Published in final edited form as:

J Phys Chem B. 2020 February 06; 124(5): 735–741. doi:10.1021/acs.jpcb.9b10123.

Assessing the One-Bond $C_\alpha - H$ Spin-Spin Coupling Constants in Proteins: Pros and Cons of Different Approaches

Agustina Arroyuelo[†], Osvaldo A. Martin[†], Harold A. Scheraga[‡], Jorge A. Vila^{†,‡}

[†]IMASL-CONICET, Universidad Nacional de San Luis, Ejército de Los Andes 950, 5700-San Luis, Argentina.

[‡]Baker Laboratory of Chemistry, Cornell University, Ithaca, NY, USA.

Abstract

In the present work we explore three different approaches for the computation of the one-bond spin-spin coupling constants (SSCC) $^1J_{CaH}$ in proteins: DFT calculation, Karplus-like equation and Gaussian Process regression. The main motivation of this work is to select the best method for a fast and accurate computation of the $^1J_{CaH}$ SSCC, for its use on everyday applications in protein structure validation, refinement and/or determination. Our initial results showed a poor agreement between the DFT computed and observed $^1J_{CaH}$ SSCC values. Further analysis lead us to the understanding that the model chosen for the DFT computations is inappropriate, and that more complex models will require a higher if not prohibitively computational cost. Finally, we show that Karplus-like equation and Gaussian Process regression provide faster and more accurate results than DFT-based calculations.

Graphical Abstract

Introduction

There is an enduring interest in the protein NMR spectroscopy field for collecting a larger number of observables in order to increase the accuracy of the NMR-determined protein

jv84@cornell.edu.

Supporting Information Available

Supporting information contains details of the parameters used in the DFT computation of $^1J_{CaH}$ SSCC. Training RMSD for the gaussian process regression is also reported.

models. Among all these observables, it has been recognized that the one-bond Carbon-Hydrogen, $^1J_{CaH}$ Spin-Spin Coupling Constants (SSCC) can be used: (i) as a conformational constraint on cyclic peptides;¹ (ii) to resolve the ambiguities caused by the non-uniqueness of the chemical shifts;² (iii) to detect sparsely populated, short-lived, protein states;³ (iv) to re-parameterize Karplus-like equations;⁴ and (v) for a determination of the tautomeric fractions of the imidazole ring of His.⁵

Analysis of the factors affecting the computation of $^1J_{CaH}$ SSCC, at different DFT-levels of theory, has been performed on 68 small organic compounds by using 35 combinations of functionals and basis sets.⁶ As a result, these authors recommend the use of the B3P86 functional in combination with the aug-cc-pVTZ-J basis set together with a 5 Hz contribution to account for the Zero Point Vibrational Contribution. Notably, the use of the popular B3LYP functional,⁷ which overestimates the computed vicinal coupling constants, yields good agreement with the observed values if it is combined with a small (TZVP) basis set. However, this agreement is attributed to error cancellation.⁸

Additionally, a statistical study of $^1J_{SSCC}$ values collected from 6 proteins was used to determine the main factors affecting the one-bond couplings involving the Ca nucleus.⁹ According to this study the main factors, in descending order, affecting $^1J_{SSCC}$ in proteins are: "... ϕ and ψ torsion angles, substituents effects, $\chi 1$ and $\chi 2$ torsional angles, tetrahedral distortion at Ca, and noise..." although, as the same authors noted, "...for practical purposes, dependence of... $^1J_{CaH}$ couplings on Ca bond-angle geometry can be considered insignificant...". These conclusions are particularly interesting and, certainly, deserve a theoretical investigation.

We begin our analysis by first studying the factors that influence the DFT computations of $^1J_{CaH}$ SSCC; secondly, we describe the method for building a DFT computed $^1J_{CaH}$ SSCC values look-up table. Finally, we address the performance of two empirical predictors: a Karplus-like equation and a fully Bayesian Gaussian Process regression and discuss the shortcomings of the DFT approach.

Methods

In this section we present the analysis made to study factors influencing the DFT computation of $^1J_{CaH}$ SSCC values. We also provide details on the application of the three methods compared in this work.

DFT computation

The computation of the $^1J_{CaH}$ SSCC values was carried out using terminally-blocked tripeptides with the sequence *Ace* – *Gly* – *Xxx* – *Gly* – *NMe*, otherwise noted. Where *Xxx* represents either a residue in the conformation of an experimentally determined protein structure or a regularized residue.

Solvent effects are important for the computation of $^1J_{CaH}$ SSCC.¹⁰ Consequently, dielectric solvent effects were taken into account during the DFT-calculations by using a Polarizable Continuum Model (PCM).^{11,12}

During the computation of the $^1J_{CaH}$ SSCC each Xxx residue was treated with a 6–31G basis set and the remaining ones with a 3–21G basis set; in addition, the B3P96 functional and $Di = 25.5$ was used. For details about the chosen parameters for the DFT-calculations, namely, the dielectric constant, basis set and functional see sections Factors Affecting the Computation of the $^1J_{CaH}$ SSCC and Analysis of the results on the basis set and functionals, in the SI.

Calculations were carried out using the Gaussian package.¹³

Evaluation of torsional angle dependence of the computed $^1J_{CaH}$ SSCC—In order to assess the torsional angle dependence, $^1J_{CaH}$ SSCC values were computed over a 10° grid in the range from -180° to 180° , for each torsional angle in terminally blocked tripeptides (see DFT computation) with Xxx = phenylalanine (Phe), histidine (His), threonine (Thr) or proline (Pro). While the remaining two torsional angles not being evaluated were fixed on values: $\psi = -40^\circ$, $\phi = -60^\circ$, $\chi_1 = -60^\circ$.

Look up table of the DFT computed $^1J_{CaH}$ SSCC—The look-up table was constructed by firstly sampling every 10° , for the ϕ and ψ torsional angles, and every 30° for χ_1 , for the model tripeptides (see DFT computation) where Xxx = any of the 19 most abundant amino acids in proteins, as Gly was excluded. Only 36 samples were generated for proline as the ϕ torsional angle was kept fixed in -60° . Secondly, the most probable conformations according to the Ramachandran map, were selected.

To evaluate the DFT computed $^1J_{CaH}$ SSCC look-up table, the observed $^1J_{CaH}$ SSCC values from 6 X-ray determined protein structures were used, namely, Flavodoxin from *Desulfovibrio vulgaris* (PDB id.: 1j8q), Ubiquitin (PDB id.: 1ubq), *Bacillus agaradhaerens* xylanase (PDB id.: 1qh7), Human frataxin (PDB id.: 1ekg), Ribonuclease T1 (PDB id.: 9rnt) and Squid ganglion DFPase (PDB id.: 1pjj).² Our choice of these 6 proteins is due to availability of observed $^1J_{CaH}$ SSCC data.

DFT computed $^1J_{CaH}$ SSCC from an experimentally determined hexapeptide versus tripeptide model—A portion of the helix sequence of protein 9rnt was selected to build an hexapeptide of the form $Ace - Thr_{18} - Ala_{19} - Gln_{20} - Ala_{21} - Ala_{22} - Gly_{23} - NMe$ (see Figure S1 in SI), where the sub index indicates the residue number in the sequence of 9rnt. The DFT computed $^1J_{CaH}$ SSCC values for Ala_{19} , Gln_{20} , Ala_{21} and Ala_{22} in the hexapeptide were then compared with the ones obtained from a tripeptide with Xxx = Ala_{19} , Gln_{20} , Ala_{21} or Ala_{22} . To avoid any influence of bond-length and bond-angle variations, the hexapeptide and tripeptides were regularized,¹⁴ i.e., the residues were replaced by the standard ECEPP/3 residues^{15,16} in which bond lengths and bond angles are fixed at the standard values and hydrogen atoms were added, if necessary. Later (see next section) we decided to test the bond-length variations on the computed $^1J_{CaH}$ SSCC.

Dependence of the DFT computed $^1J_{CaH}$ SSCC value with the bond-angle and bond-length—Schmidt et. al.⁹ proved that the influence of bond-angle variations on the $^1J_{CaH}$ SSCC is insignificant. In order to establish the influence of bond-length variations we analyzed the $Ca - H$ bond-length over every amino acid residue in the NMR structure of

protein ubiquitin (PDB id 1d3z).¹⁷ The resulting $C\alpha - H$ bond-length mean-value ($1.120 \pm 0.004 \text{ \AA}$) is in line with the relative $C\alpha - H$ bond-length distance ($1.117 \pm 0.007 \text{ \AA}$) determined, also for ubiquitin, by Ottiger & Bax.¹⁸ In order to test the effect of $C\alpha - H$ bond-length variations on the computation of the $^1J_{CaH}$ SSCC, the residue Ala_{28} in ten ubiquitin conformations (PDB id 1D3Z), was chosen to generate 10 terminally-blocked tripeptides with the sequence $Ace - Gly - Ala_{28} - Gly - NMe$. The residue Ala_{28} was selected because it pertains to the α -helix region of the protein and, therefore it is unlikely to show significant backbone torsional angle (ψ , ϕ) variations among the 10 conformers.

Gaussian Process regression

Gaussian Processes (GPs) are very flexible statistical models that can be used to perform Bayesian non-linear regression. They have robust theoretical properties and also have proven to be effective in different applied domains.^{19,20}

GPs were used to learn the non-linear function between the observed $^1J_{CaH}$ SSCC values and the protein torsional angles ψ , ϕ and/or χ_1 . A squared exponential kernel was chosen and the PyMC3 package was used to estimate the hyper-parameters of this kernel.²¹

GP regression was performed in a k-fold cross validation setting. K-fold cross validation means that each testing data fold was from a single protein while the remaining proteins data was used for training. Given the poor results already observed for Karplus-like equation and DFT calculations for proteins 1p1x and 1qh7, said proteins data was not used in the training set, only in the testing set, while the GP was trained using the remaining 4 proteins data.

Results of the GP regression are reported in terms of *testing RMSD* and *training RMSD*. The former is the Root Mean Squared Deviation between the learned function and the data set that was not used for training. The latter is the Root Mean Squared Deviation between the learned function and the data set that was used for training.

Karplus-like equation

Predicted $^1J_{CaH}$ SSCC values were computed with the following equation:

$$^1J_{CaH} = A + B \sin(\psi + \psi') + C \cos(2(\psi + \psi')) + D \cos(2(\phi + \phi')) + ^1J_{CaH - random} \quad (1)$$

where the parameters $A=1.7$, $B=1.4$, $C=4.1$, $D=1.7$, $\psi'=138$, $\phi'=30$ and $^1J_{CaH-random}$ were taken from Vuister et al.²²

Results

Torsional angle dependence of the computed $^1J_{CaH}$ SSCC

We observe that for every analyzed residue the main influence on the $^1J_{CaH}$ SSCC is by the ψ torsional angle. Results obtained for Thr show a dependence of the computed $^1J_{CaH}$ SSCC with χ_1 similar to that on ϕ torsional angle variations (see Figure 1a). Such dependence was not observed in neither His nor Phe. In particular, the results for Pro (Figure 1b) indicate that the DFT computed $^1J_{CaH}$ SSCC are very sensitive to both ϕ and ψ torsional

angle variations as well as to the cis/trans isomerization of the ω torsional angle and, to a lesser extent, to the up/down puckering of the proline side-chain.

Performance of the of $^1J_{CaH}$ SSCC lookup table

The observed $^1J_{CaH}$ SSCC values from 6 proteins, PDB id 1j8q, 1ubq, 1qh7, 1ekg, 9rnt and 1pjj, were used to validate the prediction obtained using the look up table of the computed $^1J_{CaH}$ SSCC. The results of the comparison between observed and computed $^1J_{CaH}$ SSCC values for each of these 6 proteins are shown in Figure S2 of SI.

A distribution of the observed $^1J_{CaH}$ SSCC values in the Ramachandran map is shown in Figure 2a, while the DFT computed values are shown on Figure 2b. It is worth noting that DFT computed $^1J_{CaH}$ values around the α -helical region are systematically the largest ones in the Ramachandran-map. This is not seen in the distribution of observed $^1J_{CaH}$ SSCC values.

Comparison between the DFT obtained $^1J_{CaH}$ SSCC for the tripeptide and hexapeptide models

A comparison between the DFT computed $^1J_{CaH}$ SSCC values for an experimentally determined hexapeptide and a regularized tripeptide is shown in Figure 3. Certainly, the DFT computed $^1J_{CaH}$ SSCC values from the tripeptide does not agree with those computed from the hexapeptide for any residue, except for *Gln*₂₀.

Bond-length dependence of the $^1J_{CaH}$ SSCC computed value

Results of the analysis made to assess the bond-length dependence of the DFT computed $^1J_{CaH}$ SSCC show a mean value of 157.962 Hz, with a standard deviation of 0.384 (see Table S1 of SI). This standard deviation indicates that the DFT computed $^1J_{CaH}$ SSCC values show barely visible changes, within the observed range of variations for the $C_\alpha - H$ bond-length in ubiquitin, namely 1.1200 ± 0.004 Å.

Gaussian Process regression and Karplus-like equation

As a result of the GP regression we obtained a mean function to compute the theoretical $^1J_{CaH}$ SSCC for each of the 6 following proteins, PDB id 1j8q, 1ubq, 1qh7, 1ekg, 9rnt and 1pjj (see Figure S3 of SI). The results are summarized in Table 1 as the testing RMSD between GP predicted $^1J_{CaH}$ SSCC values and the experimental ones (see Table S2 of SI for training RMSD). The reported RMSD were obtained by cross-validation, where the data of the protein indicated in each column name was used as test set (see Gaussian Process regression). We can observe the relevance, in terms of RMSD, of GP predicting the $^1J_{CaH}$ SSCC values from the torsional angles ψ , ϕ and/or χ . These results conform to experiments as well as to DFT computed evidence showing that variations of ψ capture most of the torsional angle dependence of the observed $^1J_{CaH}$ SSCC values.²² The RMSD between experimental and theoretical $^1J_{CaH}$ SSCC values, obtained using a Karplus-like equation, are shown in parentheses in Table 1.

Figure 4 shows probability densities of the differences between computed and experimental $^1J_{CaH}$ SSCC values for the three applied methods. For the six analyzed proteins, the mean of

the differences from the GP method are closer to zero and with lower variance than the differences computed using DFT. It is worth noting that DFT computed values show a systematic error of approximately -10 Hz, and thus mean predicted differences could be easily corrected by adding a constant value. Nevertheless, the variance remains higher for the DFT computed values compared to the other two methods. On the other hand, for proteins 1pjj and 1qh7, every approach applied to the computation of the $^1J_{CaH}$ SSCC is almost equally deficient.

Conclusions

After assessing the variation of the DFT computed $^1J_{CaH}$ SSCC values with the $C_\alpha H$ bondlength, we conclude that bond-length variations do not significantly influence the computation of $^1J_{CaH}$ SSCC values.

From the analysis of torsional angle dependence of the DFT computed $^1J_{CaH}$ SSCC values for Thr, His and Phe (see Figure 1a), we conclude that the largest dependence of the DFT computed $^1J_{CaH}$ SSCC is with the variation of ψ , ϕ (in agreement with experimental evidence of Vuister et al., and ab initio calculations of Edison et al.²³) and χ_1 . Thus, the ranking of the torsional angle dependence is: $\psi \gg \phi > \chi_1$. This dependence of the DFT computed $^1J_{CaH}$ SSCC with the torsional angle variations is in line with the statistical analysis of over 3,000 observed $^1J_{CaH}$ SSCC from 6 proteins, carried out by Schmidt et al.⁹ Regarding Pro (see Figure 1b), besides the already stated torsional angle influence, the cis/trans isomerization also impacts on the DFT computed $^1J_{CaH}$ SSCC value and, to a lesser extent, the up/down puckering of the proline side-chain.

As for the performance of the constructed look-up table, the agreement between observed $^1J_{CaH}$ SSCC values for 6 proteins and the DFT computed ones (see Figure S2 in SI), is subpar in every case (all RMSD > 2.44 Hz). The clear-cut separation in the Ramachandran map corresponding to higher and lower values of $^1J_{CaH}$ SSCC, between the α -helix region and the remaining regions observed in Figure 2b, is not present in the distribution of observed data (Figure 2a) indicating a problem with the methodology used for the DFT computation of the $^1J_{CaH}$ SSCC. Indeed, a comparison of the computed $^1J_{CaH}$ SSCC values on the hexapeptide and the tripeptide (see Figure 3) indicates that the tripeptide is a poor approximation to reproduce values of each residue in the hexapeptide, except for *Gln*₂₀. This results indicate that the DFT computed $^1J_{CaH}$ SSCC values are very sensitive to the environment of the peptide-model used to compute it, making an accurate prediction of their values in proteins a daunting task. Especially when the motivation is to build lookup tables for fast numerical approximations of $^1J_{CaH}$ SSCC values, as was done for chemical shifts on proteins and glycans.^{24,25} Doing something similar with $^1J_{CaH}$ SSCC will require a more complex model including non-covalent effects of the environment. This approach would be computationally very costly but most importantly, difficult to generalize to the construction of lookup tables for fast computation. In conclusion, this problem pairs that encountered in the computation of $^{13}C^\beta$ Shielding.²⁶

For the above exposed reasons, exploring alternative methods is worthwhile. Gaussian process regression allowed us to obtain the nonlinear function between torsional angles and

$^1J_{CaH}$ SSCC values. We show that GPs can predict the observed $^1J_{CaH}$ SSCC values for 6 proteins, as a function of ψ , ϕ and/or χ_1 with testing RMSD values ranging from 1.47Hz to 10.51Hz (see Table 1). Analysis of GPs results, in terms of RMSD, shows that the torsional-angle dependence of the $^1J_{CaH}$ SSCC in proteins is captured mainly by the ψ torsional-angle with a lesser contribution by ϕ . This is in good agreement with the DFT calculations and experimental evidence from Vuister et al.²² Moreover, in the presence of ψ and ϕ , the addition of χ_1 does not substantially help to get better predictions.

Overall, the computed RMSD from both the GPs and the Karplus-like equation are similar, although the GPs are slightly better for 5 of 6 proteins (see Table 1). Nevertheless, the poor agreement (RMSD > 2.0 Hz) found for proteins 1p1x and 1qh7 by using the Karplus-like equation and GP regression cannot be attributed only to a poor or inadequate modeling, instead it points out to other effects, like problems during data acquisition or experimental conditions. Results regarding GP prediction of $^1J_{CaH}$ SSCC are very promising and this type of non-parametric approach will benefit greatly from new experimental data collection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This research was supported by CONICET-Argentina (PIP-0087), UNSL-Argentina (PROICO 3-2212), and ANPCyT-Argentina (PICT-0767) (J.A.V.) and (PICT-0218) (O.A.M.), and U.S. National Institutes of Health (GM-14312) (H.A.S.).

References

- (1). Egli H; Von Philipsborn W Conformational Dependence of One-Bond $C\alpha$, H Spin Coupling in Cyclic Peptides. *Helvetica Chimica Acta* 1981, 64, 976–988.
- (2). Zweckstetter M; Bax A Single-step determination of protein substructures using dipolar couplings: aid to structural genomics. *Journal of the American Chemical Society* 2001, 123, 9490–9491. [PubMed: 11562251]
- (3). Hansen AL; Kay LE Measurement of histidine pKa values and tautomer populations in invisible protein states. *Proceedings of the National Academy of Sciences* 2014, 111, E1705–E1712.
- (4). Gapsys V; Narayanan RL; Xiang S; de Groot BL; Zweckstetter M Improved validation of IDP ensembles by one-bond $C\alpha$ – $H\alpha$ scalar couplings. *Journal of biomolecular NMR* 2015, 63, 299–307. [PubMed: 26433382]
- (5). Sudmeier JL; Ash EL; Günther UL; Luo X; Bullock PA; Bachovchin WW HCN, a triple-resonance NMR technique for selective observation of histidine and tryptophan side chains in $^{13}C/^{15}N$ -labeled proteins. *Journal of Magnetic Resonance, Series B* 1996, 113, 236–247. [PubMed: 8995843]
- (6). San Fabián J; García de la Vega J; Suardiá R; Fernández-Oliva M; Pérez C; Crespo-Otero R; Contreras RH Computational NMR coupling constants: Shifting and scaling factors for evaluating $^1J_{CH}$. *Magnetic Resonance in Chemistry* 2013, 51, 775–787. [PubMed: 24123317]
- (7). Buevich AV; Sauri J; Parella T; De Tommasi N; Bifulco G; Williamson RT; Martin GE Enhancing the utility of $^1J_{CH}$ coupling constants in structural studies through optimized DFT analysis. *Chemical Communications* 2019, 55, 5781–5784. [PubMed: 31041954]
- (8). Ruden TA; Lutnæs OB; Helgaker T; Ruud K Vibrational corrections to indirect nuclear spin–spin coupling constants calculated by density-functional theory. *The Journal of chemical physics* 2003, 118, 9572–9581.

- (9). Schmidt JM; Howard MJ; Maestre-Martínez M; Pérez CS; Löhr F Variation in protein $C\alpha$ -related one-bond J couplings. *Magnetic Resonance in Chemistry* 2009, 47, 16–30. [PubMed: 18853398]
- (10). Ruud K; Frediani L; Cammi R; Mennucci B Solvent effects on the indirect spin– spin coupling constants of benzene: the DFT-PCM approach. *International Journal of Molecular Sciences* 2003, 4, 119–134.
- (11). Tomasi J; Persico M Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chemical Reviews* 1994, 94, 2027–2094.
- (12). Tomasi J Selected features of the polarizable continuum model for the representation of solvation. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2011, 1, 855–867.
- (13). Frisch MJ et al. Gaussian16 Revision C.01 2016; Gaussian Inc Wallingford CT.
- (14). Vila JA; Villegas ME; Baldoni HA; Scheraga HA Predicting ^{13}C α chemical shifts for validation of protein structures. *Journal of biomolecular NMR* 2007, 38, 221–235. [PubMed: 17558470]
- (15). Nemethy G; Gibson KD; Palmer KA; Yoon CN; Paterlini G; Zagari A; Rumsey S; Scheraga HA Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *The Journal of Physical Chemistry* 1992, 96, 6472–6484.
- (16). Momany F; McGuire RF; Burgess A; Scheraga HA Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *The Journal of Physical Chemistry* 1975, 79, 2361–2381.
- (17). Cornilescu G; Marquardt JL; Ottiger M; Bax A Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *Journal of the American Chemical Society* 1998, 120, 6836–6837.
- (18). Ottiger M; Bax A Determination of Relative N- HN, N- C, $C\alpha$ - C, and $C\alpha$ - $H\alpha$ Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society* 1998, 120, 12334–12341.
- (19). Williams CK; Rasmussen CE Gaussian processes for machine learning; MIT press Cambridge, MA, 2006; Vol. 2.
- (20). Martin O Bayesian Analysis with Python: Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ, 2nd Edition; Packt Publishing, 2018.
- (21). Salvatier J; Wiecki TV; Fonnesbeck C Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* 2016, 2, e55.
- (22). Vuister GW; Bax A Quantitative J correlation: a new approach for measuring homonuclear three-bond J (HNH. α .) coupling constants in ^{15}N -enriched proteins. *Journal of the American Chemical Society* 1993, 115, 7772–7777.
- (23). Edison AS; Markley JL; Weinhold F Calculations of one-, two-and three-bond nuclear spin-spin couplings in a model peptide and correlations with experimental data. *Journal of biomolecular NMR* 1994, 4, 519–542. [PubMed: 8075539]
- (24). Garay PG; Vila JA; Martin OA CheSweet: An application to predict glycan's chemicals shifts. *J. Open Source Software* 2018, 3, 488.
- (25). Martin OA; Vila JA; Scheraga HA Che Shift-2: graphic validation of protein structures. *Bioinformatics* 2012, 28, 1538–1539. [PubMed: 22495749]
- (26). Vila JA; Arnautova YA; Martin OA; Scheraga HA Are accurate computations of the $^{13}\text{C}'$ shielding feasible at the DFT level of theory? *Journal of computational chemistry* 2014, 35, 309–312. [PubMed: 24403017]

Figure 1:

(a) Torsional angle dependence of the DFT computed $^1J_{CaH}$ SSCC, on the model tripeptide *Ace-Gly-Xxx-Gly-NMe* with *Xxx* = threonine (Thr), phenylalanine (Phe) and histidine (His), as a function of ϕ , ψ and χ_1 torsional angles (see Methods section for details of the calculations). (b) Torsional angle dependence of the DFT computed $^1J_{CaH}$ SSCC, on the model tripeptide *Ace-Gly-Xxx-Gly-NMe* with *Xxx* = proline (Pro), as a function of ϕ and ψ torsional angles, the up or down pucker of the proline side-chain and cis/trans isomerization.

Figure 2:

Ramachandran plot of the $^1J_{CaH}$ SSCC distribution for the: (a) observed values from 6 proteins (PDB id 1j8q, 1ubq, 1qh7, 1ekg, 9rnt and 1pjax); and (b) DFT computed values for every naturally occurring amino acids, except Gly. The color bar references the range in (Hz) of DFT computed $^1J_{CaH}$ SSCC values. Indicating lower values in dark blue to higher values in yellow.

Figure 3:

Comparison of the DFT computed $^1J_{C\alpha H}$ SSCC values for Ala_{19} , Gln_{20} , Ala_{21} , Ala_{22} in the hexapeptide $Ace-Thr_{18}-Ala_{19}-Gln_{20}-Ala_{21}-Ala_{22}-Gly_{23}-NMe$ (blue- filled square) versus the ones computed in the tripeptide $Ace-Gly-Xxx-Gly-NMe$, with $Xxx = Ala_{19}$, Gln_{20} , Ala_{21} and Ala_{22} (red-filled circle); the results are plotted in decreasing order of the computed $^1J_{C\alpha H}$ SSCC values in the hexapeptide (blue-filled square).

Figure 4:
Kernel Density Estimation plots of the difference between experimental and computed $^1J_{C\alpha H}$ SSCC for DFT calculations (blue), Gaussian Process regression (green) and Karplus-like equation (orange).

Table 1:

Testing RMSD in Hz, for $^1J_{CaH}$ SSCC values predicted using a Gaussian Process regression. The first column indicates the number and type of torsional angles considered during the prediction of the $^1J_{CaH}$ SSCC values. Values in parentheses correspond to RMSD between observed and Karplus-like equation predicted $^1J_{CaH}$ SSCC values. Values in brackets correspond to RMSD between observed and DFT computed $^1J_{CaH}$ SSCC values.

	1ubq	9rnt	1ekg	1j8q	1p1x	1qh7
φ	2.20	2.95	2.96	2.73	2.88	10.51
ψ	1.66	2.33	2.69	2.25	3.11	3.27
χ	3.27	3.66	4.36	3.18	3.24	2.89
$\varphi-\psi$	1.47 (2.35)	2.41 (2.53)	2.64 (2.97)	2.21 (2.35)	3.02 (3.06)	6.28 (3.27)
$\varphi-\psi-\chi$	1.64 [9.60]	3.28 [7.68]	3.34 [8.36]	2.06 [7.90]	3.43 [9.69]	6.85 [9.23]