Published in final edited form as:
J Comput Chem. 2014 February 5; 35(4): 309–312. doi:10.1002/jcc.23499.

Are accurate computations of the $^{13}C'$ shielding feasible at the DFT level of theory?

Jorge A. Vila[1,2,*], **Yelena A. Arnautova**[3], **Osvaldo A Martin**[1,2], and **Harold A. Scheraga**[2,*]
[1]IMASL-CONICET, Universidad Nacional de San Luis, Italia 1556, 5700-San Luis, Argentina
[2]Cornell University, Ithaca, NY, USA
[3]Molsoft L.L.C., 11199 Sorrento Valley Road, San Diego, CA 92121

Abstract

The goal of this study is twofold. First, to investigate the relative influence of the main structural factors affecting the computation of the $^{13}C'$ shielding, namely, the conformation of the residue itself and the next nearest-neighbor effects. Second, to determine whether calculation of the $^{13}C'$ shielding at the DFT level of theory, with an accuracy similar to that of the $^{13}C^{\alpha}$ shielding, is feasible with the existing computational resources. The DFT calculations, carried out for a large number of possible conformations of the tripeptide Ac-G**XY**-NMe, with different combinations of **X** and **Y** residues, enable us to conclude that the accurate computation of the $^{13}C'$ shielding for a given residue **X** depends on the: (*i*) (ϕ,ψ) backbone torsional angles of **X;** (*ii*) side-chain conformation of **X;** (*iii*) (ϕ,ψ) torsional angles of **Y;** and (*iv*) identity of residue **Y**. Consequently, DFT-based quantum mechanical calculations of the $^{13}C'$ shielding, with *all* these factors taken into account, are two orders of magnitude more CPU demanding than the computation, with similar accuracy, of the $^{13}C^{\alpha}$ shielding. Despite not considering the effect of the possible hydrogen bond interaction of the carbonyl oxygen, this work contributes to our general understanding of the main structural factors affecting the accurate computation of the $^{13}C'$ shielding in proteins and may spur significant progress in effort to develop new validation methods for protein structures.

Introduction

The influence of different factors, such as the conformation of the residue itself and the identity of the next-nearest neighbors, on the computation of the $^{13}C'$ shielding has been discussed by both Xu and Case [1] and Han et al.,[2] with a ranking of the influence of these factors being provided by Han et al.[2] However, the information provided by these two publications is not sufficient to decide whether the $^{13}C'$ shielding can be computed, at the DFT level of theory, within an accuracy of ~0.5 ppm that was obtained for the $^{13}C^{\alpha}$ nucleus,[3,4] for which the shielding is determined mainly by the residue itself without significant influence of the nearest neighbors except for residues preceding proline.[4] Here, interest is centered on determining the relative influence of those factors that affect the accuracy of the computation of the $^{13}C'$ shielding and, consequently, whether the $^{13}C'$ shielding can be computed, at the DFT level of theory, with an accuracy similar to that for the $^{13}C^{\alpha}$ shielding with existing computational resources.

To accomplish the goal of this study efficiently, a terminally-blocked tripeptide with the sequence Ac-G**XY**-NMe, with **X** and **Y** representing any of the 20 naturally occurring amino acids, was considered. Shielding calculations, at the DFT level, were carried out on a

*Corresponding authors has5@cornell.edu.

large number of possible conformations for different combinations of **X** and **Y** residues. This approach enabled us to determine the extent to which the $^{13}C'$ shielding can be computed for a given residue **X** within approximately ~0.5 ppm. The plausible dependence of the computed $^{13}C'$ shielding on the carbonyl oxygen involved in hydrogen bonds is not investigated here.

Materials and Methods

The $^{13}C'$ isotropic shielding value (σ) for each amino acid residue **X**, in a terminally-blocked tripeptide with the sequence Ac-G**XY**-NMe, was computed at the OB98/6-311+G(2d,p) level of theory with the Gaussian 03 package.[5] The remaining residues in each tripeptide were treated at the OB98/3-21G level of theory, i.e., by using the *locally-dense basis set* approach [6].

The number of conformations and residue-dependent effects required to test the structural factors affecting the computation of the $^{13}C'$ shielding, at the DFT-level of theory, are very large and, hence, very CPU-time demanding. For this reason, we focus our analyses on only the following residues: Thr, Asp, Val, Met, Trp, Tyr, Gln, Pro and Gly, with Gly replaced by Ile, and Pro removed, for the side-chain effect analysis. This set of residues contains polar, non-polar, aromatic and ionizable residues and, hence, implicit in this choice of residues is the assumption that, if there is any dominant residue-dependent effect, then it can be generalized to all the remaining naturally-occurring amino acids. The ionizable residues (Asp and Tyr) were considered neutral during the quantum chemical calculations.

To study the effect of the nature of residue **Y** on the $^{13}C'$ shielding of **X,** with X=**A** (Ala), in the terminally-blocked tripeptide Ac-G**AY**-NMe, the backbone torsional angles of **A** were sampled every 10°, the backbone torsional angles $(\phi,\psi)_{\mathbf{Y}}$ of every residue **Y** were fixed at a canonical α–helix conformation with $\phi = -60°$ and $\psi = -40°$, and the Ac-G and **Y**-NMe torsional angles were free to vary.

To study the side-chain effect of residue **X** on its $^{13}C'$ shielding, the terminally-blocked tripeptide Ac-G**XA**-NMe was used, with A=Ala. The backbone torsional angles of residue **X** were sampled every 10° and the backbone torsional angles (ϕ,ψ) of residue **A** were fixed at an extended conformation with $\phi = -140°$ and $\psi = +40°$. For each amino acid **X,** with a given fixed backbone torsional angle, three side-chain rotamers of χ^1 were considered, namely −60°, + 60° and 180° with no restrictions on the remaining side-chain torsional angles, viz., χ^2, χ^3, etc, of residue **X**.

The dependence of the $^{13}C'$ shielding of residue **X** on the backbone torsional angles of **X** and **Y**, was studied using the terminally-blocked tripeptide Ac-G$\mathbf{A_X A_Y}$-NMe; to avoid side-chain effects, only alanine was considered as **X** and **Y** residues. The (ϕ,ψ) angles of residues $\mathbf{A_X}$ and $\mathbf{A_Y}$ were sampled every 20°, while the Ac-G and **Y**-NMe torsional angles were free to vary. With one example, we illustrate the procedure used to examine the relative sensitivity of the computed $^{13}C'$ shielding of residue $\mathbf{A_X}$ to the backbone torsional-angle variations of $\mathbf{A_X}$ and $\mathbf{A_Y}$. Once the torsional angles $(\psi)_X$, $(\phi)_Y$ and $(\psi)_Y$ are fixed, the variance, of the $^{13}C'$ shielding of residue A_x, was computed for a set of $(\phi)_X$ torsional angles sampled every 20°. This procedure is repeated for different combinations of fixed $(\psi)_X$, $(\phi)_Y$ and $(\psi)_Y$ torsional angles, i.e., ~300 combinations were analyzed. This total number of torsional-angle combinations is lower than are actually possible (~18^3) because we limit the sampling to the most populated regions of the Ramachandran map, namely for the α-helical and the extended region, respectively. Finally, the averaged standard deviation, $<\sigma_{\phi_X}>$, is calculated as the square root of the sum over all the weighted variances. A weight factor is essential for an accurate estimation of the averaged standard deviation because the total

number of possible combinations of the $(\phi)_X$, $(\psi)_X$, $(\phi)_Y$ and $(\psi)_Y$ torsional angles is not a fixed number, e.g., it varies because of atomic overlapping. Overall, a comparison between averaged standard deviations, computed for each torsional angle as described above for $<\sigma_{\phi X}>$, enabled us to rank the relative sensitivity of the $^{13}C'$ shielding of residue $\mathbf{A_x}$ to the backbone torsional-angle variations of $\mathbf{A_X}$ and $\mathbf{A_Y}$.

To avoid the referencing problem, we decided to compute shielding differences (Δ) rather than chemical-shift differences. Thus, for example, to study the side-chain effect of residue **X** on its $^{13}C'$ shielding (see Table 1), the Δ values were computed as: $\Delta = (^{13}C'_{\mathbf{A}} - {}^{13}C'_{\mathbf{Xn}})$ where $^{13}C'_{\mathbf{A}}$ is the isotropic shielding value of residue **A** (Ala) in the tripeptide Ac-G**A**A-NMe, and $^{13}C'_{\mathbf{X}\boldsymbol{n}}$ is the isotropic shielding value of residue **X,** in the tripeptide Ac-G**X**A-NMe, with the side-chain torsional angle χ^1 of **X** fixed at any rotamer ***n***, with ***n*** = −60°, +60° and 180°. If the second-order difference between |Δ |'s of two rotamers is ≤ 0.5 ppm, for any residue **X** other than Ala, Gly or Pro, there is no need to consider the side-chain effect for an accurate computation of the $^{13}C'$ shielding for this residue. The adopted cutoff difference value, namely Δ ≤ 0.5 ppm, was chosen to be comparable to the average difference observed from nearest-neighbor effects on computed $^{13}C^{\alpha}$ shieldings. [4]

Results and Discussion

The results indicate that: (*i*) proper consideration of at least the χ^1 side-chain torsional-angle variations of residue **X** is crucial for an accurate computation of $^{13}C'$ shielding of residue **X** (see Table 1); (*ii*) the identity of residue **Y** could give rise to significant differences, > 0.5 ppm, in the computed $^{13}C'$ shielding of residue **X** (see Table 2); and (*iii*) the sensitivity of the computed $^{13}C'$ shielding of residue $\mathbf{A_X}$**,** to the variation of the backbone torsional angles of residues $\mathbf{A_X}$ and $\mathbf{A_Y}$, can be ranked as: $(\phi)_{\mathbf{X}} \approx (\psi)_{\mathbf{X}} > (\phi)_{\mathbf{Y}} \gg (\psi)_{\mathbf{Y}}$. This backbone torsional angle ranking is based on the relative magnitudes of the computed averaged standard deviations (as explained in the Materials and Methods section), i.e., with $<\sigma_{\phi X}> = 1.4$; $<\sigma_{\psi X}> = 1.2$; $<\sigma_{\phi Y}> = 0.9$; and $<\sigma_{\psi Y}> = 0.3$.

An analysis of the structural factors that most influence the ranking of the computed $^{13}C'$ shielding of residue $\mathbf{A_X}$, by Han et al., [2] shows a different trend for the relative influence of the backbone torsional angles, namely $(\psi)_{\mathbf{X}} > (\psi)_{\mathbf{Y}} > (\phi)_{\mathbf{X}} > (\phi)_{\mathbf{Y}}$, i.e., from the relative percentage of influence,[2] namely 13.9%, 8.6%, 5.8% and 3.6%, respectively. The most striking difference, between Hans *et al*. [2] and our ranking, happens for the torsional angle $(\psi)_{\mathbf{Y}}$ which is listed as the second most important by Han *et al*.,[2] while it is the least important, in our ranking. Because the $^{13}C'$ nucleus of $\mathbf{A_X}$ is three bonds away from the $(\psi)_{\mathbf{Y}}$ torsional angle, rather than two or one from any other torsional angle, namely from $(\phi)_{\mathbf{X}}$, $(\phi)_{\mathbf{Y}}$ or $(\psi)_{\mathbf{X}}$, respectively, a larger response of the $^{13}C'$ shielding variations to $(\psi)_{\mathbf{X}}$, $(\phi)_{\mathbf{X}}$ and $(\phi)_{\mathbf{Y}}$, rather than of the $^{13}C'$ shielding variation to $(\psi)_{\mathbf{Y}}$, is expected and consistent with our results. As to whether the relative influence of the backbone torsional angles is biased by the presence of carbonyl oxygen hydrogen bonding, not considered in our analysis, remains to be investigated.

An accurate prediction of the $^{13}C'$ shielding of residue **X**, in the tripeptide Ac-G**XY**-NMe, would require generation and computation, at the DFT-level of theory, of ~10,000,000 conformations for *each* of the 20 naturally occurring amino acids, about ~200 times larger than the number of conformations computed for each residue for the $^{13}C^{\alpha}$ chemical-shift database [3]. This large number of conformations needed to compile a detailed database of the $^{13}C'$ shieldings, was estimated based on the above analysis of the factors affecting the computation of the $^{13}C'$ shielding, assuming that the backbone torsional angles $(\phi)_{\mathbf{X}}$ and $(\psi)_{\mathbf{X}}$ are sampled every 10°, $(\phi)_{\mathbf{Y}}$ every 30°, and $(\psi)_{\mathbf{Y}}$ every 60°; and the side-chain torsional angle χ^1 every 60°, except for **X** = Pro, Gly and Ala. Finally, the resulting number

of conformations should be multiplied by 20, to take the nature of the next nearest-neighbor residue **Y** into account. In this estimation of the number of conformations, the assumption is made that the χ's of residue **Y**, except for **Y** = Pro, Ala and Gly, and the (φ,ψ) values of residue Gly (located at the N-terminus of the tripeptide) are free to vary. The fact that the $^{13}C'$ shielding of a residue could also be influenced by hydrogen bond formation of the carbonyl oxygen only exacerbates the problem.

A ranking among seven different state-of-the-art servers,[2] based on their ability to predict the observed chemical shifts in a set of 61 high-quality protein structures, shows that the lowest correlation coefficient between the observed and predicted chemical shifts, among all heavy nuclei, occurs for the $^{13}C'$ nucleus. Conceivably, a proper consideration of the factors, and their order, that influence the $^{13}C'$ shielding is the origin of the problem.

Conclusions

The DFT-based quantum mechanical calculations of the $^{13}C'$ shielding, with the relative influence of the above-mentioned structural factors taken into account properly, are two orders of magnitude more demanding of CPU time than the computation, with similar accuracy, of the $^{13}C^{\alpha}$ shielding and, hence, difficult to achieve in a reasonable amount of time with existing computational power. However, recent progress in the development of a powerful new type of quantum computers[7] opens a new venue to solve scientific problems several orders of magnitude faster than can be done today. This also means that other effects not analyzed here, such as carbonyl-oxygen hydrogen bonding, could also be considered without exacerbating the computational problem.

Overall, an accurate computation of the $^{13}C'$ shielding, and, hence, the development of new and more accurate *physics-based* validation methods for protein structures, may be possible in the near future.

Acknowledgments

This work was supported by grant GM14312 (HAS) from the National Institutes of Health, USA, PIP-112-2011-0100030 (JAV) from IMASL-CONICET, Argentina, and Project 328402 (JAV) from UNSL, Argentina. The research was conducted by using the resources of Blacklight, a facility of the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center.

References

1. Xu X-P, Case DA. Biopolymers. 2002; 65:408–423. [PubMed: 12434429]
2. Han B, Liu Y, Ginzinger SW, Wishart DS. J Biomol NMR. 2011; 50:43–57. [PubMed: 21448735]
3. Vila JA, Arnautova YA, Martin OA, Scheraga HA. Proc Natl Acad Sci U S A. 2009; 106:16972–7. [PubMed: 19805131]
4. Vila JA, Serrano P, Wüthrich K, Scheraga HA. J Biomol NMR. 2010; 48:23–30. [PubMed: 20644980]
5. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery, J., et al. Gaussian 03, Revision E.01. Inc. Wallingford CT: 2004.
6. Chesnut DB, Moore KD. J Comput Chem. 1989; 10:648–659.
7. Kassal I, Whitfield JD, Perdomo-Ortiz A, Yung M-H, Aspuru-Guzik A. Annu Rev Phys Chem. 2011; 62:185–207. [PubMed: 21166541]

Table 1

Effect of the side chain of residue **X** on computed $^{13}C'$ shielding

X[a]	Δ[b] (ppm)
Thr	4.3
	1.9
	1.5
Asp	0.9
	1.1
	1.6
Val	0.6
	0.9
	2.7
Met	0.7
	0.3
	1.6
Trp	2.3
	0.6
	0.9
Tyr	1.2
	−0.4
	1.0
Gln	1.5
	0.5
	1.5
Ile	3.2
	1.5
	1.8

[a] Results of the computed $^{13}C'$ shielding for residue **X** in the tripeptide: Ac-G**XA**-NMe, for **X** in column 1. All the listed results were obtained by assuming that the backbone torsional angles (φ,ψ) of residue **A** were fixed (as indicated in the Materials and Methods section) at an extended conformation, namely φ = −140° and ψ = +140°. However, among all the possible values of the backbone torsional angles (φ,ψ) of residue **X**, sampled every 10°, the set φ = −60° and ψ = −30° was chosen because a computed $^{13}C'$ shielding value exists for *all* the listed residue**s** for this particular set of backbone torsional angles.

[b] For each residue **X** there are three Δ values (see Methods section); each of these values corresponds to a difference, Δ computed by using a given χ^1 side-chain rotamer of **X**, namely −180°, −60° or +60°. For all the listed residues there is, at least, one second-order difference, between Δ values, > 0.5 ppm, indicating the need to consider the side-chain effect for an accurate computation of the $^{13}C'$ shielding of residue **X**, as mentioned in the Materials and Methods section.

Table 2

Effect of the nature of residue Y on computed $^{13}C'$ shielding of **A**

Ac-GAY-NMe[a]			
Y[b]	$\Delta_{(\phi = -60;\ \psi = -40)}$[c]	$\Delta_{(\phi = -60;\ \psi = -60)}$[c]	$\Delta_{(\phi = -140;\ \psi = +140)}$[c]
Thr	***1.0***	***1.7***	***0.6***
Asp	***1.4***	***0.9***	−0.3
Val	***0.7***	***1.5***	0.3
Met	***0.7***	***1.1***	0.1
Trp	0.4	***0.7***	0.1
Tyr	0.2	0.2	−0.2 (***0.8***)[d]
Gln	***−1.0***	−0.5	−0.3
Pro	***−2.3***	***1.1***	***−1.1***
Gly	***0.6***	***1.2***	−0.1

[a] All the listed results, in terms of Δ, were obtained assuming that the backbone torsional angles (φ,ψ)**Y** of residue Y are fixed at a canonical α–helix conformation, namely φ = −60° and ψ = −40°. The Δ values were computed as: $\Delta = (^{13}C'\mathbf{A} - ^{13}C'\mathbf{Y})$ where $^{13}C'\mathbf{A}$ is the isotropic shielding value of residue **A** (Ala) in the tripeptide Ac-G**A**A-NMe, and $^{13}C'\mathbf{Y}$ is the isotropic shielding value of residue **A** in the tripeptide Ac-G**A**Y-NMe, with the identity of residue Y listed in column 1.

[b] Identity (by using a three letter code) of residue **Y** in the tripeptide Ac-GA**Y**-NMe.

[c] The sub index of Δ represents the particular backbone (φ,ψ) torsional angles chosen for the residue **A**, among all possible ones from the Ramachandran map. Those values for which |Δ| > 0.5 ppm are highlighted in boldface and italics.

[d] In parentheses, an alternative value for Δ, (0.8), was computed for a slightly-shifted set of backbone torsional angles, namely (φ = −40°; ψ = +20°) rather than (φ = −40°; ψ = +40°); this result illustrates that the same absolute value of |Δ| computed for Tyr (0.2), in each column of this Table, is just a coincidence, and that the nature of residue Y, matters.