J|A|C|S
COMMUNICATIONS

Published on Web 06/19/2003

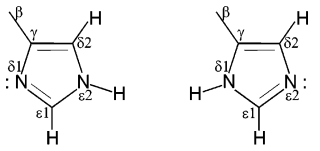
Identification of Histidine Tautomers in Proteins by 2D $^{1}H/^{13}C^{\delta 2}$ One-Bond Correlated NMR

James L. Sudmeier, Elizabeth M. Bradshaw, Kristin E. Coffman Haddad, Regina M. Day,[†] Craig J. Thalhauser, Peter A. Bullock, and William W. Bachovchin*

Department of Biochemistry, Tufts University School of Medicine, Boston, Massachusetts 02111

Received January 8, 2003; E-mail: william.bachovchin@tufts.edu

The neutral imidazole side chain of histidine, generally found at high pH, exists in equilibrium between two forms: the $N^{\epsilon 2}-H$ tautomer (**1**) and the $N^{\delta 1}-H$ tautomer (**2**). In proteins, and in smaller His compounds, nature exhibits a distinct preference for **1**, while reserving **2** for special tasks.



An example of tautomer **2** is the catalytic His 57 ($pK_a \approx 7$) in the active site of the serine protease α-lytic protease (α-LP). The existence of **2** in α-LP at high pH was first proven by using direct observation of ^{15}N NMR resonances in ^{15}N His-labeled enzyme.[1,2] In all serine protease active sites, tautomer **2** is stabilized by H-bonding between the $N^{\delta 1}-H$ proton and the carboxylate group of the adjacent "catalytic triad" residue.[3] Other examples of **2** are "structural" His residues, generally H-bonded, shielded from solvent,[4] and pH-invariant (i.e. $pK_a < 3$), such as His 226 in subtilisin BPN′.

We have recorded the 2D $^{1}H/^{13}C$ correlated NMR spectra shown in Figure 1 of various samples containing uniformly $^{13}C,^{15}N$-labeled histidines, utilizing delays optimized for the directly bonded $^{1}J_{CH}$ spin-coupling constants of ∼200 Hz exhibited by His side chains. Panel A is from the spectrum of monomeric [U-$^{13}C,^{15}N$]-L-His (Cambridge Isotope Labs) at pH 12 and −55 °C in 80% d_6-ethanol/20% water. As shown using ^{15}N NMR,[5] under these conditions the rapid equilibrium between the two His tautomers can be frozen, and panel A in Figure 1 displays only the $^{1}H/^{13}C^{\delta 2}$ correlation resonance of tautomer **2**, constituting 26% of the mixture by 1D ^{13}C spectral integrations. The dumbbell-shaped resonances in Figure 1 are caused by the large one-bond $C_\gamma-C_\delta$ spin-coupling constants, $^{1}J_{C\gamma C\delta}$, showing interesting variations (62−92 Hz) compared to monomer (71 Hz), of unknown origin. As illustrated in panel B for His 187 of "TBD", the origin-specific DNA binding domain from simian virus 40 T-antigen (T-ag-$OBD_{131-259}$),[6] these spin-couplings can readily be "tuned out", yielding a single resonance with increased sensitivity by use of a constant time heteronuclear single quantum coherence (ct-HSQC) pulse program.[7]

Of the six His residues in subtilisin BPN′ (MW = 27.5, high-stability mutant N155A),[8] the four $C^{\delta 2}$ nuclei that resonate between 113.3 and 115.7 at pH 10.5 (Table 1S) are all assigned to tautomer **1**. The two remaining His residues in subtilisin BPN′, shown in Figure 1, panel C, exhibit $^{13}C^{\delta 2}$ resonances at 125.0 and 124.2 ppm—the former the nontitrating signal assignable to His 226, the latter the titrating "catalytic triad" active site His 64—both predicted from the X-ray crystal structure[4] to be tautomer **2** and confirmed here in solution. His 64 undergoes a large downfield shift in its $^{13}C^{\delta 2}$ resonance from 118.0 to 124.2 ppm as the pH is increased from 5.5 to 10.5—a difference of 6.2 ppm between cation and tautomer **2**.

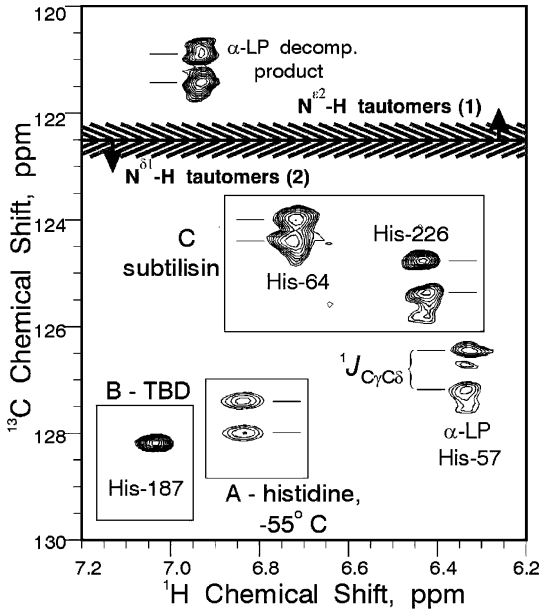


Figure 1. Selected portions of 2D $^{1}H/^{13}C^{\delta 2}$ correlation NMR spectra of four uniformly $^{13}C,^{15}N$-labeled His-containing samples: Underlying is a 500 MHz HMQC spectrum of α-lytic protease grown with [U-$^{13}C,^{15}N$]-L-His, pH 9.0 and 25 °C. (A) 500 MHz HMQC of [U-$^{13}C,^{15}N$]-L-His at pH 12 and −55 °C in 80% d_6-ethanol/20% H_2O; (B) 600 MHz ct-HSQC of U-$^{13}C,^{15}N$-labeled TBD at pH 5.5 and 30 °C (constant delay 13.5 ms). TBD has no tryptophans. (C) 600 MHz HMQC of subtilisin BPN′ (N155A) grown with [U-$^{13}C,^{15}N$]-L-His, pH 10.5, 25 °C. The hatched barrier shows the divide between tautomers **1** and **2**.

The use of $^{13}C^{\delta 2}$ NMR chemical shifts for discrimination between His tautomers was first performed by ^{13}C NMR spectra of small histidine-containing compounds.[9] Using His derivatives 1-methyl- and 3-methyl-L-histidine, these researchers predicted an upfield $^{13}C^{\delta 2}$ deprotonation shift of −2.1 ppm for tautomer **1** and downfield +7.1 ppm for tautomer **2**—a differential of 9.2 ppm. This assumes the methyl group to be a valid proton substitute for ^{13}C chemical shift contributions. Our low-temperature results (Table 2S) for monomeric [U-$^{13}C,^{15}N$]-L-His confirm these trends, showing an even higher tautomer difference of 11.1 ppm. $^{13}C^{\gamma}$ chemical shifts produce about −8 ppm tautomer difference,[10] which has been successfully employed for His tautomer discrimination in proteins by direct ^{13}C observation in very large samples of natural isotopic abundance ribonuclease,[11] but this method lacks general applicability.

[†] Current address: Pulmonary and Critical Care Division, New England Medical Center/Tupper Research Institute, Boston, MA 02111.

10.1021/ja034072c CCC: $25.00

α-Lytic protease (MW = 19.8) contains only one His, the catalytic His 57, whose $^{13}C^{\delta 2}$ resonance undergoes a large downfield shift, from 120.6[3] to the 126.8 ppm shown in Figure 1 (the same increase of 6.2 ppm as in subtilisin BPN′) during a pH increase from 4.0 to 9.0. Thus the presence of $N^{\delta 1}$ tautomer (**2**) is easily distinguished on the basis of $^{13}C^{\delta 2}$ chemical shifts, providing further confirmation of the assignments of the $C^{\delta 2}-^{1}H$ (and indirectly the intensity-paired $C^{\epsilon 1}-^{1}H$) protons of His 57 as the authentic catalytic triad signals.[12] The resonance of the major α-LP autolysis product, which grows steadily with time, is shown in Figure 1 and Table 1S. This $C^{\delta 2}-^{1}H$ proton resonance, ranging from 7.35 ppm at pH 4 to 6.93 ppm at pH 9 and its companion $C^{\epsilon 1}-^{1}H$ proton ranging from 8.65 ppm at pH 4 to 7.70 at pH 9 had been accepted for decades as the authentic His 57 signal (called "fresh"[13]), but from the $^{13}C^{\delta 2}$ chemical shift is shown here to be primarily the wrong tautomer **1**. Correct assignment of the anomalous ^{1}H resonances in α-LP was a vital step in confirming in solution[12] the presence of a $C^{\epsilon 1}-H$-donated H-bond predicted from X-ray crystallographic data,[14] and led to a proposed new reaction-driven imidazole ring flip mechanism in serine protease catalysis.[12]

Of the six His residues in TBD (MW = 15.4), five $C^{\delta 2}$ nuclei resonate between 116.7 and 118.9 ppm at pH 5.5, and the sixth, His 187, the only tautomer **2**, appears at 128.1 ppm (Figure 1, panel B). The only protein in this communication prepared by uniform $^{13}C,^{15}N$-labeling of all residues, TBD contains no tryptophans, otherwise interference from Trp $^{1}H/^{13}C^{\delta 1}$ might occur. Originally the $^{13}C^{\delta 2}$ carbon and $C^{\delta 2}-^{1}H$ proton resonances of His 187 were unassigned,[15,16] in part due to degeneracy caused by the unusual $C^{\epsilon 1}-H$ proton chemical shift (6.97 ppm). Thus, we assumed His 187 to be tautomer **1**, leading to a reversal of the $^{15}N^{\delta 1}$ and $^{15}N^{\epsilon 2}$ assignments.[16] ^{15}N chemical shifts are generally the most informative NMR probe of His protonation, tautomerism, and H-bonding, whether obtained by direct ^{15}N observation[3] or indirect observation using either 2D $^{1}H/^{15}N$ two-bond correlation[17,18] or 3D $^{1}H/^{13}C/^{15}N$ "HCN" sequential one-bond correlation.[16] In the case of TBD His 187, however, the $^{13}C^{\delta 2}$ carbon chemical shift elucidated the tautomeric equilibrium, which in turn corrected the ^{15}N assignments.

Tautomer identification by direct observation of histidine ^{15}N requires large samples, ≥1.5 mL of ∼2 mM protein, but has the advantage that line widths of 100 Hz or more due to high molecular weight, aggregation, or acid−base fast-exchange broadening[19] can be tolerated.[1] The indirect ^{15}N observation methods offer the advantage of higher ^{1}H sensitivity and multiple correlation of intraresidue ^{1}H, ^{15}N, and for HCN, ^{13}C resonances. A disadvantage is their reliance upon relatively small spin-coupling constants, i.e., $^{2}J_{NH} \leq |11\ Hz|$ and, for HCN, $^{1}J_{CN} \leq |16\ Hz|$, which limits the tolerable line widths.

In summary, if the $^{13}C^{\delta 2}$ chemical shift of neutral ("high pH") His is >122 ppm, a predominance of $N^{\delta 1}-H$ tautomer (**2**) is indicated; if <122 ppm, mostly $N^{\epsilon 2}-H$ tautomer (**1**) is indicated. The method is easy to implement, requiring only bioincorporation of [U-^{13}C] (or the more readily available [U-^{13}C, ^{15}N])-histidine. Standard HMQC or HSQC NMR pulse programs then yield the chemical shifts of His $^{13}C^{\delta 2}$ with the benefit of high ^{1}H sensitivity. Because of the large one-bond spin-coupling ($^{1}J_{CH} \approx 200$ Hz), the method should extend to proteins having large ^{1}H and ^{13}C line widths, including very high molecular weights. A TROSY HSQC pulse program specifically for aromatic $^{1}H/^{13}C$ correlation[20] is available for very large proteins and higher magnetic fields, although it was not necessary for the proteins up to 27.5 kD at 600 MHz in this work.

Acknowledgment. This work was funded in part by NIH Grant GM55397. The Bruker AMX500 NMR spectrometer was funded by the NIH, and the Bruker DRX600, by an NIH/NSF Shared Instrumentation Grant. R.M.D. is the recipient of a National American Heart Association Scientist Development Grant.

Supporting Information Available: Table 1S lists ^{1}H and $^{13}C^{\delta 2}$ chemical shifts, pK_a values, and other data on the His residues of the three proteins presented here; Table 2S lists all NMR parameters of monomeric [U-^{13}C,^{15}N]-His in 80% ethanol (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bachovchin, W. W.; Roberts, J. D. *J. Am. Chem. Soc.* **1978**, *100*, 8041−8047.
(2) Bachovchin, W. W. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7948−7951.
(3) Bachovchin, W. W. *Magn. Reson. Chem.* **2001**, *39*, S199−S213.
(4) Alden, R. A.; Birktoft, J. J.; Kraut, J.; Robertus, J. D.; Wright, C. S. *Biochem. Biophys. Res. Commun*. **1971**, *45*, 337−344.
(5) Farr-Jones, S.; Wong, W. Y. L.; Gutheil, W. G.; Bachovchin, W. W. *J. Am. Chem. Soc.* **1993**, *115*, 6813−6819.
(6) Luo, X.; Sanford, D. G.; Bullock, P. A.; Bachovchin, W. W. *Nat. Struct. Biol.* **1996**, *3*, 1034−1039.
(7) Vuister, G.; Bax, A. *J. Magn. Reson.* **1992**, *98*, 428−435.
(8) 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.* **2003**, *12*, 794−810.
(9) Reynolds, W. F.; Peat, I. R.; Freedman, M. H.; Lyerla., J. R., Jr. *J. Am. Chem. Soc*. **1973**, *95*, 328−331.
(10) Goux, W.; Allerhand, A. *J. Biol. Chem.* **1979**, *254*, 2210−2213.
(11) Walters, D. E.; Allerhand, A. *J. Biol. Chem*. **1980**, *255*, 6200−6204.
(12) Ash, E. L.; Sudmeier, J. L.; Day, R. M.; Vincent, M. P.; Torchilin, E. V.; Haddad, K. C.; Bradshaw, E. M.; Sanford, D. G.; Bachovchin, W. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10371−10376.
(13) Markley, J. L.; Neves, D. E.; Westler, W. M.; Ibanez, I. B.; Porubcan, M. A.; Baillargeon, M. W. *Front. Protein Chem.* **1980**, *19*, 31−61.
(14) Derewenda, Z. S.; Derewenda, U.; Kobos, P. M. *J. Mol. Biol.* **1994**, *241*, 83−93.
(15) Luo, X. Ph.D. Dissertation, Tufts University, Sackler School, Boston, MA, 1997.
(16) Sudmeier, J. L.; Ash, E. L.; Gunther, U. L.; Luo, X.; Bullock, P. A.; Bachovchin, W. W. *J. Magn. Reson. B* **1996**, *113*, 236−247.
(17) Van Dijk, A. A.; Scheek, R. M.; Dijkstra, L.; Wolters, G. K.; Robillard, G. T. *Biochemistry* **1992**, *31*, 9063−9072.
(18) Pelton, J. G.; Torchia, D. A.; Meadow, N. D.; Roseman, S. *Protein Sci.* **1993**, *2*, 543−558.
(19) Sudmeier, J. L.; Evelhoch, J. L.; Jonsson, N. B.-H. *J. Magn. Reson.* **1980**, *40*, 377−390.
(20) Meissner, A.; Sorensen, O. W. *J. Magn. Reson.* **1999**, *139*, 439−442.

JA034072C