JOC Article

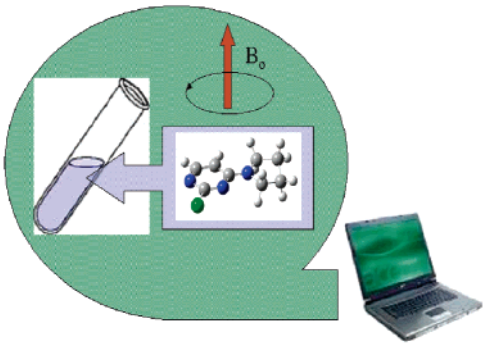
Accuracy vs Time Dilemma on the Prediction of NMR Chemical Shifts: A Case Study (Chloropyrimidines)

Manuel Pérez,[§] Torren M. Peakman,[§] Alexander Alex,[§] Paul D. Higginson,[§] John C. Mitchell,[‡] Martin J. Snowden,[‡] and Iñaki Morao*,[§,‡]

Pfizer Ltd., Global Research and Development, Sandwich Laboratories, Sandwich, Kent, CT13 9NJ, United Kingdom, and University of Greenwich, Medway Sciences, Chatham Maritime, ME4 4TB, United Kingdom

inaki.morao@pfizer.com

Received January 4, 2006



The nuclear magnetic shieldings of two chloropyrimidine species have been predicted and analyzed by means of ab initio and DFT methods. The results have been compared with the experimental values and with those from other database-related approaches. These dataset-based techniques are found to be particularly valuable because of the accurate and instantaneous prediction of the ^{13}C chemical shifts. On the other hand, only a few quantum chemistry based approaches were showed to be the most precise to predict ^{1}H chemical shifts and to elucidate unequivocally the ^{1}H NMR spectra of the regioisomeric mixture under study. Special emphasis was put on incorporating the solvent effect, implicitly, or explicitly. The influence of the level of theory and basis set in the predicted values has also been discussed.

Introduction

NMR spectroscopy is one of the most popular and powerful analytical tools for the structural characterization of molecules in solution.[1] Consequently, different techniques for the accurate and relatively quick prediction of chemical shifts and coupling constants have been developed over the last few decades.[2] Nowadays, there are two main and opposite approaches to achieve such a goal. The fastest and most used solution in industry is the dataset-based approximation. This method includes several algorithms to split the molecule of interest into unique fragments and to compare them with the internal database, with the predicted chemical shift finally computed by incremental rules. For instance, the popular ACD/Labs software package contains a database of ca. 1.5 millions experimental ^{1}H chemical shifts.[3,4] The accuracy of this approach is in general good but depends on the similarity between the molecule under study and the structurally related compounds contained in the dataset. Other limitations are that only the 2D information of the molecule is processed and therefore no stereochemistry is examined. On the other hand, ab initio methods based on coupled perturbation Hartree−Fock calculate the magnetic properties with no need of experimental information.[5] These approaches have been intensively developed in the

[§] Pfizer, Ltd.

[‡] Medway Sciences.

(1) *NMR and Chemistry*: *An introduction to modern NMR Spectroscopy*; Akitt, J. W., Mann, B. E., Eds.; CRC Press: London, UK, 2000.

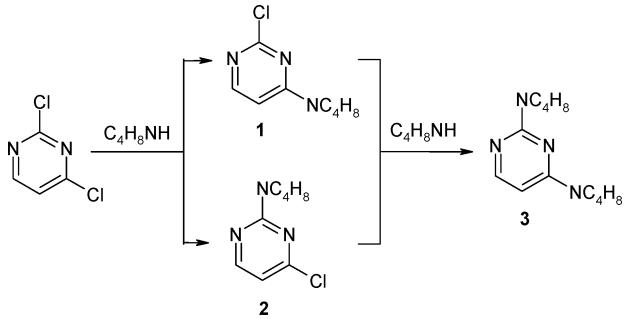
(2) For recent reviews see: (a) Williams, A. *Curr. Opin. Drug Discuss. Dev.* **2000**, *3*, 298−305. (b) Heine, T.; Corminboeuf, C.; Seifert, G. *Chem. Rev.* **2005**, *105*, 3889−3910.

(3) *ACD/Labs*, version 8.0; Advanced Chemistry Development, Inc.: Toronto ON, Canada, 2004; www.acdlabs.com.

(4) Users can add their own data into the commercially supplied database. For instance, in our database most of Pfizer and intermediate compounds are included. However, in this paper and in order to reproduce the results in other laboratories we explicitly excluded the Pfizer database.

10.1021/jo0600149 CCC: $33.50
Published on Web 03/23/2006

SCHEME 1. Synthesis of Compounds 1, 2, and 3



past decades and the main challenges (gauge invariance, computer time, large systems, solvation) have productively been overcome.[6] Additionally and somewhere in the middle between these two extreme positions, a semiempirical methodology has successfully been developed.[7] This approach relies initially on a dataset of compounds and rationalizes the influence of each functional group over a given atom by using equations of different properties (e.g., anisotropy, ring current). The unknowns in the generated equations are parametrized and repeatedly calculated by an iterative process until a good agreement between calculated and experimental values is obtained.

Pyrrolidinylpyrimidine analogues **3** (Scheme 1) were synthesized recently to provide a probe for the metabolite identification of Tirilazad.[8,9] The corresponding intermediates **1** and **2** were also fully characterized by NMR experiments. These regioisomers present very similar ^{13}C and ^{1}H chemical shifts even though the chemical surrounding of these atoms is relatively different. From a computational standpoint this mixture is also an ideal medium size and without conformational issues to carry out a comprehensive comparison along the aforementioned approaches. Thus, in this paper we present ab initio and DFT predictions of ^{13}C and ^{1}H NMR chemical shifts for these molecules (**1**, **2**) in both gas phase and solution (dimethyl sulfoxide or chloroform). These quantum methodologies together with well-known database related packages have been compared with the experimental results. This study clearly shows the need of an intelligent use of both approaches for routine comparisons with the experiments and also the importance of solvent effects to achieve more accuracy in the ^{1}H NMR predictions.

Computational Methods

Most of the quantum mechanical results shown in this work have been obtained by means of the GAUSSIAN03 series of programs,[10] using a standard PC desktop Pentium III (1GB of RAM). For practical reasons we were mainly interested in standard and relatively rapid calculations without the need for supercomputers. Geometric optimizations have been carried out at the Hartree−Fock (HF) level of theory. Given that the systems under study are aromatic the electron correlation is expected to be critical and post Hartree−Fock and density functional theory (DFT) optimizations were also performed. In particular, second-order Møller−Plesset (MP2) keeping the core electrons frozen and three DFT variants were chosen. Unless otherwise noted the hybrid functional B3LYP was always used.[11] For comparison purposes between DFT methods, mPW1PW91[12] and pure BLYP[13] were also analyzed in a few representative cases (named DFT2 and DFT3, respectively, in this paper). To determine the importance of the basis set in the NMR chemical shifts three different basis sets were used: the standard 6-31G(d,p), the medium 6-311+G(d,p), and the large and expensive 6-311++G(3df,2dp), which in this paper are denoted as bs1, bs2, and bs3, respectively.[14] Atomic charges were calculated by using the NBO natural bonding analysis.[15] Nuclear magnetic shieldings and nucleus-independent chemical shifts (NICS) were computed by gauge including atomic orbital (GIAO).[16] Nonspecific solvent effects in both geometry optimizations and NMR calculations were partially taken into account by means of the polarizable continuum model (PCM).[17] Of these methods, the integral equation formalism (IEF) approach was selected.[18] The two solvents used in the experiments were dimethyl sulfoxide and chloroform (labeled as s1 and s2, respectively). For instance, DFT(bs1)s2//DFT(bs1)s2 stands for the GIAO evaluation at the B3LYP/6-31G(d,p) level including the chloroform solvent by means of IEF-PCM methodology over a geometry optimized at the B3LYP/6-31G(d,p) level in chloroform solution by the IEF-PCM method. Solvent effects were also investigated considering the discrete solvation model ("supermolecule" or "cluster" approach) by including explicit solvent molecules around the solute. The solute was placed in the center of an equilibrated cube (30 Å^3) of (dimethyl sulfoxide or chloroform) solvent molecules. The MM optimization was performed with the OPLS force field.[19] After initial minimization of the whole system, MD simulations were also carried out by heating the system to 300 K, using standard parameters (1.5 fs time step, 1.0 ps equilibration time, 10 ps simulation time, the PRCG method within

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

(6) (a) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251−8260. (b) Wolinski, K.; Haacke, R.; Hinton, J. F.; Pulay, P. *J. Comput. Chem.* **1997**, *18*, 816−825. (c) Ochsenfeld, C.; Kussmann, J.; Koziol, F. *Angew. Chem., Int. Ed.* **2004**, *43*, 4485−4489. (d) Cammi, R. *J. Chem. Phys.* **1998**, *109*, 3185−3196.

(7) (a) Abraham, R. J. *Prog. Nucl. Reson. Spectrosc*. **1999**, *35*, 85−152. (b) Hunter, C. A.; Packer, M. J.; Zonta, C. *Prog. Nucl. Reson. Spectrosc*. **2005**, *47*, 27−39.

(8) Tirilazad is a nonglucocorticoid, 21-aminosteroid that was developed for the acute treatment of ischemic stroke and subarachnoid hemorrhage: Bath, P. M. W. et al. *Stroke* **2000**, *32*, 2257−2265.

(9) Streeper, R. T.; Pearson, P. G.; Zhao, Z.; Mizsak, S. A.; Vrbanac, J. J. *J. Labelled Compd. Radiopharm.* **1998**, *XLI*, 577−584.

(10) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(11) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (b) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(12) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675.

(13) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

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

(15) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899−926.

(16) Ditchfield, R. *Mol. Phys*. **1974**, *27*, 789−807.

(17) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.

(18) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.

(19) Jorgensen, J. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc*. **1996**, *118*, 11225−11236.

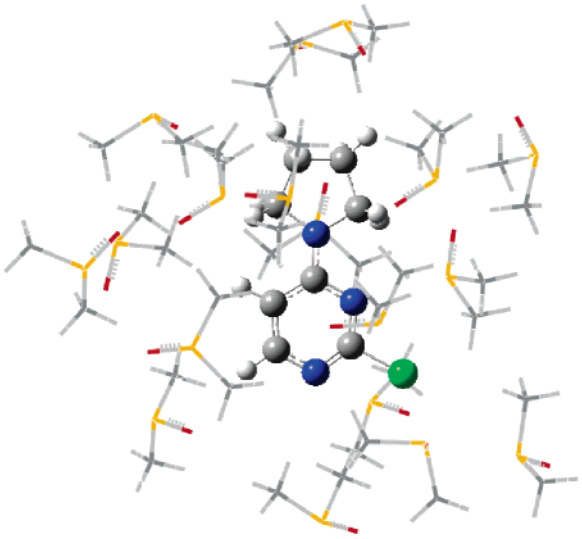


FIGURE 1. First solvation shell of chloropyrimidine **1** in DMSO solution. For ONIOM calculations the solute is computed at B3LYP (ball and stick) and the solvent molecules (wireframe) at HF level.

stochastic dynamics, the SHAKE algorithm applied to bonds to hydrogens).[20] To analyze the solute at the same level of theory as the isolated molecule an additional QM/MM calculation (B3LYP/6-31G(d,p)//OPLS2001) was performed with the QSite program.[21] Then, the modified box was reduced to a smaller and tractable system that includes the solute and the closest solvent molecules within a radius of 4.0 Å. This operation yields the solute surrounded by a well-defined first solvation shell (approximately 20 solvent molecules).[22] Given that the nuclear magnetic shielding calculations of these supermolecules are still computationally expensive, a comparative study between the full DFT level and the two-layer GIAO-ONIOM(DFT:HF) method was carried out.[23] The system partitioning was straightforward, considering the solute (molecule of interest) at the high level (B3LYP) and the remaining (solvent molecules) in a cheaper HF region (Figure 1).

Results and Discussion

^{13}C Chemical Shifts of Compounds 1 and 2. All the carbon nuclear magnetic shielding values obtained both experimentally and computationally are collected in Table 1. Figure 2 shows the scaled mean absolute errors (CMAE) of all computed methods for compounds **1** and **2** in both solvents.[24] First, it can be seen from entries 1 and 2 that though the molecules are polar (dipolar moments of 6.4 and 2.6 D at B3LYP/6-31G(d,p) level for **1** and **2**, respectively) the ^{13}C chemical shifts are relatively insensitive to the solvent used (differences are less than 2.6 ppm). The experimental results were first compared with the ACD/Lab software prediction. For compound **1** the maximum discrepancy is only 4.3 ppm (at C3 atom) and the scaled mean absolute error (Figure 2) is comparable or less than the CMAE found in most of the ab initio calculations (vide infra). Instead, the C4 atom of compound **2** is considerably overestimated (8 ppm discrepancy). On the other hand, the shielding order observed in the experiment for compound **1** (C1 and C2 nuclei with the same and highest chemical shift) is not maintained when using this program. Similar conclusions in terms of relative order can also be drawn for compound **2**. In agreement with the experiment the C3 atom in **1** is relatively shielded compared to the C4 atom in **2**. The parametrized semiempirical NMR-Predict software gives much better results (about half of the errors) than ACD for compounds **1** and **2** in both solvents. Again, it does not reproduce the experimental order for C1, C2, and C4 atoms regardless of the variant employed (either network or hose, entries 4 and 5).[25] In both cases, the network version was found to be more accurate than the hose variant. The estimation of the chemical shifts by using a trained back-propagation neural net from fast PM3 semiempirical MO calculations was relatively poor.[26] The following entries (7−15) analyze the effect of the level of theory and basis set of quantum NMR predictions, both in the geometry optimization and in GIAO calculations.[27] The standard B3LYP level of theory with use of the standard 6-31G(d,p) basis set (entry 7) gives quite promising results, with only slightly worse statistics than those from dataset based approaches. Interestingly, the increase of the basis set in the GIAO calculation at the same geometry (entry 8) did not improve the results by any means. Indeed, disappointing discrepancies at C1 of 24 and 14 ppm were found for compounds **1** and **2**, respectively. A complete basis set (CBS) approach has been proposed to improve accuracy but at least in this case it would not be appropriate.[28] Given that B3LYP/6-311+G(d,p) can be considered a nice tradeoff between accuracy and speed for larger molecules we have also considered this basis set in both geometry and GIAO calculations (entry 9).[29] Again, the results are deceivably similar to the previously computed method, showing that bs1 is sufficient enough for geometry optimizations and the variation in NMR parameters from bs2 to bs3 is expected to be small. It is worth noting in these last two entries that the chemical shifts of all carbon atoms are overestimated. To study exclusively the influence of the level of theory in GIAO calculations entries 10, 11, and 12 were also computed, keeping the basis set constant and the same DFT level of optimization. The Hybrid MPW1PW91 level has proved to be one of the best DFT levels of theory to analyze the magnetic features of molecules but no relevant differences with

(20) (a) *MacroModel 8.6*; Schrödinger, Inc.: Portland ,OR, 2004. (b) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440−467.

(21) (a) *Qsite 3.0*; Schrödinger, Inc.: Portland, OR, 2004. (b) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468−1494. (c) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *16*, 1442.

(22) Mennucci, B.; Martinez, J. M.; Tomasi, J. *J. Phys. Chem.* **2001**, *105*, 7287−7296.

(23) Karadakov, P. B.; Morokuma, K. *Chem. Phys. Lett.* **2000**, *317*, 589−596. σ^C[ONIOM2(DFT-GIAO:HF-GIAO] = σ^C(DFT-GIAO,model) + σ^C-(HF-GIAO,real) − σ^C(HF-GIAO,model).

(24) Statistical parameters considered in this paper: slope (*a*), intercept (*b*), and linear correlation coefficient (*r*2) are obtained by a linear fit of the calculated vs experimental ^{13}C NMR chemical shifts; mean absolute error (MAE = $\sum[|\delta_{exp} - \delta_{cal}|]/n$), mean error (ME = $\sum[\delta_{exp} - \delta_{cal}]/n$), scaled mean absolute error (CMAE = $\sum[|\delta_{exp} - \delta_{scaled}|]/n$, being $\delta_{scaled} = (\delta_{cal} - b)/a$)) and STEYX function is the standard error of the regression. See Supporting Information for full details.

(25) http://www.modgraph.co.uk; NMRPredict. Modgraph Consultants Ltd.

(26) (a) Clark, T.; Alex, A.; Beck, B.; Burkhardt, F.; Chandrasekhar, J.; Gedeck, P.; Horn, A. H. C.; Hutter, M.; Martin, B.; Rauhut, G.; Sauer, W.; Schindler, T.; Steinke, T. *VAMP 8.1*; Erlangen, 2002. (b) Clark, T.; Rauhut, G.; Breindl, A. *J. Mol. Model.* **1995**, *1*, 22−35.

(27) The reference used to calculate the relative shifts was the methane molecule at the same level of theory used in the GIAO calculation. Given that TMS is the experimental reference, a correction (δ = 0.23 and −2.3 ppm) was finally made in order to determine the chemical shifts for ^{1}H and ^{13}C atoms, respectively.

(28) Wiberg, K. B. *J. Comput. Chem.* **2004**, *25*, 1342−1346.

(29) (a) 6-311++G(d,p): Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frish, M. J. *J. Chem Phys.* **1996**, *104*, 5497−5509. (b) 6-311+G-(2d,p): Kupka, T.; Ruscic, B.; Botto, R. E. *J. Phys. Chem. A* **2002**, *106*, 10396−10407.

TABLE 1. Experimental and Calculated ^{13}C and ^{1}H Chemical Shifts for Compounds 1 and 2

		1[f]						**2**[f]					
no.	methodologies used[a]	C1	C2	C3	C4	H3	H4	C1	C2	C3	C4	H3	H4
1	exptl (s1,DMSO)[b]	160.1	160.6	101.9	155.7	6.19	7.98	160.7	159.9	158.4	108.0	8.18	6.50
2	exptl (s2,$CDCl_3$)[b]	159.2	160.3	102.9	156.2	6.48	8.01	163.3	161.9	159.6	108.0	8.25	6.64
3	ACD 8.0[c]	159.8	156.9	107.2	159.5	6.57	7.86	161.6	158.4	159.9	116.3	8.09	6.31
4	*Pred*: H-s1; network-C[d]	159.2	160.0	100.1	159.5	6.47	8.76	162.3	151.4	159.7	111.5	8.76	6.59
5	*Pred*: H-s2; hose-C[d]	161.7	154.7	100.1	159.5	6.47	8.76	161.1	156.2	159.0	109.0	8.76	6.59
6	VAMP 8.1[e]	183.6	166.9	69.9[g]	125.3			175.9	159.3	158.4	99.7		
7	DFT(bs1)//DFT(bs1)	168.1	155.0	98.9	152.4	5.77	7.87	166.6	155.2	153.6	106.7	7.97	6.13
8	DFT(bs3)//DFT(bs1)	182.7	171.4	109.2	169.0	6.07	8.22	177.0	168.4	165.3	112.6	8.30	6.56
9	DFT(bs2)//DFT(bs2)	180.9	170.3	108.7	167.3	5.92	8.18	180.2	169.9	168.6	116.6	8.22	6.39
10	DFT2(bs1)//DFT(bs1)	167.4	155.4	99.1	153.4	5.92	8.02	166.2	155.6	154.8	106.6	8.13	6.27
11	DFT3(bs1)//DFT(bs1)	167.7	150.8	98.3	148.8	5.66	7.71	165.1	152.1	149.8	106.0	7.81	6.01
12	MP2(bs1)//DFT(bs1)	174.6	170.7	99.0	167.0	5.59	8.08	176.2	168.1	169.0	106.0	8.21	5.92
13	MP2(bs1)//MP2(bs1)	176.5	170.5	98.1	167.5	5.54	8.05	177.0	169.0	169.9	105.6	8.19	5.82
14	HF(bs1)//HF(bs1)	162.1	160.4	90.8	156.5	5.40	7.80	165.4	156.0	158.8	97.6	7.97	5.77
15	HF(bs3)/HF(bs1)	173.6	172.9	97.0	168.9	5.74	8.15	177.3	168.7	171.2	104.2	8.27	6.24
16	DFT(bs2)s1//DFT(bs2)	178.9	171.1	111.9	167.4	6.54	8.39	180.0	169.9	170.8	116.6	8.56	6.81
17	DFT2(bs2)s1//DFT(bs2)	176.1	169.3	110.6	166.4	6.73	8.54	177.6	168.0	167.5	115.5	8.36	6.52
18	DFT3(bs2)s1//DFT(bs2)	179.2	168.2	112.5	165.0	6.49	8.29	179.5	167.7	167.9	117.0	8.45	6.73
19	DFT(bs2)s1//DFT(bs2)s1	179.2	171.3	112.4	168.2	6.57	8.41	180.0	170.3	171.6	116.6	8.63	6.91
20	DFT(bs1)s1//DFT(bs1)s1	166.9	156.0	102.0	153.1	6.37	8.08	166.2	155.7	156.2	106.5	8.35	6.58
21	DFT(bs1)s2//DFT(bs1)s2	166.7	155.6	100.1	152.8	6.17	8.01	166.3	155.6	155.4	106.5	8.23	6.44
22	HF(bs1)s1//HF(bs1)	160.4	161.2	92.9	156.6	6.02	8.09	165.3	156.0	161.1	97.7	8.38	6.21
23	HF(bs1)s2//HF(bs1)	160.9	161.0	92.2	156.6	5.82	8.00	165.2	156.0	160.4	97.5	8.25	6.07
24	cluster s1 ONIOM (bs1)	161.3	161.7	93.0	154.2	5.96	8.47	164.7	158.8	160.5	94.6	8.53	6.26
25	cluster s1 DFT (bs1)	167.6	156.3	101.6	156.5	6.17	8.59	165.4	156.2	158.3	106.8	8.68	6.71
26	cluster s2 ONIOM (bs1)	163.3	160.9	94.1	156.9	6.55	8.65	164.5	157.5	159.7	98.5	8.47	5.98
27	cluster s2 DFT (bs1)	167.9	157.1	105.9	153.6	7.08	8.53	166.3	156.4	155.9	110.3	8.26	6.17

[a] See the Computational Methods section for details. [b] Experimental values were retrieved from literature (ref 9). [c] Values obtained from ACD version 8.0. [d] NMRPredict software values showing ^{1}H chemical shifts in DMSO and chloroform and ^{13}C chemical shifts with use of HOSE and Networks variants (ref 25). [e] PM3 (VAMP) values for carbon magnetic shieldings (ref 26). [f] Atom numbering of compounds **1** and **2** according to Figure 3. [g] Out of range predicted value (diagnostic message printed).

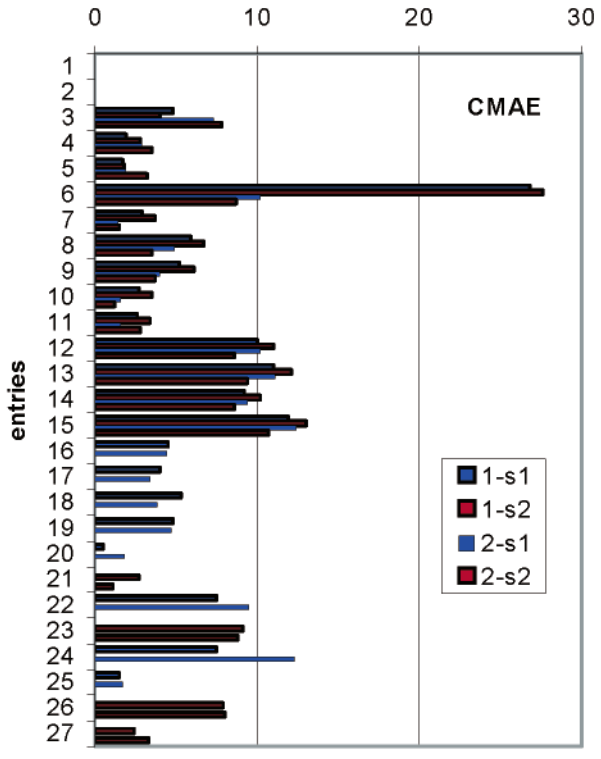


FIGURE 2. Scaled mean absolute errors analyzing ^{13}C chemical shifts for compounds **1** and **2**. The values obtained from gas-phase calculations were compared with both solvents. Each entry stands for the corresponding computed method shown in Table 1.

B3LYP are found in our study case (entry 10).[30] Although the pure (and faster) BLYP level presents similar CMAEs to the other DFT variants the maximum errors in certain carbons are significantly worse (entry 11).[31] As previously reported in the literature, the MP2 procedure is not as successful as DFT methods and provided relatively worse results.[32] The B3LYP and MP2 optimization levels afford identical geometries except for the C−Cl distance (Figure 3) and no significant magnetic variations were observed (entries 12 and 13). Several GIAO calculations on HF optimizations were examined to explore how the neglect of electron correlation in the geometries affects the magnetic parameters. Surprisingly, a simple HF//HF methodology (entry 14) reproduced the chemical shifts of almost all carbons, except for the values at C3 and C4 for compounds **1** and **2**, respectively. An explanation for this relies on the well-known overestimated contribution of the ionic resonance forms to the real hybrid structure at the HF level of theory (Scheme 2).[33] As a consequence, the negative charges of the aforementioned carbons are exaggerated and the chemical shifts are computed at higher fields. Figure 3 noticeably shows the structural and electrostatic differences between HF and correlated B3LYP levels. It can also be analyzed by a magnetic point of view, computing the aromaticity of these six-membered rings by means of the NICS criterion.[34] In both compounds B3LYP optimizations exhibit higher aromatic character than HF geometries.[35] In a related study of amino pyrimidines and

(30) Cimino, P.; Gomez-Paloma, L.; Duca, D.; Riccio, R.; Bifulco, G. *Magn. Reson. Chem.* **2004**, *42*, S26−S33.

(31) See ref 29a for ^{13}C chemical shifts at different DFT levels. BLYP is a pure DFT method where the coupled-perturbed Hartree−Fock equations are avoided; therefore nuclear magnetic shielding calculations are faster with this method than with HF and B3LYP variants. For the size of this study, the relative speed was about B3LYP=1.25*HF and BLYP=0.75*HF.

(32) Significant underestimation of the paramagnetic terms has been generally observed. See: Wiberg, K. B. *J. Comput. Chem.* **1999**, *20*, 1299−1303.

(33) Carpenter, J. H.; McGrath, M. P.; Hehre, W. J. *J. Am. Chem. Soc.* **1989**, *111*, 6154.

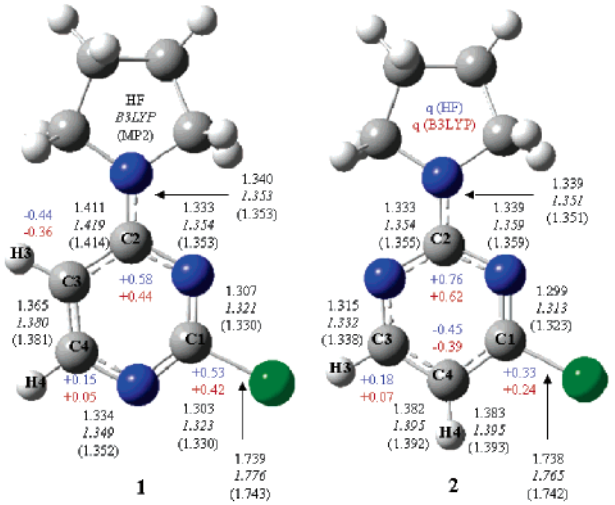
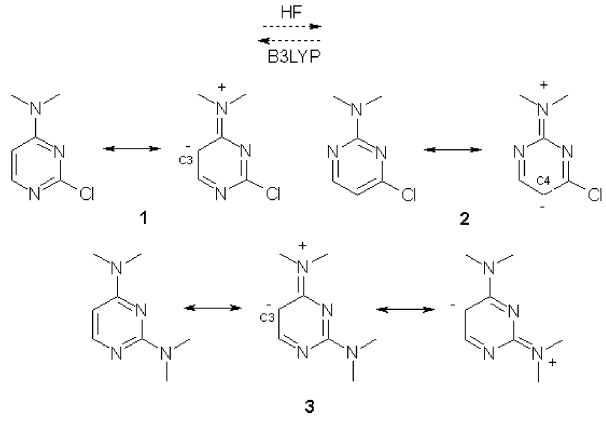


FIGURE 3. Bond distances (Å) and NBO charges *q*(e) of compounds **1** and **2** computed at different theoretical levels.

SCHEME 2. Resonance Forms for the Studied Compounds 1, 2, and 3



benzenes correlations of chemical shifts with 2pz electron densities have been observed, showing that the experimental chemical shifts can be expressed as a linear function of BPW91/STO-3G Mulliken 2pz orbital populations.[36] An examination of NBO and Mulliken atomic charges of the carbons in compounds **1** and **2** at both HF and B3LYP levels showed this trend in most cases (see the Supporting Information). Gryff-Keller et al. have pointed out recently the systematic discrepancy of ^{13}C chemical shifts for chlorinated aromatic carbons at the B3LYP level, which is not manifested either at the Hartree−Fock level or for non-chlorinated analogues.[37] A comparison between calculated and crystal structures from the Cambridge Structural Database (CSD) shows that the C−Cl bond distance is well-reproduced at HF and MP2 levels, but not at the DFT level (Figure 3).[38] This can explain the large discrepancy for the C1 atom at the DFT level relative to the remaining aromatic carbons for **1** and **2**.[39] The use of a much larger basis set (entry 15) results in much less shielding, giving worse statistical errors. From this initial gas-phase analysis the hybrid DFT methodology with a standard basis set is the best quantum approach, at a comparable accuracy level to the parametrized semiempirical NMR prediction, to match the experimental carbon results.

To analyze the solvent effects on magnetic shielding calculations, continuum dielectric medium and discrete cluster solvation models were considered (entries 16−27). While reaction field (continuum) methods provide an effective way to take into account purely electrostatic long-range interactions, the solute surrounded by a number of explicitly treated solvent molecules (usually named as cluster or super molecule) is the preferred methodology to describe the short-range interactions, in particular when the solvent under study is very apolar.[40] Although the combination of both approaches is desirable it was not considered in this paper for practical and computationally prohibited reasons. The dependence on the basis set in GIAO calculations using the polarizable continuum model has been studied and found to behave similar to that found for in vacuo calculations, defining the 6-311+G(d,p) basis as the best compromise between accuracy and computational demand.[41] Therefore, the first analysis carried out corresponds to the influence of solvent effects by means of PCM-like methods in the nuclear magnetic constants. Although DMSO is a solvent with a large dipole moment and high dielectric constant (entry 9 vs entry 16) the ^{13}C nucleus is not very sensitive to this external field (less than 2 ppm variation). This finding is in accordance with the experimental data (entry 1 vs entry 2). In fact, only a small improvement is achieved in terms of CMAEs after including the solvent effect. In addition, no significant differences are obtained (at this optimization level) from the GIAO analysis at the different functionals (entries 16, 17, and 18). Even for a quite polar molecule such as **1**, the additional geometry optimization of the solute in solution does not appear to be necessary (less than 1 ppm variation, entry 19). Similar conclusions are derived when using the standard 6-31G(d,p) basis set (entry 7 vs entries 20 and 21), except for the improvement due to the solvent effect now being more pronounced. Given that the dielectric constant of chloroform solvent is lower than that of DMSO, the relative variations of the chemical shifts are even smaller. The calculation of the ^{13}C nuclear magnetic shieldings within the framework of the PCM at the Hartree−Fock level affords the same trend found at the DFT level (entry 14 vs entries 22 and 23). Finally, a combination of molecular dynamics simulations and quantum mechanical calculations was performed to mimic the solution state surrounding the solute. This computational strategy has been successfully applied to conformational preference studies in

(34) (a) Schleyer, P. v. R.; Maerker, C.; Dransfeld, A.; Jiao, H.; Hommes, N. J. R. v. E. *J. Am. Chem. Soc*. **1996**, *118*, 6317−6318. (b) Chen, Z.; Wannere, C. S.; Corminboeuf, C.; Puchta, R.; Schleyer, P. v. R. *Chem. Rev*. **2005**, *105*, 3842−3888.

(35) The NICS values are −6.4 (−5.6) and −5.9 (−5.4) ppm for B3LYP and HF geometries respectively for compound **1** (**2**), calculated at the (3,+1) ring point of electron density according to Bader at the GIAO-B3LYP/6-31G(d,p) level. See: Morao, I.; Cossio, F. P. *J. Org. Chem*. **1999**, *64*, 1868−1874.

(36) Barfield, M.; Fagerness, P. *J. Am. Chem. Soc*. **1997**, *119*, 8699−8711.

(37) Gryff-Keller, A.; Molchanov, S. *Mol. Phys*. **2004**, *102*, 1903−1908.

(38) Ma, B.; Lii, J.-H.; Schaefer, H. F., III; Allinger, N. L. *J. Phys. Chem*. **1996**, *100*, 8763−8769.

(39) The CSD search provides 23 and 29 entries for **1** and **2**, respectively, using the corresponding C4N3Cl scaffold. The experimental C−Cl bond length is found to be ca. 1.73 Å. See Figure 3 for comparison with computed values at different levels of theory. Indeed, a partial optimization at the same level of theory of compound **1** restraining the C−Cl distance to the experimental value showed a better agreement (168.2 vs 166.3 ppm).

(40) GIAO-PCM methodology also fails if the solute−solvent system exhibits cooperative charge transfer, see: Klein, R. A.; Mennucci, B.; Tomasi, J. *J. Phys. Chem*. **2004**, *108*, 5851−5863.

(41) Cammi, R.; Mennucci, B.; Tomasi, J. *J. Chem. Phys*. **1999**, *110*, 7627−7638.

chloroform solution.[42] To define the radius size of this complex, radial pair distribution RPD functions can be used. A well-defined first solvation shell is obtained by using a cutoff distance over 4.0 Å (see the Computational Methods section for details).[22] Keeping the optimization of the compound of interest at the same level of theory, the effect of the environment in isolated and solvated systems can be easily analyzed. It is noteworthy that though the real situation is dynamically complicated with lots of different structures coexisting only an equilibrated conformation of such a phenomenon for each case was computationally considered here. By comparing entry 7 with entries 25 and 27 modest differences (more than 2 ppm) can be observed, but only on those carbons (C3 and C4) that are linked to a hydrogen atom (C3 and C4). Indeed, given that the hydrogens are the external atoms of the molecule some polarization of these C−H bonds facing the environment (solvent) is expected. Surprisingly, similar or poorer CMAE values are obtained after including the surrounding microenvironment. Another important fact is that the shielding (−) or deshielding (+) of each carbon during the gas-to-solution transition is consistent in both clusters. Moreover, this consistency is also kept with use of PCM-like methods. However, the magnitude of these shifts does not correlate with the dielectric constant. Therefore, although electrostatic forces are present, significant other nonelectrostatic interactions are equally effective. While the continuum solvation models showed a better agreement with the experiment than gas-phase calculations, in the discrete solvation approach this statement is only correct for compound **1**, not for **2**. Given that these clusters imply around 20 solvent molecules (over 200 atoms) an additional multilayer ONIOM approach has also been considered to achieve significant computational savings.[23] In fact, ONIOM-GIAO methodology has been used to predict accurately ^{13}C chemical shifts of organic species embedded in large systems.[43] When the ONIOM isotropic NMR chemical shielding approach[23] at the B3LYP(bs1)//HF(bs1) level (entries 24 and 26) is used, the values obtained are an average number between full HF and B3LYP methods. The CMAE values found in these entries are considerably worse than those from full DFT discrete solvated calculations (entries 25 and 27).

^{1}H Chemical Shifts of Compounds 1 and 2. All the proton nuclear magnetic shieldings are reported in Table 1. Figure 4 shows the mean absolute errors (MAE) of all computed methods for compounds **1** and **2** in both solvents. In contrast to ^{13}C chemical shifts, proton chemical shifts are very sensitive to solvent effects. Indeed, the H3 atom for compound **1** is deshielded 0.3 ppm, passing from chloroform to DMSO (entries 1 and 2). ACD software does not take into account the solvent used, showing a better agreement with the values obtained in chloroform. On the other hand, the NMRPredict program allows selection of the solvent (only for proton but not for carbon atoms). However, in this particular case, the predictions are found to be the same in both solvents and particularly poor (ca. 0.75 and 0.5 ppm discrepancies for compounds **1** and **2**, respectively). The remaining entries are related to the use of different quantum methodologies for the prediction of the proton

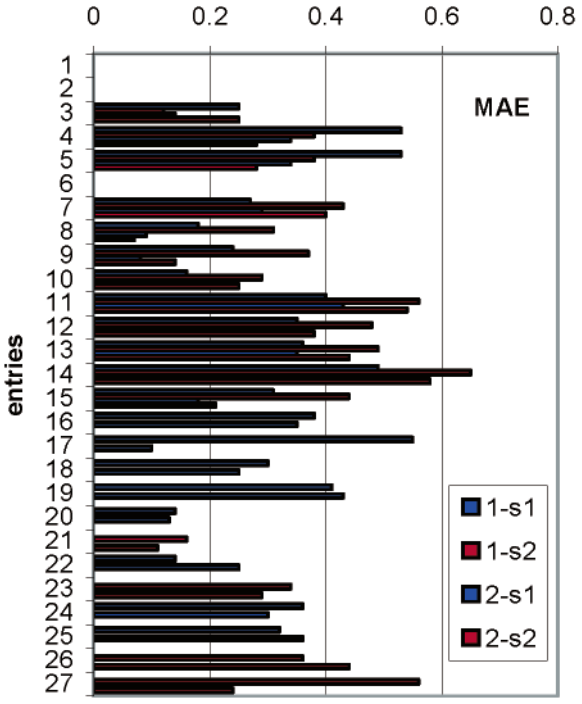


FIGURE 4. Mean absolute errors analyzing ^1H chemical shifts for compounds **1** and **2**. The values obtained from gas-phase calculations were compared with both solvents. Each entry stands for the corresponding computed method shown in Table 1.

chemical shifts. The standard B3LYP level of theory using the standard 6-31G(d,p) basis set (entry 7) underestimates the chemical shifts, with an unacceptable maximum deviation of 0.7 ppm. Unlike the ^{13}C chemical shifts, the increase of the basis set in the GIAO calculation at the same geometry (entry 8) shows much better results, comparable to those found with database-related ACD software. The use of an intermediate basis set (bs2) presents a midway error between bs1 and bs3. Therefore, the agreement with the experimental values consistently improves as the basis set is enlarged.[44] However, and regardless of the solvent analyzed, the largest error is not associated with the most acidic (i.e., deshielded) proton.[45] The analysis of the influence of the level of theory in GIAO calculations entries 10, 11, and 12 was also compared. The alternative hybrid DFT level (MPW1PW91) shows significantly better results, with almost half mean errors than B3LYP values.[30] In contrast, the pure BLYP method presents worse mean errors than those from the B3LYP level, in both molecules and solvents.[31] Therefore, although the BLYP variant is computed at lower cost, an important consideration as the molecular size increases, the results at least in this example are not encouraging. The MP2 procedure showed slightly inferior results than the B3LYP method (entry 12). The B3LYP and MP2 geometries are practically identical and consequently no significant magnetic variations are found (less than 0.1 ppm, entry 13). When the GIAO calculations are performed at the HF level over HF geometries the results found are exceptionally inaccurate, with a disparity of over 1 ppm. Therefore, the electron correlation in the geometries and GIAO calculations is of paramount importance in aromatic proton chemical shift predictions. In line

(42) Casanovas, J.; Namba, A. M.; Leon, S.; Aquino, G. L. B.; Jose da Silva, G. V.; Aleman, C. *J. Org. Chem.* **2001**, *66*, 3775−3782.

(43) (a) Zheng, A.; Chen, L.; Yang, J.; Yue, Y.; Ye, C.; Lu, X.; Deng, F. *Chem. Commun.* **2005**, 2474−2476. (b) Gascon, J. A.; Sproviero, E. M.; Batista, V. S. *J. Chem. Theory Comput.* **2005**, *1*, 674−685. (c) Benzi, C.; Crescenzi, O.; Pavone, M.; Barone, V. *Magn. Reson. Chem.* **2004**, *42*, S57−S67.

(44) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357−7363.

(45) Perczel, A.; Csaszar, A. G. *J. Comput. Chem.* **2000**, *21*, 882−900.

TABLE 2. Statistical Analysis of ^{13}C and ^{1}H Chemical Shifts for Compounds 1 and 2[a–c]

		carbon						proton					
		1		**2**		**A**	**B**	**1**		**2**		**A**	**B**
no.	methodologies used	E_{11}	E_{12}	E_{22}	E_{21}	y/n	y/n	E_{11}	E_{12}	E_{22}	E_{21}	y/n	y/n
3	(s1) ACD 8.0	4.8	*0.8*	7.3	*11.3*	n	n	0.25	*0.20*	0.14	*0.11*	n	n
3	(s2) ACD 8.0	4.0	*2.0*	7.8	*10.6*	n	n	0.12	*0.23*	0.25	*0.13*	n	n
4	(s1)*Pred*: network	1.9	*7.0*	2.9	*7.3*	y	y	0.53	*0.31*	0.34	*0.59*	n	n
4	(s2) *Pred*: network	2.8	*6.2*	3.5	*6.5*	y	y	0.38	*0.34*	0.28	*0.43*	n	n
5	(s1)*Pred*: hose	1.7	*5.7*	1.9	*6.1*	y	y	0.53	*0.31*	0.34	*0.59*	n	n
5	(s2) *Pred*: hose	1.8	*5.1*	3.7	*5.3*	y	y	0.38	*0.34*	0.28	*0.43*	n	n
20	DFT(bs1)s1//DFT(bs1)s1	0.5	*5.4*	1.8	*3.8*	y	y	0.12	*0.12*	0.13	*0.38*	y	y
21	DFT(bs1)s2//DFT(bs1)s2	2.7	*6.2*	1.1	*3.1*	y	y	0.16	*0.36*	0.11	*0.13*	y	y

[a] For the statistical definitions see ref 24. [b] Italic numbers stand for cross CMAE (comparison between the theoretical values for X and the experimental values of the regioisomer Y). [c] Two different criteria (A and B) to assign unambiguously the isomeric mixture (see text). y/n means that the method is (yes/not) able to assign the compounds unequivocally.

with carbon chemical shift results, the worse values are found at H3 and H4 for compounds **1** and **2**, respectively. As mentioned before, the overestimated weighting of the ionic valence forms to the genuine structure at the HF level of theory is the main reason for disagreement. Again here, the larger the basis set the higher the precision obtained (entries 14 and 15).[44,45] An examination of NBO and Mulliken charges of the hydrogen atoms in compounds **1** and **2** at both HF and B3LYP levels shows that an increased cationic character of the proton leads to a higher upfielded value. From the gas-phase analysis the hybrid DFT methods using the largest basis set were found to be the best quantum approaches. At this point it is also worth noting that the DFT magnetic shieldings calculated with standard exchange-correlation (EC) functionals still have not shown decisive improvement over the HF results for nonaromatic compounds.[44−46] The conventional DFT methods were parametrized to be accurate for potential surfaces and geometries but not for nuclear magnetic shieldings. Alternatives to optimize DFT methods for NMR shielding calculations have been proposed recently.[47] However, in this paper only the three common aforementioned DFT variants have been analyzed.

As carried out during the investigation of carbon magnetic shieldings, the solvent effects on magnetic shielding calculations were studied by means of continuum dielectric medium and discrete cluster solvation models. The first analysis corresponds to the influence of solvent effects by means of PCM-like methods on GIAO calculations with DMSO as solvent. In contrast to the behavior of ^{13}C, dramatic variations are found (entry 9 vs 16) between the gas phase and DMSO solution (in the 0.2−0.6 ppm range). This finding is in accordance with the experimental values in DMSO and chloroform (entry 1 vs entry 2). It is notable that the agreement between the computed and experimental results is much better with the gas-phase values (entry 9) rather than the DMSO solution results (entry 16). Contrary to the gas phase, MPW1PW91 and BLYP variants show comparable or marginally better results that those from B3LYP. Furthermore, the additional geometry optimization of the compounds in solution does not turn into a required step (ca. 0.1 ppm variation, entry 19). The calculation of the proton magnetic shieldings at the PCM-HF level affords much superior results than in vacuo HF results in both solvents (entries 22 and 23 vs entry 14). Apart from continuum models the cluster approach was considered at both full DFT and two-level ONIOM levels. The computational details are explained in the computational method and carbon chemical shifts subsections. A comparison between entry 7 and entries 25 and 27 shows that the presence of surrounding solvent molecules implies a constant significant shielded deviation. As in ^{13}C chemical shifts, poorer CMAE values were found after including the surrounding microenvironment. The higher the chemical shift value the higher the divergence is due to the solvent. Also, the variations are much pronounced in DMSO rather than chloroform. Moreover, these features are also set aside by using PCM-like methods. In general, the ONIOM approach at the B3LYP(bs1)//HF(bs1) level (entries 24 and 26) gives statistically better results than the corresponding full studies at the B3LYP(bs1) level (entries 25 and 27) only for compound **1**.

Unequivocal Assignment of Compounds 1 and 2 Analyzing ^{13}C and ^{1}H Chemical Shifts. In the last two sections we have compared the results from quantum and dataset based methods with the experimental values. In fact, a quantitative analysis of how accurate the predicted values are helps to develop better and more precise procedures. However, an unambiguous assignment (which compound is which) within the isomeric mixture is also of paramount importance. In our case study, we already have the experimental chemical shifts of two isomers **1** and **2** in two deuterated solvents. Therefore, the next step consists of analyzing the possibility to identify unequivocally these compounds with use of the predicted values from the previously aforementioned methods. This identification must be correct regardless of the solvent used. Thus, methodologies that do not take into account solvent effects would not be very efficient, especially in proton chemical shifts where the solvent is a critical factor. To distinguish each compound from the other using our predicted values we also need to calculate the cross errors, i.e., the differences between the predicted values for compound X with the experimental values of the isomeric compound Y. The combination of own and cross errors for each compound and in each solvent is reported in Table 2.[48,49] To the best of our knowledge there is not a well-established criterion for an unequivocal assignment. We offer here two criteria for such an objective. The first (named A) consists of the individual comparison between cross and own errors for each compound.

(46) (a) Wang, B.; Fleischer, U.; Hinton, J. F.; Pulay, P. *J. Comput. Chem.* **2001**, *22*, 1887−1895. (b) Wang, B.; Hinton, J. F.; Pulay, P. *J. Comput. Chem.* **2002**, *23*, 492−497.

(47) Magyarfalvi, G.; Pulay, P. *J. Chem. Phys.* **2003**, *119*, 1350−1357.

(48) For clarity in Table 2, the only quantum methodology found to assign correctly ^{1}H and ^{13}C chemical shifts in both solvents (DFT(bs1)s1−2// DFT-(bs1)s1−2) is shown in this table (entries 20 and 21). See the Supporting Information for the data of all 27 methodologies computed.

(49) We have used the scaled mean absolute error (CMAE) as the parameter to quantify the residuals between prediction and experiment. For proton chemical shifts the mean absolute error (MAE) was used instead as only two hydrogen atoms (aromatic) are analyzed.

The assignment is correct if at least the own error is smaller than the cross error for each compound: $E_{11} < E_{12}$ and $E_{22} < E_{21}$.

However, these intuitive but separate conditions are not taking into consideration the trend of both compounds at the same time. To achieve this, a better collective criterion (B) based on the interclass distance computed in SIMCA (Soft Independent Modeling of Class Analogy) is also presented here.[50−52] In parallel with the method of calculating the distance between two classes by using the cross and own residual variances, the second criterion B for an unambiguous assignment is defined as follows[53]

$$\left(\frac{{E_{12}}^2 + {E_{21}}^2}{{E_{11}}^2 + {E_{22}}^2}\right) > 3$$

In this study both criteria draw the same conclusions. In the ^{13}C case, semiempirical and most DFT approaches are qualitatively good enough to differentiate each isomer. However, only the continuum solvated hybrid DFT methodology yields a correct assignment when analyzing the proton chemical shifts in both solvents.

Conclusions

In this paper ^{1}H and ^{13}C nuclear magnetic shieldings of two chloropyrimidine species have been calculated with use of ab initio and DFT methodologies. The results have been compared with the experimental values and with those from ACD/Lab and NMRPredict software packages. The instant predictions made by these two programs are generally found to be accurate enough for both carbon and proton chemical shifts. Only a few quantum chemistry based approaches have been shown to be as accurate for determining ^{13}C and chemical shifts and superior in elucidating unequivocally the ^{1}H NMR spectra of the regioisomeric mixture under study. In this context it should be noted that calculations in chloroform are comparable with those in the gas phase but significantly different from those in DMSO-d_6. Therefore, the solvent must be considered explicitly or implicitly in the NMR calculations. Self-consistent reaction field (SCRF) models are found to be more precise (and faster) than the discrete solvation models (super molecule). Unlike ^{13}C chemical shifts, high correlated levels of theory and large basis sets are equally very important for the accurate prediction of proton chemical shieldings. In conclusion, the accuracy vs time dilemma seems to be only apparent for the prediction of ^{1}H chemical shifts, where quantum based methods can give us a better description of the molecule than the other approaches. However, to achieve this, the quantum method must include a correlated level of theory with a relatively large basis set and taking into account the solvent effect implicitly.

Acknowledgment. Pfizer, Inc. is acknowledged for continuing support to this research. This paper is dedicated to Prof. José M. Aurrecoechea for his pioneering work to incorporate molecular modeling in Organic Chemistry teaching. We would like to thank Jeff Seymour from Modgraph Consultans Ltd. for the NMR predictions using the NMRPredict software.

Supporting Information Available: Cartesian coordinates and absolute energies of compounds **1** and **2** at all levels of optimization, correlations between chemical shifts and atomic charges (Figures S1−S4), and all statistical parameters computed in this work (Tables S1−S9). This material is available free of charge via the Internet at http://pubs.acs.org.

JO0600149

(50) SIMCA is a modeling technique that builds a box for each category separately on the basis of a specified number of principal components and a critical distance with probabilistic meaning. The interclass distance D, the rate of misclassifications, and the rate of samples not classified into all categories determine the accuracy of the classification. $D^{1,2} = [(R_{12}^2 + R_{21}^2)/(R_{11}^2 + R_{22}^2)]^{1/2} - 1$

(51) Brereton, R. G. *Multivariate Pattern Recognition in Chemometrics*; Elsevier: Amsterdam, The Netherlands, 1992.

(52) Ebert, C.; Gianferrara, T.; Linda, P.; Masotti, P. *Magn. Reson. Chem.* **1990**, *28*, 397−407.

(53) As a metric rule of thumb, if the value of interclass distance between categories is over 3 these classes are considered well classified. We here propose a much less strict criterion. It would be the hypothetical case where both cross errors are ca. double their own errors. The higher the number (we suggest here 3 as a minimum) the higher is the confidence in the assignation.