_{10}$ lead to accurate predictions in the case of **1** $-$ Cl$_5$ and highlighted the need for more experimental investigation of $^{13}$C NMR of **1** $-$ Cl$_{10}$ prepared by conventional methods. We note that this effect appears to be a phenomenon related to halogens in general, from additional computations done with fluorine and bromine. Indeed, when a certified method is used to synthesize **1** $-$ Cl$_{10}$, satisfactory agreement between experiment and theory is obtained. At a minimum we would conclude that the previously reported literature spectrum of **1** $-$ Cl$_{10}$, was misassigned. Identification of such errors in experimental data would not have been possible using simple HF computations. Equally, most prop-

erties of organic molecules are poorly predicted at simple HF levels. As such, claims that lower level HF computations do a better job of modeling corannulene derivatives are misleading and should be neglected.

# References

1. Cimino P, Gomez-Paloma L, Duca D, Riccio R, Bifulco G (2004) Magn Reson Chem 42:526
2. Rychnovsky SD (2006) Org Let 8:2895
3. Seiders TJ, Baldridge KK, Siegel JS (1999) J Am Chem Soc 121:7439
4. Seiders TJ, Elliott EL, Grube GH, Siegel JS (1999) J Am Chem Soc 121:7804
5. d'Antuono R, Botek E, Champagne B, Spassova M, Denkova P (2006) J Chem Phys 125:144309
6. Gryff-Keller A, Molchanov S (2004) Mol Phys 102:1903
7. Kaupp M, Malkina O, Malkin V (1997) Chem Phys Lett 265:55
8. Cheng P-C (1996) Corannulene: Novel synthesis and reactions (Fullerenes). Dissertation, Boston College p 254
9. Samdal S, Hedberg L, Hedberg K, Richardson AD, Bancu M, Scott LT (2003) J Phys Chem 107:411
10. Bancu M, Rai AK, Cheng P-C, Gilardi RD, Scott LT (2004) Synlett 0173
11. Huang R, Huang W, Wang Y, Tang Z, Zheng L (1997) J Am Chem Soc 119:5954
12. Gordon MS, Merrill GN (1999) J Chem Phys 110:6154
13. Gordon MS, Truhlar DG (1986) J Am Chem Soc 108:5412
14. Kaupp M, Bühl M, Malkin VG (2004) Wiley-VCH, Weinheim, p 603
15. Bühl M, Kaupp M, Malkina OL, Malkin VG (1999) J Comp Chem 20:91
16. Wilson PJ (2003) Ann Rep NMR Spect 49:117
17. Gauss J, Stanton JF (2004) ; In: Kaupp M, Bühl M, Malkin V (eds) Calculation of NMR and EPR Parameters. Wiley-VCH, Weinheim p 123
18. Gauss J, Stanton JF (2002) Adv chem Phys 123:355
19. Meiler J, Meusinger R, Will M (1999) Monatsh Chem 130:1089
20. Neudert R, Penk M (1996) J Chem Inf Comput Sci 36:244
21. Pretsch E, Fürst A, Badertscher M, Bürgin R (1992) J Chem Inf Comput Sci 32:291
22. Meiler J, Sanli E, Junker J, Meusinger R, Lindel T, Will M, Maier W, Köck M (2002) J Chem Inf Comput Sci 42:241
23. Forsyth DA, Sebag AB (1997) J Am Chem Soc 119:9483
24. Baldridge KK, Siegel JS (1999) J Phys Chem 103:4038
25. Giesen DJ, Sumbulyadis N (2002) Phys Chem Chem Phys 4:5498
26. Weston J, Ahlbrecht H (1997) J Chem Soc Perkin Trans 2:1003
27. Wiitala KW, Hoye TR, Cramer CJ (2006) J Chem Theory Comput 2:1085
28. Laws DD, Le H, De Dios AC, Havlin RH, Oldfield E (1995) J Am Chem Soc 117:9542
29. Schmidt MW, Baldridge KK, Boatz JA, Jensen JH, Koseki S, Gordon MS, Nguyen KA, Windus TL, Albert ST (1990) Quantum Chemistry Program Exch Bull 10:10
30. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomergy JA, Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishido M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin J, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Doannernberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, KFarkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DG, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challocombe M, Gill PMW, Johnson BG, chen W, Wong MW, Gonzalez C, Pople JA (2004). Gaussian Inc., Wallingford
31. Møller C, Plesset MS (1934) Phys Rev 46:618
32. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, New York
33. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785
34. Miehlich B, Savin A, Stoll H, Preuss H (1989) Chem. Phys. Lett. 157:200
35. Adamo C, Barone V (1998) J Chem Phys 108:664
36. Perdew JP, Wang Y (1992) Phys Rev B 45:13244
37. Becke AD (1996) J Chem Phys 104:1040
38. Adamo C, Barone V (1998) J Comput Chem 19:418
39. Ditchfield R, Hehre WJ, Pople JA (1971) J Chem Phys 54:724
40. Hehre WJ, Lathan WA (1972) J Chem Phys 56:5255
41. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) J Chem Phys 72:650
42. Dunning TH (1970) Chem Phys Lett 7:423
43. Dunning TH (1989) J Chem Phys 90:1007
44. Kendall RA, Dunning TH, Harrison RJ (1992) J Chem Phys 96:6796
45. Woon DE, Dunning TH (1993) J Chem Phys 98:1358
46. Seiders TJ, Baldridge KK, Gleiter R, Siegel JS (2000) Tetrahedron Lett 41:4519
47. Seiders TJ, Baldridge KK, Grube GH, Siegel JS (2001) J Am Chem Soc 123:517
48. Seiders TJ, Baldridge KK, O'Connor JM, Siegel JS (1997) J Am Chem Soc 119:4781
49. Seiders TJ, Baldridge KK, O'Connor JM, Siegel JS (2004) Chem Commun 8:950
50. Seiders TJ, Baldridge KK, Siegel JS (1996) Journal Of the American Chemical Society 118:2754
51. Seiders TJ, Baldridge KK, Siegel JS, Gleiter R (2000) Tetrahedron Lett 41:4519
52. Seiders TJ, Elliot E, Baldridge KK, Siegel JS (1999) Abstr Pap Am Chem Soc 217:U166
53. Douglas M, Kroll NM (1974) Ann Phys 82:89
54. Hess BA (1985) Phys Rev A 32:756
55. Hess BA (1986) Phys Rev A 33:3742
56. Wolf A, Reiher M, Hess BA (2002) J Chem Phys 117:9215
57. Melo JI, Ruiz de Azua MC, Peralta JE, Scuseria GE (2005) J Chem Phys 123:204112
58. Chesnut DB (1997) Chem Phys 314:73
59. Chestnut DB (1989) In: Webb GA (ed) Annual reports on NMR Spectroscopy, vol 21. Academic Press, San Diego
60. Chestnut DB (1994) In: Webb GA (ed) Annual reports on NMR Spectroscopy, vol 29. Academic Press, San Diego
61. Chestnut DB (1997) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, vol 8. VCH New York
62. Keith TA, Bader RFW (1993) Chem Phys Lett 210:223

63. Keith TA, Bader RFW (1992) Chem Phy Lett 194:1
64. Ditchfield R (1972) J Chem Phys 56:5688
65. Wolinski K, Hinton JF, Pulay P (1990) J Am Chem Soc 112:8521
66. Petrukhina MA, Andreini KW, Mack J, Scott LT (2005) J Org Chem 70:5713
67. Biedermann PU, Pogodin S, Agranat I (1999) J Org Chem 64:3655
68. Potier Y, Siegel JS, Baldridge KK in preparation
69. Costa VEU, Alifantes J, Axt M, Mollmann MES, Seidl PR (1999) J Braz Chem Soc 10:341
70. Glaser R, Chen N, Wu H, Knotts N, Kaupp M (2004) J Am Chem Soc 126:4412
71. Eloranta J, Hu J, Suontamo R, Kolehmalnen E, Knuutinen J (2000) Magn Res Chem 38:987
72. Shoolery JN (1995) Prog Nuc Magn Spectr 28:37
73. Stahl M, Schopfer U, Frenking G, Hoffmann RW (1996) J Org Chem 61:8083
74. Stahl M, Schopfer U, Frenking G, Hoffman RW (1997) J Org Chem 62:8083
75. Stahl M, Schopfer U, Frenking G, Hoffman RW (1997) Mol Phys 92:569
76. Evans J, Norton JR (1974) Inorg Chem 13:3043
77. Siegel JS, Baldridge KB, Dorta R (2006) J Am Chem Soc 128:10644