Predicting Carbon-13 Nuclear Magnetic Resonance Chemical Shielding Tensors in Zwitterionic L-Threonine and L-Tyrosine via Quantum Chemistry†

Angel C. de Dios, David D. Laws, and Eric Oldfield*

Contribution from the Department of Chemistry, University of Illinois at Urbana—Champaign, 505 South Mathews Avenue, Urbana, Illinois 61801

Received April 6, 1994®

Abstract: We report the *ab initio* evaluation of the carbon-13 nuclear magnetic resonance shielding tensors for each carbon atom in crystalline, zwitterionic, L-threonine and L-tyrosine, using a gauge-including atomic orbital (GIAO) quantum chemical approach, with and without charge-field perturbation (CFP). For isolated molecules, there is a correlation coefficient, R^2, of 0.975 between experimental shift and computed shielding, with a slope of −1.03 and an rmsd of 12.3 ppm. This error is due primarily to large deviations in the C^o σ_{11} (in the CO sp^2 plane and perpendicular to C^α–C^o) and σ_{22} (perpendicular to the sp^2 plane). Incorporation of a point-charge lattice to represent the local charge field results in a decrease in rmsd to 6.4 ppm, due primarily to changes in σ_{11} and σ_{22}. In the icosahedral representation and with charge field perturbation, we find an overall rmsd of 4.4 ppm over a 200 ppm chemical shift range (slope = −0.992, R^2 = 0.997), while for the isotropic shifts alone the rmsd reduces to 3.8 ppm. Thus, combined use of charge-field perturbation and a gauge-including atomic orbital approach permits excellent prediction of carbon-13 isotropic chemical shifts and principal shift tensor elements in two zwitterionic polar amino acids. The charge-field approach is particularly useful since it allows for inclusion of environmental effects on shielding without adding to the number of contracted functions. Moreover, the polarization effects are primarily limited to C^o, supporting the idea that for ^{13}C, long-range electrostatic field contributions to shielding will be small, especially for sp^3 carbons. The ability to successfully predict ^{13}C shielding tensor elements in highly polar (zwitterionic, hydroxyl-containing) amino acids provides strong additional support for the adequacy of GIAO/CFP-GIAO methods in predicting ^{13}C chemical shifts in proteins, and other macromolecules as well.

Introduction

With the recent development of relatively affordable high-performance reduced instruction set computer workstations, together with the observation that nuclear magnetic resonance (NMR) chemical shifts in macromolecules such as proteins are extremely sensitive to local conformation or structure,[1,2] there has been considerable interest in using *ab initio* quantum chemical methods to predict nuclear magnetic resonance spectra of proteins[3-6] and simpler model systems.[7] Once this hurdle is crossed, then the process can begin to be reversed, and chemical shifts can be used to deduce structure.[8] However, in some cases there may be questions as to the importance of more long-range interactions—such as electrostatic field effects—which for some nuclei may contribute appreciably to shielding.[3] We have thus investigated ^{13}C shielding in two zwitterionic, polar amino acid crystals, L-threonine and L-tyrosine, to evaluate the extent of these long-range contributions to shielding in two small polar molecules whose high-resolution neutron diffraction structures are known[9,10] and whose ^{13}C shielding tensors have been reported.[11,12]

The NMR chemical shielding, with its three principal elements and their orientation with respect to the molecular frame, contains six pieces of information, all of which can be independent in molecules that lack symmetry. Although our present ability to reproduce isotropic shielding values of individual sites in proteins from *ab initio* methods[3-6] is certainly very useful, a precise calculation of shielding tensor elements and the directions of the principal components is also desirable. This is primarily due to the fact that there may be accidental cancellations of errors in the theoretically evaluated individual tensor elements. However, experimentally it is very difficult to determine accurate shift tensors in proteins, due to sensitivity limitations. Thus, in order to assess the present quality of computed shielding tensor values, we have chosen to study two crystalline amino acids, threonine and tyrosine. Both contain "functional" OH groups, which could complicate accurate shielding calculations due to hydrogen bonding, and the use of zwitterionic crystals, containing $^+NH_3$ and CO_2^- moieties, also presents a fairly stringent test of our ability to handle rather strong electrostatic field contributions to shielding in real systems. Thus, the ability to accurately predict all principal shielding tensor elements—even including the (charged) carboxylates—in the zwitterionic forms of two polar amino acids, can be taken as a good test of our ability to evaluate ^{13}C shifts of polar residues in macromolecules. In particular, for threonine and tyrosine we need to evaluate 39 individual shielding tensor elements, covering a 250 ppm chemical shift range.[11,12]

† This work was supported by the United States National Institutes of Health (Grant No. HL-19481) and (in part) by the American Heart Association (Grant No. AHA 92-013340) with funds provided in part by the American Heart Association, Illinois Affiliate, Inc., and by equipment grants from the International Business Machines Corporation, Shared University Research Instrumentation Program, and the Graduate Research Board of the University of Illinois.

* Abstract published in *Advance ACS Abstracts*, July 15, 1994.

(1) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490.
(2) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol* **1991**, *222*, 311.
(3) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491.
(4) Laws, D. D.; de Dios, A. C.; Oldfield, E. *J. Biomol. NMR* **1993**, *3*, 607.
(5) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *J. Am. Chem. Soc.* **1993**, *115*, 9768.
(6) de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*. 5307.
(7) Jiao, D.; Barfield, M.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 10883.
(8) Pearson, J. G.; Le, H.; de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.*, submitted for publication.
(9) Ramanadham, M.; Sikka, S. K.; Chidambaram, R. *Pramana* **1973**, *1*, 247.
(10) Frey, M. N.; Koetzle, T. F.; Lehmann, M. S.; Hamilton, C. *J. Chem. Phys.* **1973**, *58*, 2547.
(11) Janes, N.; Ganapathy, S.; Oldfield, E. *J. Magn. Reson.* **1983**, *54*, 111.
(12) Frydman, L.; Chingas, G. C.; Lee, Y. K.; Grandinetti, P. J.; Eastman, M. A.; Barrall, G. A.; Pines, A. *Isr. J. Chem.* **1992**, *32*, 161.

0002-7863/94/1516-7784$04.50/0

Results and Discussion

Although the sensitivity of a given sp^3 carbon to electrostatic field effects, or polarization, is expected to be quite small,[13] aromatic carbons might be expected to be more polarizable, and charged residues (CO_2^-) would be expected to undergo large shielding changes due to intermolecular interactions, in this case with neighboring $^+NH_3$ groups. Thus, the effects of neighboring molecules in the zwitterionic lattice must be taken into account. As observed by neutron diffraction[9,10] both L-threonine and L-tyrosine participate in three dimensional networks of hydrogen bonds, with both the OH and CO_2^- groups being involved. We thus need to construct an appropriate lattice so that the effects of intermolecular interactions can be evaluated.

Both L-threonine and L-tyrosine crystallize in the space group $P2_12_12_1$. There are four molecules in the unit cell and their coordinates (x,y,z) were generated by the following symmetry operations: (1) x,y,z; (2) $^1/_2-x$, $1-y$, $^1/_2+z$; (3) $1-x$, $^1/_2+y$, $^1/_2-z$; and (4) $^1/_2+x$, $^1/_2-y$, $1-z$. Unit cell translations were then used to generate 28 additional molecules in the vicinity of the unit cell. Shielding calculations were performed with the TEXAS90 program of Pulay, Wolinski, and Hinton, which makes use of an efficient implementation[14] of the gauge-including atomic orbital (GIAO) methods introduced earlier.[15,16] A uniform 6-31G** basis was used for all calculations. Only one molecule had basis sets assigned to its atoms, with neighboring molecules being represented by point charges obtained from AMBER.[17] The adequacy of using point charges, as opposed to a full *ab initio* calculation, has already been demonstrated in model systems involving fluorobenzene interacting with a number of hydrogen fluoride molecules.[18] Therefore, so long as the environmental effects on shielding arise from electrical polarization, the types of shielding calculation we have used previously, which take into account the inhomogeneous electrostatic field the nuclei find themselves in, can be used, and one particularly convenient way of introducing the electrical perturbation is to represent neighboring molecules with point charges—the charge field perturbation—gauge-including atomic orbital (CFP-GIAO) approach.[3,18]

The combined results for the ^{13}C shielding tensors in L-threonine and L-tyrosine obtained by first using a single isolated molecule are illustrated graphically in Figure 1, in which we plot the observed shifts (experimental chemical shifts in ppm from TMS) *versus* theoretical shielding (ppm from the bare atom shielding value of 0 ppm). The slope of the fitted line in Figure 1 is −1.03 while the intercept is about 206 ppm (R^2 = 0.987, rmsd from the fitted line = 12.3 ppm). The points that significantly deviate from the fitted line are the tensor components σ_{11} and σ_{22} of the carboxyl ^{13}C site. The σ_{11} component lies in the sp^2 plane and is perpendicular to the C^α–C^o bond, while σ_{22} lies perpendicular to the sp^2 plane. Upon comparison with experiment, σ_{11} (A,B in Figure 1) is underestimated, while σ_{22} (C,D in Figure 1) is overestimated. Consequently, the discrepancies are less dramatic when only the isotropic values are considered. With regard to the intercept, 206 ppm is higher than the experimental absolute shielding value of 188.1 ppm measured by Jameson and Jameson for a cylindrical sample of neat TMS at 300 K.[19] This discrepancy, which appears as a constant offset, is largely due to the small basis set used in the calculations, since as shown by computations with larger basis sets (6-311++G(2d,2p)), values for the intercept as low as 193 ppm can be obtained, albeit on somewhat smaller molecules.[6] On the other hand, a smaller basis set (of double-ζ quality) without polarization functions yields an even higher value for TMS, 212 ppm, in the case of IGLO calculations.[20] However, although such absolute values are somewhat basis set dependent, the excellent correlations between theory and experiment shown here and elsewhere are not.[3-6]

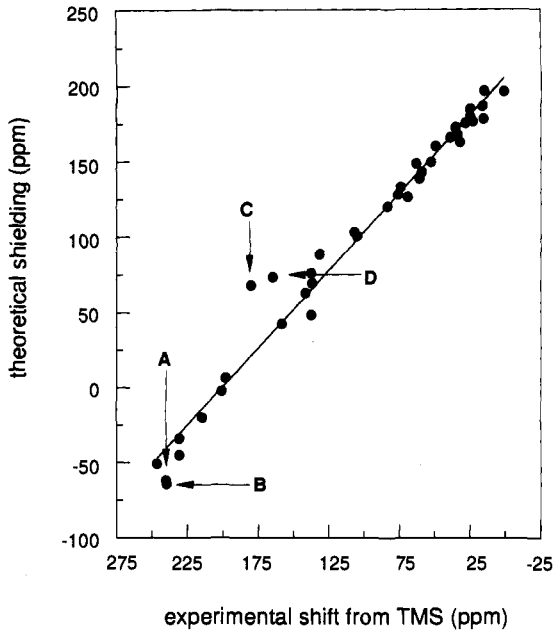

Figure 1. Graph showing the plot of experimental ^{13}C chemical shift (in ppm from TMS) versus theoretical shielding (in ppm) for each ^{13}C principal tensor component for L-threonine and L-tyrosine (zwitterionic forms; experimental data from refs 11 and 12): (A) σ_{11}[Tyr C°], (B) σ_{11}[Thr C°], (C) σ_{22}[Tyr C°], and (D) σ_{22}[Thr C°]. A uniform basis (6-31G**) was employed for both calculations, which were carried out on a cluster of IBM RISC/6000 workstations equipped with 40 Gbytes of disc space and operating at a peak theoretical speed of 1 Gflop. It required ~10 h to evaluate all tensor elements. Slope = −1.03, intercept = 206 ppm, R^2 = 0.975, rmsd = 12.3 ppm.

Considerable improvements are achieved when the calculations are performed in the presence of point charges, as shown in Figure 2 (slope = −1.04, intercept = 207 ppm, R^2 = 0.996, rmsd = 6.4 ppm), where the previous outlying points for σ_{11} and σ_{22} (A–D) are now much closer to the fitted line. The electrostatic contributions lead to a significant increase in the value of σ_{11} of C^o (16 ppm for L-tyrosine and 11 ppm for L-threonine) and a major decrease in σ_{22} (37 ppm for L-tyrosine and 27 ppm for L-threonine).

From the neutron diffraction structures, it can be seen that the carboxyl oxygens act as hydrogen bond acceptors involving three hydrogen atoms, and from the calculated shieldings presented in this work it is apparent that the primary electrostatic effect from these neighboring hydrogens is a reduction in the anisotropy and range of the carboxyl ^{13}C shielding tensor. The electrostatic field effects on the shielding of the alkyl ^{13}C sites are all very small (<2 ppm), while the aromatic sites in L-tyrosine exhibit a moderate sensitivity (3–8 ppm) to polarization.

As mentioned earlier, the shielding as a tensor involves not only its principal components but also the directions of these components with respect to a laboratory or molecular frame. Fortunately, the NMR work on L-threonine we reported previously involved a single-crystal study,[11] so that the experimental orientations of the ^{13}C shielding tensors are known, and if one compares the experimental and calculated direction cosines with respect to the crystallographic axes of the ^{13}C shielding tensors, a respectable agreement is seen (GIAO rmsd = 0.135). With inclusion of point charges, there is even better accord (rmsd =

(13) Batchelor, J. G. *J. Am. Chem. Soc.* **1975**, *97*, 3410.
(14) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251.
(15) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789.
(16) London, F. *J. Phys. Radium* **1937**, *8*, 397.
(17) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(18) de Dios, A. C.; Oldfield, E. *Chem. Phys. Lett.* **1993**, *205*, 108.
(19) Jameson, A. K.; Jameson, C. J. *Chem. Phys. Lett.* **1987**, *134*, 461.
(20) Schindler, M.; Kutzelnigg, W. *J. Am. Chem. Soc.* **1983**, *105*, 1360.

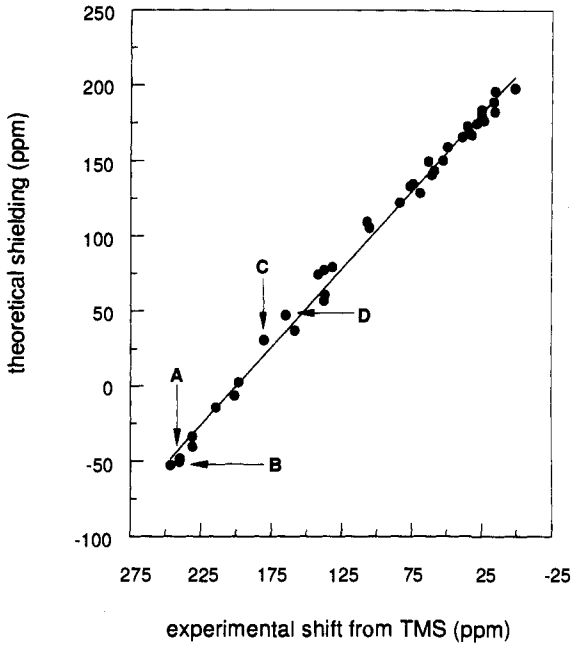

Figure 2. Experimental and theoretical ^{13}C NMR chemical shift/shielding principal tensor elements for L-threonine and L-tyrosine, with charge field perturbation: (**A**) σ_{11}[Tyr C°], (**B**) σ_{11}[Thr C°], (**C**) σ_{22}[Tyr C°], and (**D**) σ_{22}[Thr C°]. There is considerably better agreement for σ_{11} and σ_{22} with CFP. Slope = –1.04, intercept = 207 ppm, R^2 = 0.996, rmsd = 6.4 ppm.

0.092). Unfortunately, comparing tensor orientations in terms of direction cosines does not properly gauge the calculations presented here, since in addition to being difficult to visualize, errors in direction cosines do not translate directly to errors in terms of radians or degrees. Moreover, for tensors that are close to being spherical, errors can be magnified when taken in terms of direction cosines. To solve this problem, as suggested earlier by Alderman et al.[21] and Facelli and Grant,[22] use of an icosahedral representation affords a more convenient method for comparing the information content of experimental shift and theoretical shielding tensors, and this approach is discussed in detail in ref 21. In Figure 3, we show a comparison between the theoretical shieldings and experimental shifts for the ^{13}C sites in L-threonine, this time in the icosahedral representation. As can be seen, in this representation, the improvement with addition of point charges is similarly evident (CFP-GIAO slope = –0.992, R^2 = 0.997, rmsd = 4.4 ppm versus GIAO slope = –0.966, R^2 = 0.987, rmsd = 9.4 ppm), with the advantage that a separate comparison between direction cosines is not required. For a 220 ppm overall shielding tensor element range, the CFP-GIAO method yields an excellent slope, correlation, coefficient, and rms deviation from the fitted line of only 4.4 ppm—excellent agreement given 24 tensor elements (six for each of the four carbons in threonine). If isotropic chemical shifts alone are considered, the rmsd for threonine and tyrosine drops to 3.8 ppm, over a 160 ppm shift

(21) Alderman, D. W.; Sherwood, M. H.; Grant, D. M. *J. Magn. Reson.* **1993**, *101*, 188.
(22) Facelli, J. C.; Grant, D. M. *Nature* **1993**, *365*, 325.

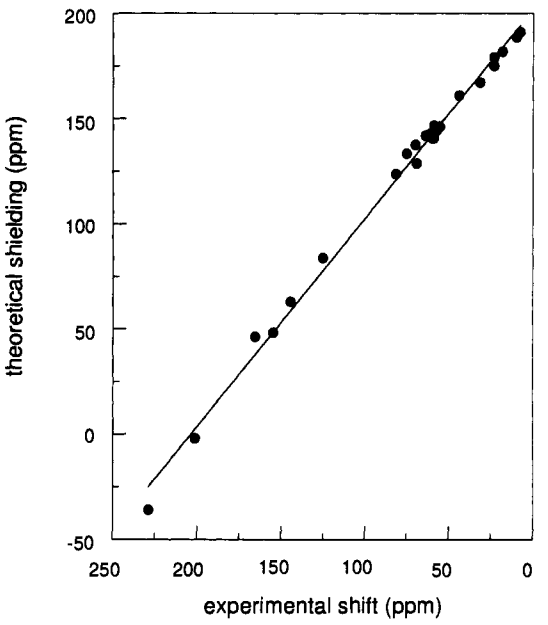

Figure 3. Experimental and theoretical ^{13}C NMR chemical shift/shielding tensor elements for L-threonine in the icosahedral representation, with charge field perturbation. Slope = –0.992, intercept = 202 ppm, R^2 = 0.997, rmsd = 4.4 ppm.

range, with CFP. Moreover, even without CFP, the predicted shieldings are in good accord with the experimental observations for all but the (charged) carboxylate groups.

Conclusions

The results presented here illustrate the adequacy of present theoretical methods for predicting chemical shieldings. The current CFP method of introducing electrostatic field effects into the shielding computations allows for inclusion of environmental effects without increasing the number of contracted functions. Thus, a significant increase in the quality of calculated shielding values is achieved without an appreciable increase in disk storage or computation time. Indeed, from a practical standpoint, the extremely small rmsds we find now approach the uncertainties in experimental chemical shifts reported for proteins in solution, due to referencing uncertainties between different groups.[23] The successful prediction of 39 shielding tensor elements for two polar amino acids in zwitterionic crystal lattices strongly supports the utility of GIAO and CFP-GIAO methods for predicting chemical shifts in proteins[3-6] and, *inter alia*, the use of chemical shifts to predict or refine protein or peptide structure in solution,[8] or in the semisolid or solid state.

Acknowledgment. We are grateful to Professors Peter Pulay and James Hinton and Dr. Krzysztof Wolinski for providing us with their TEXAS90 program, to Professors C. J. and A. K. Jameson, for helpful discussions, to Dr. Scott Wilson, for providing L-threonine and L-tyrosine coordinate files, and to the reviewers for their helpful comments.

(23) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171.