Limiting Values of the one-bond C–H Spin-Spin Coupling Constants of the Imidazole Ring of Histidine at High-pH

Jorge A. Vila^{1,*} and Harold A. Scheraga²

¹IMASL-CONICET, Universidad Nacional de San Luis, Ejército de Los Andes 950, 5700-San Luis, Argentina; ²Baker Laboratory of Chemistry, Cornell University, Ithaca, NY, USA.

* Corresponding author

E mail address: vila@unsl.edu.ar

Keywords: SSCC; one-bond spin-spin coupling constant; high-pH limiting values; Histidine tautomers; protonated form of His; solvent effects.

Summary

Assessment of the relative amounts of the forms of the imidazole ring of Histidine (His), namely the protonated (H^+) and the tautomeric $\text{N}^{\epsilon 2}\text{-H}$ and $\text{N}^{\delta 1}\text{-H}$ forms, respectively, is a challenging task in NMR spectroscopy. Indeed, their determination by direct observation of the ^{15}N and ^{13}C chemical shifts or the one-bond C-H , $^1J_{\text{CH}}$, Spin-Spin Coupling Constants (SSCC) requires knowledge of the “canonical” limiting values of these forms in which each one is present to the extent of 100%. In particular, at high-pH, an accurate determination of these “canonical” limiting values, at which the tautomeric forms of His coexist, is an elusive problem in NMR spectroscopy. Among different NMR-based approaches to treat this problem, we focus here on the computation, at the DFT level of theory, of the high-pH limiting value for the $^1J_{\text{CH}}$ SSCC of the imidazole ring of His. Solvent effects were considered by using the polarizable continuum model approach. The results of this computation suggest, first, that the value of $^1J_{\text{C}\epsilon 1\text{H}} = 205 \pm 1.0$ Hz should be adopted as the canonical high-pH limiting value for this SSCC; second, the variation of $^1J_{\text{C}\epsilon 1\text{H}}$ SSCC during tautomeric changes is minor, *i.e.*, within ± 1 Hz; and, finally, the value of $^1J_{\text{C}\delta 2\text{H}}$ SSCC upon tautomeric changes is large (15 Hz) indicating that, at high-pH or for non-protonated His at any pH, the tautomeric fractions of the imidazole ring of His can be predicted accurately as a function of the observed value of $^1J_{\text{C}\delta 2\text{H}}$ SSCC.

Introduction

The role of Histidine (His) in many biological functions and activities is very well documented [1-4], and the reason for such versatility can be found in three distinctive properties of the His amino acid residue: (i) existence of two neutral, chemically-distinct forms ($\text{N}^{\delta 1}\text{-H}$ and $\text{N}^{\epsilon 2}\text{-H}$ tautomers, also known as π and τ tautomers, respectively [5], and a charged H^+ form, shown in Figure 1), with one form favored over the other by the protein environment and pH [6]; (ii) the only ionizable residue (the charged form) that titrates around neutral pH has a pK^0 of 6.6 [7] and (iii) appearance of a population of ~50% in all enzyme active sites [8].

Despite these well-known facts, the physical properties of neutral His are difficult to characterize experimentally [9], making a proper determination of the fractions of the tautomeric forms of the imidazole ring of His a challenging problem in NMR spectroscopy. Among the experimental methods in current use, are those based on the observed: (a) ^{15}N chemical shifts [10-12]; (b) $^{13}\text{C}^{\gamma}$ and $^{13}\text{C}^{\delta 2}$ chemical shifts [6,13]; and (c) $^1J_{\text{CH}}$ Spin-Spin Coupling Constant (SSCC) of the imidazole ring of His [13-15]. As with any experimental method, *all* these mentioned approaches possess shortcomings: (i) the tautomeric fractions obtained by the ^{15}N -based method may differ significantly depending on the adopted canonical limiting values of the ^{15}N chemical shift [16]; (ii) the $^{13}\text{C}^{\gamma}$ and $^{13}\text{C}^{\delta 2}$ chemical shifts cannot always be observed. In fact, only 106 $^{13}\text{C}^{\gamma}$, versus 4,703 $^{13}\text{C}^{\delta 2}$, chemical shifts of the imidazole ring of histidine have been deposited in the Biological Magnetic Resonance data Bank (BMRB) [17]. Hence, problems in the determination of the chemical shifts for these nuclei, such as that for the ground state of His 40 in the protein Im7 [14], often prevent the use of this methodology; and (iii) the observed one bond C–H SSCC value at the high-pH limit is ambiguous, as will be discussed below.

It should be noted, from Figure 1, that there are *only* two one-bond C–H's, $^1J_{\text{CH}}$, SSCC's of the imidazole ring of Histidine, namely the $^{13}\text{C}^{\epsilon 1}\text{-H}$, $^1J_{\text{C}\epsilon 1\text{H}}$, and the $^{13}\text{C}^{\delta 2}\text{-H}$, $^1J_{\text{C}\delta 2\text{H}}$, SSCC, respectively. Absences of an accurate value for each of these SSCC's, at the high-pH limit, gives rise to two different kinds of problems as explained below.

The first problem pertains to the use of $^1J_{\text{C}\epsilon 1\text{H}}$ SSCC to determine the protonation fraction of His, e.g., to detect sparsely populated, short-lived, protein states [14]. In detail, the

low-pH limiting value for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC appears to be quite well defined (221 ± 1.0 Hz [14]), for the $^1J_{\text{C}\epsilon\text{1H}}$ SSCC pH-dependence of four titrating His residues (His 6, His 13, His 26, His 87) of the PLCC γ SH2 protein domain [14,18]. However, the observed high-pH limit for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC differs among *five* His residues, of the PLCC γ SH2 protein domain, by up to ~ 6 Hz [14], i.e., four titrating His residues converge to a high-pH limiting value of 207 ± 1.0 Hz while the remaining one (His 57), which is the *only* non-titrating His residue, shows an almost flat, pH-independent, value of 203 ± 1.0 Hz [14]. The existence of two possible high-pH limiting values for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC, namely 207 ± 1.0 Hz or 203 ± 1.0 Hz [14], is a source of ambiguity. A similar ambiguity is found for four non-titrating His residues of subtilisin BPN' having $^1J_{\text{C}\epsilon\text{1H}}$ SSCC in the range of ~ 205 Hz to ~ 209 Hz [19].

The second problem pertains to a potential contradiction between an assumption [15] and existing evidence about the variation of $^1J_{\text{C}\delta\text{2H}}$ SSCC upon tautomeric change [13]. Platzer *et al.* [15] had proposed that the variations of $^1J_{\text{C}\delta\text{2H}}$ SSCC should be independent of the forms of the His tautomer, as for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC. On the other hand, there is experimental evidence for L-histidine at pH 12 in 80% d6-ethanol/20% H₂O at -55 °C [13], showing the existence of a large, rather than a small change, of $^1J_{\text{C}\delta\text{2H}}$ SSCC upon changes of the tautomeric forms.

As can be inferred from the above, a common problem, in both $^1J_{\text{CH}}$ SSCC's and the ^{15}N -based methods, is the need for accurate knowledge of the "canonical" limiting values of the imidazole ring of His in which each form of His, namely the protonated (H^+) and the tautomeric $\text{N}^{\epsilon\text{2}}\text{-H}$ and $\text{N}^{\delta\text{1}}\text{-H}$ forms, respectively, is present to the extent of 100%. In this regard, the canonical limiting values of the ^{15}N chemical shift have already been analyzed [16] and, hence, here we will determine the high-pH limiting values for both $^1J_{\text{CH}}$ SSCC's of the imidazole ring of His. By doing this, we will be able to: (i) eliminate any possible ambiguity about the actual value of $^1J_{\text{C}\epsilon\text{1H}}$ SSCC; and, (ii) resolve a possible contradiction associated with the variations of $^1J_{\text{C}\delta\text{2H}}$ SSCC upon changes in the relative amounts of the tautomeric forms.

Materials and Methods

Calculations details

All DFT-calculations of the two $^1J_{\text{CH}}$ SSCC's, of the imidazole ring of **His** in the Ac-**His**-NMe molecule, were carried out by using the Gaussian 09 suite of programs [20]; the Keywords used in Gaussian 09 (listed here for assessing the reproducibility of the calculations) were: “NMR=Mixed”, with the options “CPHF=Conv=10” and “Int=ultrafine” [21]. Additional Keywords, such as “Readatoms”, were also tested (see Results and Discussion section).

There are four contributions to the NMR coupling constants [22], namely, the Fermi Contact (**FC**), the Spin Dipolar (**SD**), the Paramagnetic Spin-Orbit (**PSO**), and the Diamagnetic Spin-Orbit (**DSO**) contribution, respectively. All four are known as the Ramsey contributions. For this reason, in each of the Tables, we have listed: (i) the sum (Σ) of *all* four Ramsey contributions to each DFT-computed $^1J_{\text{CH}}$ SSCC, as computed with the Gaussian 09 suite of programs; and (ii) the predicted values for the $^1J_{\text{CH}}$ SSCC's, listed in the last column of each Table, are obtained after adding an *ad-hoc* contribution of 5Hz, to the Σ term, due to the Zero Point Vibrational Contribution [23].

All the results in Table 1 correspond to gas-phase DFT calculations, while the ones in Tables 2-4 include DFT-calculations in both gas-phase and in the presence of solvent (see *Solvent Effects* section below). For the latter, all the results obtained with solvent are highlighted in italics and bold face.

Structural geometry of His

All the $^1J_{\text{CH}}$ SSCC calculations, at the DFT level of theory, were carried out by using the histidine geometry as defined in the Empirical Conformational Energy Program for Peptides and Proteins (ECEPP) in which their bond-lengths and bond-angles were parameterized by Momany *et al.* [24] and updated by Némethy *et al.* [25] by using a high-resolution ($R = 0.037$) X-ray crystal structure of histidine. In this regard, (i) less than 10% of more than 500,000 crystal structures deposited in the Cambridge Structural Database have an R -factor < 0.04 [26]; and (ii) a comparison between DFT-computed and observed $^{13}\text{C}^\alpha$ chemical shifts of two different structures of Ubiquitin [27], one that possesses non-regular geometry which has been obtained by X-ray diffraction at 1.8 Å resolution (PDB id 1UBQ [29]) and the other structure with regularized ECEPP geometry [27], 1UBQ_{reg}, with both, in terms of rmsd, leading to 3.28 ppm and 2.38 ppm, respectively. Supplementary analysis of

the agreement between these structures with the deposited electron density data of 1UBQ, in terms of the *R*-factor, leads to 19.2% and 23.1% for 1UBQ and 1UBQ_{reg}, respectively, while the all-heavy-atom rmsd between these two structures is only 0.142 Å.

Overall, the better agreement, in terms of $^{13}\text{C}^\alpha$ chemical-shifts, obtained with 1UBQ_{reg}, rather than 1UBQ, is consistent with the well-known recognition that the bond lengths and bond angles of both X-ray and NMR-derived structures of proteins are not defined as highly accurately as in studies of small molecules [29] with which the ECEPP geometry has been parameterized [24,25].

Solvent effects

It is well known that sizable solvent effects are important for the computation of $^nJ_{\text{CH}}\text{SSCC}$ only for $n = 1$ [30], with n being the number of intervening bonds between the Carbon and the Hydrogen. Consequently, dielectric solvent effects were taken into account during the DFT-calculations by using a Polarizable Continuum Model (PCM) [31,32] as implemented in the Gaussian package [20]. Because we are interested primarily in proteins in water, we decided to model the dielectric medium of such a system by using theoretical evidence [23] indicating that the average dielectric constants (D_i) inside and at-the-surface of a protein are around 6-7 and 20-30, respectively. Hence, the dielectric solvent effects on the DFT-computation of the $^1J_{\text{CH}}\text{SSCC}$ for the imidazole ring of His were computed by using the PCM approach with an *effective* D_i of 6.5 and 25.5, respectively.

It is worth noting that there is a *direct* and an *indirect* contribution of the solvent effects on the $^1J_{\text{CH}}\text{SSCC}$ [30], and both effects were taken into account here. The *direct* contribution, due to the polarization of the molecular electronic structure by the solvent, was considered here by using the PCM approach which, during the calculation of $^1J_{\text{CH}}\text{SSCC}$, takes into account the induced surface charges derived from the boundary conditions at the cavity surface. The *indirect* effect, caused by the change of the molecular geometry due to the solvent, was also considered here because we used a high-resolution X-ray crystal structure of histidine and, as is well known, the crystals contain, on average, ~50% of solvent [34].

Determination of the factors affecting the computation of $^1J_{\text{CH}}\text{SSCC}$

Among the factors affecting the computation of the four Ramsey contributions to $^1J_{\text{CH}}$ SSCC, we first analyzed their dependence on the functionals *and* on the basis sets chosen for the DFT calculations. This dependence is discussed below.

Selecting the functional

Maximoff *et al.* [35] ranked 20 functionals for their ability to reproduce the observed $^1J_{\text{CH}}$ SSCC of 31 chemical compounds containing 11 aromatic molecules. Among all the 20 functionals, we selected *only* those showing a mean absolute error $< 3\text{Hz}$ (see Table 1). In Table 1, we highlight in bold face the results from those functionals giving closer predictions to the observed $^1J_{\text{C}\epsilon\text{1H}}$ SSCC for the non-titrating His 57 of the PLCC γ SH2 protein domain, namely $\sim 203\text{Hz}$ [14]. These selected *best* functionals, namely, OPW91 and OPBE, are in agreement with the conclusion reached by Maximoff *et al.* [35] after testing 20 functionals. To decide which one of these two functionals should be adopted for our DFT calculations, the following complementary analysis was carried out. There is only one aromatic group ($\text{C}_4\text{H}_4\text{N}_2$), among 11 tested by Maximoff *et al.*, [35] possessing a C–H between 2 nitrogen atoms, in a similar chemical arrangement as that of $^{13}\text{C}^{\epsilon 1}$ of the imidazole ring of His (see Figure 1). For this particular aromatic group ($\text{C}_4\text{H}_4\text{N}_2$), the results obtained by Maximoff *et al.* [35] with the OPW91 (0.2Hz) and the OPBE (2.4 Hz) functionals, in terms of the difference, Δ , between the observed and DFT-computed $^1J_{\text{CH}}$ SSCC, show that OPW91 matches the experimental data better than OPBE (see Supporting Information in Table 1 from Maximoff *et al.* [35]). This result indicates that OPW91 is the best functional with which to predict the observed values of $^1J_{\text{CH}}$ SSCC for the $\text{C}_4\text{H}_4\text{N}_2$ aromatic group and, hence, OPW91 rather than OPBE was chosen for the DFT calculations in this work.

Selecting a basis set

It is well known that $^1J_{\text{CH}}$ SSCC calculations are dominated by the Fermi-Contact (FC) contribution which depends strongly on the electron density close to the nuclei and, hence, demands high quality of the basis set chosen [36]. Consequently, *all* the DFT-calculations in this work were carried out by using the “aug-cc-pVTZ-J” basis set [37] which is specially designed to study NMR properties. In addition, there is an important consensus

that this basis set gives satisfactory close agreement between observed and computed $^1J_{\text{CH}}$ SSCC [21,35,36,38,39,40].

A locally-dense basis set approximation [41] was also used to sense the response of the DFT-computed $^1J_{\text{CH}}$ SSCC to the basis set distribution chosen, i.e., different combinations of “aug-cc-pVTZ-J” basis sets on the nuclei of interest of the imidazole ring of **His**, and a “6-31G” or “3-21G” basis set on the remaining nuclei of the Ace-**His**-NMe molecule, were tested here; the results are shown in Tables 2 and 3.

Results and Discussion

In general, the gas-phase results from Tables 2 and 3 enable us to conclude that the computed $^1J_{\text{C}\epsilon\text{1H}}$ SSCC with a locally-dense basis set approximation is indistinguishable from the one obtained by using a *uniform* basis set distribution, i.e., a “aug-cc-pVTZ-J” basis set on *all* the nuclei of the Ac-**His**-NMe molecule, but with a considerable reduction in computational time. In addition, a comparison of the DFT-computed $^1J_{\text{C}\epsilon\text{1H}}$ SSCC values, from rows 4-7 of Table 2, shows that the results do not depend on the ϕ , ψ , χ_1 and χ_2 torsional angles of His. It is worth noting that, for a given basis set distribution, the use of the keyword “*Readatoms*”, that focuses the computation of $^1J_{\text{CH}}$ SSCC on the chosen pair of nuclei, can speed up the calculation by up to ~7 times.

Consideration of solvent effects, by using the PCM approach with two D_i , namely 6.5 and 25.5, increases the gas-phase results for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC on the $\text{N}^{\epsilon 2}$ -H and the $\text{N}^{\delta 1}$ -H tautomers by up to ~3Hz (see the highlight in italics in rows 8-9 and 5-6 of Tables 2 and 3, respectively). In view of all the assumptions and round off made, the most accurate DFT-computed value for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC is $205\text{Hz} \pm 1.0\text{Hz}$. Without doubt, within ± 1 Hz, the values of $^1J_{\text{C}\epsilon\text{1H}}$ SSCC for the tautomers of the imidazole ring of His, are indistinguishable. This likely high-pH limiting value ($205\text{Hz} \pm 1.0\text{Hz}$) for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC is within the range of the *lowest* high-pH limiting value observed from non-titrating His on both the subtilisin BPN' and the PLCC γ SH2 protein domain, explicitly ~205Hz and 203 ± 1.0 Hz, respectively. Moreover, it is also in very good agreement with the observed values (~205Hz for each of the His tautomers) obtained from NMR experiments for L-histidine at pH 12 in 80% d6-ethanol/20% H₂O at -55 °C [13].

The computed $^1J_{C\delta 2H}$ SSCC in the gas-phase and in the presence of solvent, by using the PCM approach, with an “aug-cc-pVTZ-J” basis set on *all* nuclei of the imidazole ring of His and a “6-31G” basis set on the remaining nuclei of the Ac-His-NMe molecule, are listed in Table 4 for both the $N^{\epsilon 2}$ -H and the $N^{\delta 1}$ -H tautomer, respectively. In the presence of solvent effects (highlighted in italics in Table 4) the most remarkable of the results obtained for $^1J_{C\delta 2H}$ SSCC, is the existence of a large difference between His tautomers, namely 15Hz. This result is counter to the assumption made by Platzner *et al.* [15] that $^1J_{C\delta 2H}$ SSCC should behave like $^1J_{C\epsilon 1H}$ SSCC upon tautomeric changes, i.e., showing minimal or no difference. Even more important, our result for the difference between tautomers, in terms of $^1J_{C\delta 2H}$ SSCC, is in excellent agreement with existing NMR-based evidence on L-histidine at pH 12 in 80% d6-ethanol/20% H_2O at $-55\text{ }^{\circ}C$ [13], showing that the difference in $^1J_{C\delta 2H}$ SSCC between His tautomers is indeed 15Hz.

Despite the excellent agreement mentioned above, there is a conflict between the observed and the DFT-computed value for $^1J_{C\delta 2H}$ SSCC for each of the tautomeric forms of the neutral His, i.e., the computed values in the presence of solvent for the $N^{\epsilon 2}$ -H and the $N^{\delta 1}$ -H tautomer are, as shown in Table 4, 165Hz and 180Hz, respectively, while the observed values are in the reverse order, namely, 180Hz and 165Hz, respectively (see Table 2S of Sudmeier *et al.* [13]). An irrefutable test to resolve this discrepancy would be to repeat the NMR experiments on L-histidine at pH 12 in 80% d6-ethanol/20% H_2O at $-55\text{ }^{\circ}C$. However, this is beyond the scope of this work. For this reason, and in order to validate that the predictions in Table 4 actually belong to the referred His tautomer, the following analyses were carried out, first a validation in terms of the observed chemical shifts, second, a computation of the difference in the $^1J_{C\delta 2C\gamma}$ SSCC values ($\Delta_{C\delta 2C\gamma}$) upon tautomeric changes and, third, a computation of the $^1J_{C\delta N\epsilon 2}$ and the $^2J_{C\delta N\delta 1}$ SSCC in the $N^{\epsilon 2}$ -H and in the $N^{\delta 1}$ -H tautomers, respectively.

The $^{13}C^{\gamma}$ and $^{13}C^{\delta 2}$ shieldings for each tautomer of the imidazole ring of His, mentioned in Table 5, were computed at the DFT-level of theory because these nuclei are very sensitive probes with which to confirm the tautomeric forms of His accurately [6,11]. Chemical shifts are related to shieldings as differences with respect to a reference shielding.

Hence, from Table 5, we can straightforwardly infer the following chemical-shifts inequalities:

$$(^{13}\text{C}^\gamma)_{\text{N}\epsilon^2\text{-H}} \gg (^{13}\text{C}^\gamma)_{\text{N}\delta^1\text{-H}} \text{ and } (^{13}\text{C}^{\delta^2})_{\text{N}\delta^1\text{-H}} \gg (^{13}\text{C}^{\delta^2})_{\text{N}\epsilon^2\text{-H}} \quad (1)$$

The chemical-shift inequalities between tautomers given by Eq. (1) are in full agreement with both the observed chemical shifts (see Table 2S of Sudmeier *et al.* [13]) and the DFT-computed shielding values in model peptides for the tautomeric forms of the imidazole ring of histidine (see Figure 1 of Vila *et al.* [6]).

The computation, at the DFT-level of theory, of the difference ($\Delta_{\text{C}\delta^2\text{C}\gamma}$) for the $^1J_{\text{C}\delta^2\text{C}\gamma}$ SSCC upon tautomeric changes, viz., with $\Delta_{\text{C}\delta^2\text{C}\gamma} = ({}^1J_{\text{C}\delta^2\text{C}\gamma}|^{\epsilon^2} - {}^1J_{\text{C}\delta^2\text{C}\gamma}|^{\delta^1})$, where ${}^1J_{\text{C}\delta^2\text{C}\gamma}|^\lambda$ with $\lambda = \epsilon^2$ or δ^1 represents the computed $^1J_{\text{C}\delta^2\text{C}\gamma}$ SSCC value for the $\text{N}^{\epsilon^2}\text{-H}$ and the $\text{N}^{\delta^1}\text{-H}$ tautomers, respectively, was carried out by using the His structures, the functional (OPW91) and the basis-set described in Table 4. The computed difference, $\Delta_{\text{C}\delta^2\text{C}\gamma} = +2.6$ Hz, is in good agreement, but opposed in sign, with the observed difference, $\Delta_{\text{C}\delta^2\text{C}\gamma} = -3.0$ Hz (see Table 2S of Sudmeier *et al.* [13]). Although no effort was made to optimize the functional and basis set for the computed one-bond C–C, $^1J_{\text{CC}}$,SSCC, an analyses with *all* the functionals shown in Table 1, with a locally-dense basis set distribution mentioned in Table 4, shows consistency with the result obtained with OPW91, namely $\Delta_{\text{C}\delta^2\text{C}\gamma} = +1.8$ Hz, $+2.0$ Hz, $+2.6$ Hz, $+1.8$ Hz and $+1.9$ Hz, for the B3LYP, B3P86, OPBE, B972 and PB86 functionals, respectively.

For a long time, it has been recognized that the Carbon–Nitrogen spin-spin coupling constants (J_{CN}) can be used for an accurate determination of the tautomeric forms of His [8,42]. For this reason, and despite the fact that no effort was made to optimize the functional or the basis set with which to compute the one- and two-bond C–N, J_{CN} , an analysis of the computed J_{CN} values was carried out by using the His structures, the functional (OPW91), and the basis set described in Table 4. Among *all* possible C–N SSCC of the imidazole ring of His, we focused the calculations on both the $^1J_{\text{C}\delta^2\text{N}\epsilon^2}$ and the $^2J_{\text{C}\delta^2\text{N}\delta^1}$ because these SSCC were suggested as the probes of choice with which to determine the ratio of the $\text{N}^{\epsilon^2}\text{-H}$ and the $\text{N}^{\delta^1}\text{-H}$ tautomers, accurately [8]. Consequently, the results obtained here for $^1J_{\text{C}\delta^2\text{N}\epsilon^2}$ (13.2 Hz), in

the N^{ϵ^2} -H tautomer (see Figure 1a), and $^2J_{C\delta^2N\delta^1}$ (4.7 Hz), in the N^{δ^1} -H tautomer (see Figure 1b), show the same trend as that of the observed absolute values for these tautomers, namely ~ 13 Hz and ~ 5 Hz, respectively [8,42]. Even though this analysis does show internal consistency with the previous calculations of both the chemical shifts and $^1J_{C\delta^2H}$ SSCC, the validation of the theoretical results remains elusive without farther experimental evidence.

Overall, under the *only* condition that His must be non-protonated and, assuming the correctness of the theoretical predictions for $^1J_{C\delta^2H}$ SSCC of the N^{ϵ^2} -H and the N^{δ^1} -H tautomers, it is possible to determine the fraction of the N^{δ^1} -H tautomeric form of the imidazole ring of His, f^{δ^1} , as a function of the observed $^1J_{C\delta^2H}$ SSCC value. So, assuming that $J^{obs} = f^{\epsilon^2} J_{high-pH}^{\epsilon^2} + f^{\delta^1} J_{high-pH}^{\delta^1}$, and $f^{\epsilon^2} + f^{\delta^1} = 1$ the following equation can be used for this purpose:

$$f^{\delta^1} = \frac{J^{obs} - J_{high-pH}^{\epsilon^2}}{J_{high-pH}^{\delta^1} - J_{high-pH}^{\epsilon^2}} = \frac{J^{obs} - 165}{15} \quad (2)$$

where J refers to $^1J_{C\delta^2H}$ SSCC, “*obs*” is the value observed, $J_{high-pH}^{\epsilon^2}$ and $J_{high-pH}^{\delta^1}$ refer to $^1J_{C\delta^2H}$ SSCC obtained, at the high-pH limiting value, from the N^{ϵ^2} -H and N^{δ^1} -H His tautomers, respectively; as shown in Table 4, “165” is the value for the $J_{high-pH}^{\epsilon^2}$ SSCC of the N^{ϵ^2} -H tautomer, and “15” is the difference, 180-165, between the computed $^1J_{C\delta^2H}$ SSCC for the tautomeric forms of His at the high-pH limit. Naturally, the fraction of the N^{ϵ^2} -H tautomeric forms is given by: $f^{\epsilon^2} = 1 - f^{\delta^1}$.

It is worth noting that, even if the Sudmeier *et al.* [13] observations were confirmed Equation (2) will still be valid after the substitution $\delta^1 \rightarrow \epsilon^2$.

Conclusions

The main results obtained for $^1J_{CH}$ SSCC in the N^{ϵ^2} -H and N^{δ^1} -H tautomers of the imidazole ring of His indicate that: (i) $^1J_{C\epsilon^1H} = 205 \pm 1.0$ Hz should be adopted as the canonical high-pH limiting value for this SSCC; (ii) $^1J_{C\epsilon^1H}$ SSCC is not sensitive enough to

be used as a probe with which to determine the tautomeric states of the imidazole ring of His; and (iii) $^1J_{C\delta 2H}$ SSCC shows a large difference ($\sim 15\text{Hz}$) between tautomers and, hence, this SSCC emerges as a very sensitive probe with which to identify the tautomeric fractions of non-protonated His.

Overall, the theoretical confirmation showing that $^1J_{C\epsilon 1H}$ SSCC is *not* sensitive upon tautomeric change means that the use of this SSCC will enable us to predict the protonation fraction of His, as a function of the pH, accurately, because the prediction will not depend on the fraction of the tautomeric forms. In addition, the fraction of the tautomeric forms of the imidazole ring of His, at either high-pH or for non-protonated buried His, can be predicted accurately by using Equation (2) for $^1J_{C\delta 2H}$ SSCC.

Acknowledgments

This research was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB10-19767) (HAS), and PIP-112-2011-0100030 from IMASL-CONICET-Argentina, Project 3-2212 from UNSL-Argentina, and PICT-2014-0556 from AMPCyT-Argentina (JAV).

References

- [1] Cheng, F.; Sun H.; Zhang, Y.; Mukkamala, D.; Oldfield, E. *J. Am. Chem. Soc.* 127 (2005) 12544-12554.
- [2] Jensen, M.R.; Has, M.A.S.; Hansen, D.F.; Led, J.J. *Cell. Mol. Life Sci.* 64 (2007) 1085-1104.
- [3] Hass, M.A.S.; Hansen, D.F.; Christensen, H.E.M.; Led, J.J.; Kay, L.E. *J. Am. Chem. Soc.* 130 (2008) 8460-8470.
- [4] Kumar, M.; Venkatnathan A. *J. Phys. Chem. B* 119 (2015) 3213-3222.
- [5] Li, S.; Hong, M. *J. Am. Chem. Soc.* 133 (2011) 1534-1544.
- [6] Vila, J.A.; Arnautova, Y.A.; Vorobjev, Y.N.; Scheraga, H.A. *Proc. Natl. Acad. Sci. USA* 108 (2011) 5602-5607.
- [7] Demchuk, E.; Wade, R.C. *J. Phys. Chem.* 100 (1996) 17373–17387.
- [8] Shimba, N.; Serber, Z.; Ledwidge, R.; Miller, S.M.; Craik, C.S.; Dotsch, V. *Biochem.* 42 (2003) 9227-9234.
- [9] Bermúdez, C.; Mata, S.; Cabezas, C.; Alonso, J.L. *Angew. Chem. Int. Ed.* 53 (2014) 11015-11018.
- [10] Blomberg, W.; Rüterjans, H. *J. Am. Chem. Soc.* 99 (1977) 8149-8159.
- [11] Pelton, J.G.; Torchia, D.A.; Meadow, N.D.; Roseman S. *Prot. Sci.* 2 (1993) 543-558.
- [12] Kolehmainen, E.; Ośmiałowski, B. *Int. Rev. Phys. Chem.* 31 (2012) 567-629.
- [13] Sudmeier, J.L.; Bradshaw, E.M.; Haddad, E.C.; Day, R.M. Thalhauser, C.J.; Bullock, P.A.; Bachovchin, W.W. *J. Am. Chem. Soc.* 125 (2003) 8430-8431.
- [14] Hansen, A.L.; Kay, L.E. *Proc. Natl. Acad. Sci. USA* 111 (2014) E1705-E1712.
- [15] Platzer, G.; Okon, M.; McIntosh, L.P. *J. Biomol. NMR* 60 (2014) 109-129.
- [16] Vila, J.A. *J. Phys. Chem. B* 116 (2012) 6665-6669.
- [17] Ulrich, E.L.; Akutsu, H.; Doreleijers, J.F.; Harano, Y.; Ioannidis, Y.E.; Lin J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C.F.; Tolmie, D.E.; Wenger, R.K.; Yao, H.; Markley, J.L. *Nucleic Acids Res.* 36 (2008) D402-D408.
- [18] Singer, A.U.; Forman-Kay, J.D. *Prot. Sci.* 6 (1997) 1910-1919.

- [19] Day, R.M.; Thalhauser, C.J.; Sudmeier, J.L.; Vincent, M.P.; Torchilin, E.V.; Sanford, D.G.; Bachovchin, C.W. Bachovchin, W.W. *Protein Sci.* 12 (2003) 794-810.
- [20] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J.J.; Dapprich, S.; Daniels, A.D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, revision A.1; Gaussian, Inc.: Wallingford, CT, **2009**.
- [21] Deng, W.; Cheeseman, J.R.; Frisch, M.J. *J. Chem. Theory. Comput.* 2 (2006) 1028-1037.
- [22] Ramsey, N.F. *Phys. Rev.* 91 (1953) 303-307.
- [23] Ruden, T.A.; Lutnæs, O.B.; Helgaker, T.; Ruud, K.J. *Chem. Phys.* 118 (2003) 9572-9581.
- [24] Momany, F.A.; McGuire, A.F.; Burgess, A.W.; Scheraga, H.A. *J. Phys. Chem.* 79 (1975) 2361-2381.
- [25] Némethy, G.; Gibson, K.D.; Palmer, K.A.; Yoon, C.N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H.A. *J. Phys. Chem.* 96 (1992) 6472-6484;
- [26] Oda, K.; Koyama, H.A. *Acta Cryst. B* 28 (1972) 639-642.
- [27] Arnautova, Y.A.; Vila, J.A.; Martin, O.A.; Scheraga, H.A. *Acta Cryst.* D65 (2009) 697-703.
- [28] Vijay-Kumar, S.; Bugg, C.E.; Cook, W.J. *J. Mol. Biol.* 194 (1987) 531-544.

- [29] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic Acids Research* 28 (2000) 235-242.
- [30] Ruud, K.; Frediani, L.; Cammi, R.; Mennucci, B. *Int. J. Mol. Sci.* 4 (2003) 119-134.
- [31] Tomasi, J.; Persico, M. *Chem. Rev.* 94 (1994) 2027-2094.
- [32] Tomasi, J. *Comput. Mol. Sci.* 1 (2011) 855-867.
- [33] Li, L.; Li, C.; Zhang, Z.; Alexov, E. *J. Chem. Theory. Comput.* 9 (2013) 2126-2136.
- [34] Andersson, K.M.; Hovmöller, S. *Acta Cryst.* D56 (2000) 789-790.
- [35] Maximoff, S.N.; Peralta, J.E.; Barone, V.; Scuseria, G.E. *J. Chem. Theory. Comput.* 1 (2005) 541-545.
- [36] Peralta, J.E.; Scuseria, G.E.; Cheeseman, J.R.; Frisch, M.J. *Chem. Phys. Lett.* 375 (2003) 452-458.
- [37] Provasi, P.F.; Aucar, G.A.; Sauer, S.P.A. *J. Chem. Phys.* 115 (2001) 1024-1333.
- [38] Contreras, R.H.; Peralta, J.E.; Giribet, C.G.; de Azua, M.C.; Facelli, J.C. *Annu. Rep. NMR Spectrosc.* 41 (2000) 55-184.
- [39] Krivdin, L.B.; Contreras, R.H. *Ann. Repts. NMR Spectrosc.* 61 (2007) 133-245.
- [40] San Fabian, J.; Garcia de la Vega, J.M.; Suardiaz, R.; Fernández-Oliva, M.; Pérez, C.; Crespo-Otero, R.; Contreras, R.H. *Magn. Reson. Chem.* 51 (2013) 775-787.
- [41] Chesnut, D.B.; Moore, K.D. *J. Comp. Chem.* 10 (1989) 648-659.
- [42] Sudmeier, J.L., Ash, E.L., Günter, U.L., Luo, X., Bullock, P.A., Bachovchin, W.W. *J. Mag. Res. B* 113 (1996) 236-247.

Table 1: Test of functionals for the DFT computations of $^1J_{\text{C}\epsilon\text{1H}}$ SSCC of the $\text{N}^{\epsilon 2}\text{-H}$ tautomer^a

Functional	$\Sigma_{\{\text{FC,SD,PSO,DSO}\}}$ (Hz)	$^1J_{\text{C}\epsilon\text{1H}}$ (Hz)
OPW91	199.02	~204
B3LYP	231.65	~237
B3P86	215.54	~220
OPBE	198.10	~203
B972	212.57	~218
BP86	211.35	~216

- a) All gas-phase DFT calculations of $^1J_{\text{C}\epsilon\text{1H}}$ SSCC were carried out on Ac-His-NMe by using the Gaussian 09 suite of programs [20]; the chosen ϕ , ψ , ω , $\chi 1$ and $\chi 2$ torsional angles for His correspond to a local-minimum of the ECEPP force-field [25] in the α -helical region of the Ramachandran map: -73.563° , -35.197° , -179.856° , 66.389° and -62.607° , respectively; for each functional we used an “aug-cc-pVTZ-J” basis set [37] on *all* nuclei of the imidazole ring of His, and a “6-31G” basis set on the remaining nuclei of Ac-His-NMe. The total (Σ) is a sum over the four Ramsey contributions [22], as given by the output of the Gaussian 09 suite of programs [20], namely, the Fermi Contact (**FC**), the Spin Dipolar (**SD**), the Paramagnetic Spin-Orbit (**PSO**), and the Diamagnetic Spin-Orbit (**DSO**) contribution, respectively; the last column lists the predicted value for the one-bond $^1J_{\text{C}\epsilon\text{1H}}$ SSCC after adding 5 Hz to the Σ term (second column), due to the Zero Point Vibrational Contribution [23].

Table 2: High-pH limiting value for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC of the $\text{N}^{\epsilon 2}\text{-H}$ tautomer^a

$\Sigma_{\{\text{FC,SD,PSO,DSO}\}}$ (Hz)	$^1J_{\text{C}\epsilon\text{1H}}$ (Hz)
199.20	$\sim 204^{\text{b}}$
199.29	$\sim 204^{\text{c}}$
199.10	$\sim 204^{\text{d}}$
199.02	$\sim 204^{\text{e}}$
198.91	$\sim 204^{\text{f}}$
199.12	$\sim 204^{\text{g}}$
198.62	$\sim 204^{\text{h}}$
201.26	$\sim 206^{\text{i}}$
201.78	$\sim 207^{\text{j}}$

- a) All gas-phase DFT calculations of $^1J_{\text{C}\epsilon\text{1H}}$ SSCC, unless otherwise noted, for which see items *i-j* below, were carried out for Ac-His-NMe by using the Gaussian 09 suite of programs [20]; the chosen ϕ , ψ , ω , $\chi 1$ and $\chi 2$ torsional angles for His correspond to a local-minimum of the ECEPP force-field [25] in the α -helical region of the Ramachandran map: -73.563° , -35.197° , -179.856° , 66.389° and -62.607° , respectively, unless otherwise noted, for which see items *f-h* below; the total (Σ) is a sum over the four Ramsey contributions (see footnote *a* of Table 1) [22]; column 2 lists the predicted values for the one-bond $^1J_{\text{C}\epsilon\text{1H}}$ SSCC after adding 5 Hz to the Σ term (in column 1), due to the Zero Point Vibrational Contribution [23].
- b) Result obtained by using a *uniform* basis set “aug-cc-pVTZ-J” [37] for *all* the nuclei of Ac-His-NMe.
- c) Result obtained by using an “aug-cc-pVTZ-J” basis set for the $^{13}\text{C}^{\epsilon 1}$, $^1\text{H}^{\epsilon 1}$, $\text{N}^{\delta 1}$, $\text{N}^{\epsilon 2}$ and $\text{H}^{\epsilon 2}$ nuclei of the imidazole ring of His and a “3-21G” basis set for *all* the remaining nuclei of Ac-His-NMe, i.e., by using a “locally-dense” basis set approximation [41].
- d) Same as (c) with a “6-31G” rather than a “3-21G” basis set.

- e) Result obtained by using an “aug-cc-pVTZ-J” basis set for *all* nuclei of the imidazole ring of His, and a “6-31G” basis set for the remaining nuclei of Ac-His-NMe, i.e., by using a “locally-dense” basis set approximation.
- f) Same as (e) with the ϕ , ψ , ω , χ_1 and χ_2 torsional angles of His from a local-minimum, of the ECEPP force-field [25], in the extended region of the Ramachandran map: -142.183° , 153.652° , -179.870° , -58.886° and 113.207° , respectively.
- g) Same as (e) with the ϕ , ψ , ω , χ_1 and χ_2 torsional angles of His from a local-minimum, of the ECEPP force-field,²⁵ in the extended region of the Ramachandran map: -156.922° , 159.176° , 179.634° , 60.066° and 96.365° , respectively.
- h) Same as (e) with the ϕ , ψ , ω , χ_1 and χ_2 torsional angles of His from a local-minimum, of the ECEPP force-field [25], in the extended region of the Ramachandran map: -154.460° , 157.818° , -178.614° , -152.343° and -67.537° , respectively.
- i) Same as (c) with solvent effects computed by using the PCM [31,32] approach with $D_i = 6.5$.
- j) Same as (i) with $D_i = 25.5$

Table 3: High-pH limiting value for $^1J_{\text{C}\epsilon\text{1H}}$ SSCC of the $\text{N}^{\delta\text{1}}\text{-H}$ tautomer^a

$\Sigma_{\{\text{FC,SD,PSO,DSO}\}}$ (Hz)	$^1J_{\text{C}\epsilon\text{1H}}$ (Hz)
197.98	$\sim 203^{\text{b}}$
198.02	$\sim 203^{\text{c}}$
197.84	$\sim 203^{\text{d}}$
19787	$\sim 203^{\text{e}}$
199.32	$\sim 204^{\text{f}}$
199.66	$\sim 205^{\text{g}}$

- a) All gas-phase DFT calculations of $^1J_{\text{C}\epsilon\text{1H}}$ SSCC, unless otherwise noted, for which see items *f-g* below, were carried out for Ac-His-NMe by using the Gaussian 09 suite of programs [20]; the total (Σ) is a sum over the four Ramsey contributions (see footnote *a* of Table 1); the chosen ϕ , ψ , ω , χ_1 and χ_2 torsional angles for His correspond to a local-minimum of the ECEPP force-field [25] in the α -helical region of the Ramachandran map: -74.737° , -39.192° , 179.928° , -67.874° , and 28.534° , respectively; column 2 lists the predicted values for the one-bond $^1J_{\text{C}\epsilon\text{1H}}$ SSCC after adding 5 Hz to the Σ term (in column 1), due to the Zero Point Vibrational Contribution [23].
- b) Result obtained by using a *uniform* basis set “aug-cc-pVTZ-J” [37] for *all* the nuclei of Ac-His-NMe.
- c) Result obtained by using an “aug-cc-pVTZ-J” basis set for the $^{13}\text{C}^{\epsilon\text{1}}$, $^1\text{H}^{\epsilon\text{1}}$, $\text{N}^{\delta\text{1}}$, $\text{N}^{\epsilon\text{2}}$ and $\text{H}^{\epsilon\text{2}}$ nuclei of the imidazole ring of His, and a “3-21G” basis set for *all* the remaining nuclei of Ac-**His**-NMe, i.e., by using a “locally-dense” basis set approximation [41].
- d) Same as (c) with “6-31G” rather than a “3-21G” basis set.
- e) Result obtained by using an “aug-cc-pVTZ-J” basis set for *all* nuclei of the imidazole ring of His, and a “6-31G” basis set on the remaining nuclei of Ac-His-NMe, i.e., by using a “locally-dense” [41] basis set approximation.
- f) Same as (e) with solvent effects computed by using the PCM [31,32] approach with $D_i = 6.5$.
- g) Same as (f) with $D_i = 25.5$.

Table 4: High-pH limiting value for $^1J_{\text{C}\delta\text{2H}}$ SSCC^a

Tautomer	$\Sigma_{\{\text{FC,SD,PSO,DSO}\}}$ (Hz)	$^1J_{\text{C}\delta\text{2H}}$ (Hz)
$\text{N}^{\varepsilon\text{2}}\text{-H}$	158.54	$\sim 164^{\text{b}}$
	159.87	$\sim 165^{\text{c}}$
	160.08	$\sim 165^{\text{d}}$
$\text{N}^{\delta\text{1}}\text{-H}$	176.14	$\sim 181^{\text{b}}$
	175.38	$\sim 180^{\text{c}}$
	175.07	$\sim 180^{\text{d}}$

- a) All gas-phase DFT calculations of $^1J_{\text{C}\delta\text{2H}}$ SSCC, unless otherwise noted, for which see items *c-d* below, were carried out on Ac-His-NMe by using the Gaussian 09 suite of programs [20]; the chosen ϕ , ψ , ω , χ1 and χ2 torsional angles for the $\text{N}^{\varepsilon\text{2}}\text{-H}$ and $\text{N}^{\delta\text{1}}\text{-H}$ tautomers of His correspond to a local-minimum in the α -helical region of the Ramachandran map, and are given in footnote *a* of Tables 2 and 3, respectively; the total (Σ) is a sum over the four Ramsey [22] contributions (see item *a* of Table 1); column 3 list the predicted values for the one-bond $^1J_{\text{C}\delta\text{2H}}$ SSCC after adding 5 Hz to the Σ term (in column 2), due to the Zero Point Vibrational Contribution [23].
- b) Result obtained by using an “aug-cc-pVTZ-J” basis set [37] for *all* nuclei of the imidazole ring of His and “6-31G” on the remaining nuclei of Ac-His-NMe.
- c) Same as (*b*) with solvent effects computed by using the PCM [31,32] approach with $D_i = 6.5$
- d) Same as (*c*) with $D_i = 25.5$

Table 5: Chemical shieldings for the $^{13}\text{C}^{\delta 2}$ and $^{13}\text{C}^{\gamma}$ nuclei ^a

Nucleus	N ^{$\varepsilon 2$} -H tautomer Shielding (ppm)	N ^{$\delta 1$} -H tautomer Shielding (ppm)
$^{13}\text{C}^{\gamma}$	42.1	61.9
$^{13}\text{C}^{\delta 2}$	78.1	60.8

- a) All gas-phase DFT-calculations were computed, by using the Gaussian 09 suite of programs [20], for Ace-His-NMe with the OB98 functional [6] and a “6-311+G(2d,p)” basis set [6] for *all* the nuclei of His and a “6-31G” basis set on the remaining nuclei of the molecule, i.e., by using a “locally-dense” basis set approximation.⁴¹

Figure 1.- Ball and stick representation of the forms of the imidazole ring of His, namely the: **(a)** $N^{\epsilon 2}$ -H, or τ tautomer [5], **(b)** $N^{\delta 1}$ -H, or π tautomer [5], and **(c)** H^+ form, respectively.

Histidine in solution at high-pH

$N^{\delta 1}$ -H, or π tautomer

$$\Delta^1 J_{C^{\epsilon 1}H} \sim 0 \text{ Hz}$$

$$\Delta^1 J_{C^{\delta 2}H} \sim 15 \text{ Hz}$$

$N^{\epsilon 2}$ -H, or τ tautomer