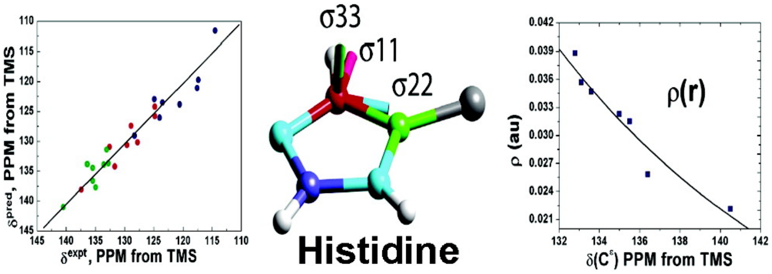
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

A Solid State C NMR, Crystallographic, and Quantum Chemical Investigation of Chemical Shifts and Hydrogen Bonding in Histidine Dipeptides

Feng Cheng, Haihong Sun, Yong Zhang, Dushyant Mukkamala, and Eric Oldfield

J. Am. Chem. Soc., **2005**, 127 (36), 12544-12554• DOI: 10.1021/ja051528c • Publication Date (Web): 11 August 2005



ACS Publications
High quality. High impact.

Journal of the American Chemical Society is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

J|A|C|S
ARTICLES

A Solid State ^{13}C NMR, Crystallographic, and Quantum Chemical Investigation of Chemical Shifts and Hydrogen Bonding in Histidine Dipeptides

Feng Cheng,[†] Haihong Sun,[‡] Yong Zhang,[†] Dushyant Mukkamala,[‡] and
Eric Oldfield*,[†,‡]

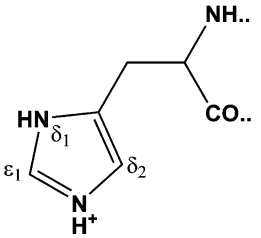
Contribution from the Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, and Department of Biophysics, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Received March 9, 2005; E-mail: eo@chad.scs.uiuc.edu

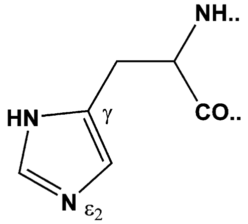
Abstract: We report the first solid-state NMR, crystallographic, and quantum chemical investigation of the origins of the ^{13}C NMR chemical shifts of the imidazole group in histidine-containing dipeptides. The chemical shift ranges for C^{γ} and C^{δ_2} seen in eight crystalline dipeptides were very large (12.7−13.8 ppm); the shifts were highly correlated ($R^2 = 0.90$) and were dominated by ring tautomer effects and intermolecular interactions. A similar correlation was found in proteins, but only for buried residues. The imidazole ^{13}C NMR chemical shifts were predicted with an overall rms error of 1.6−1.9 ppm over a 26 ppm range, by using quantum chemical methods. Incorporation of hydrogen bond partner molecules was found to be essential in order to reproduce the chemical shifts seen experimentally. Using AIM (atoms in molecules) theory we found that essentially all interactions were of a closed shell nature and the hydrogen bond critical point properties were highly correlated with the N···H···O (average $R^2 = 0.93$) and N^{ϵ_2}···H···N (average $R^2 = 0.98$) hydrogen bond lengths. For C^{ϵ_1}, the ^{13}C chemical shifts were also highly correlated with each of these properties (at the N^{ϵ_2} site), indicating the dominance of intermolecular interactions for C^{ϵ_1}. These results open up the way to analyzing ^{13}C NMR chemical shifts, tautomer states (from C^{δ_2}, C^{ϵ_1} shifts), and hydrogen bond properties (from C^{ϵ_1} shifts) of histidine residue in proteins and should be applicable to imidazole-containing drug molecules bound to proteins, as well.

Introduction

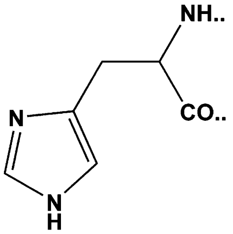
Histidine is an important amino acid in proteins since it is involved in catalysis by, for example, serine and cysteine proteases, as well as being a frequent ligand to metals such as Fe, Cu, and Zn.[1−5] Histidine can exist in four different forms: a protonated, imidazolium form (**1**), two neutral tautomers containing N^{δ_1}−H (**2**), or N^{ϵ_2}−H (**3**), as well as a formally anionic imidazolate form (**4**):



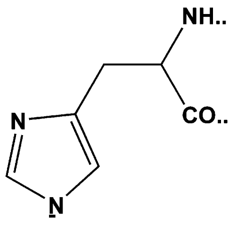
1



2



3



4

Nuclear magnetic resonance (NMR) spectroscopy is a potentially powerful tool with which to investigate these protonation states in proteins, via ^{1}H, ^{13}C, and ^{15}N chemical shifts[6−8] as well as from ^{13}C−^{15}N scalar or *J*-couplings,[9] and in early work on model systems it was proposed that the chemical shifts of $^{13}C^{\gamma}$ in **2** were about 2 ppm more shielded than those of **1**,

---

[†] Department of Chemistry.

[‡] Department of Biophysics.

(1) Fuhrmann, C. N.; Kelch, B. A.; Ota, N.; Agard, D. A. *J. Mol. Biol.* **2004**, *338*, 999.

(2) Dardenne, L. E.; Werneck, A. S.; de Oliveira Neto, M.; Bisch, P. M. *Proteins* **2003**, *52*, 236.

(3) Vallone, B.; Nienhaus, K.; Matthes, A.; Brunori, M.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17351.

(4) Firi, A. V.; Anishetty, S.; Gautam, P. *BMC Bioinf.* **2004**, *5*, 127.

(5) Lipton, A. S.; Heck, R. W.; Ellis, P. D. *J. Am. Chem. Soc.* **2004**, *126*, 4735.

(6) Wüthrich, K. *NMR in Biological Research: Peptides and Proteins*; North-Holland: Amsterdam, 1976.

(7) (a) Reynolds, W. F.; Peat, I. R.; Freedman, M. H.; Lyerla, J. R., Jr. *J. Am. Chem. Soc.* **1973**, *95*, 328. (b) Deslauriers, R.; McGregor, W. H.; Sarantakis, D.; Smith, I. C. P. *Biochemistry* **1974**, *13*, 3443.

(8) Harbison, G.; Herzfeld, J.; Griffin, R. G. *J. Am. Chem. Soc.* **1981**, *103*, 4752.

(9) (a) Shimba, N.; Takahashi, H.; Sakakura, M.; Fujii, I.; Shimada, I. *J. Am. Chem. Soc.* **1998**, *120*, 10988. (b) Shimba, N.; Serber, Z.; Ledwidge, R.; Miller, S. M.; Craik, C. S.; Dotsch, V. *Biochemistry* **2003**, *42*, 9227.

10.1021/ja051528c CCC: $30.25

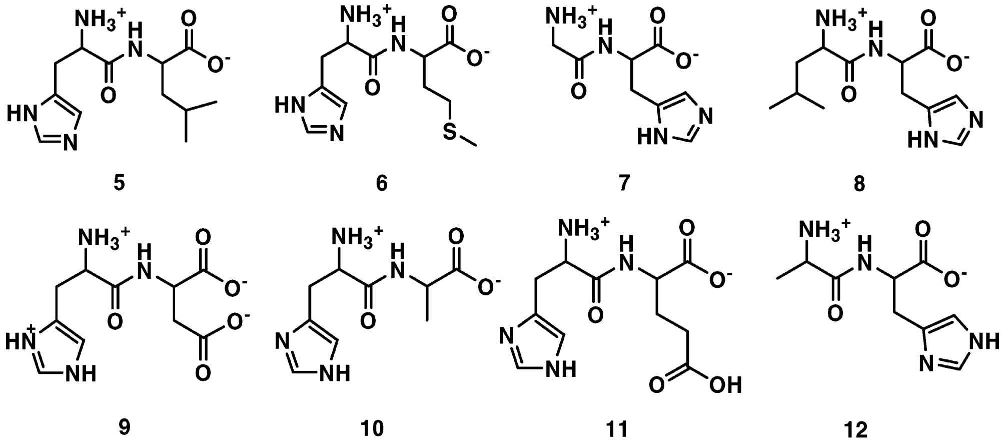


Figure 1. Structures of the histidine-containing dipeptides investigated. **5**, His-Leu; **6**, His-Met; **7**, Gly-His; **8**, Leu-His; **9**, His-Asp; **10**, His-Ala; **11**, His-Glu; **12** Ala-His.

while the ^{13}C$^{\gamma}$ shifts of **3** were about 6 ppm more deshielded.[7] These observations then led to investigations of ^{13}C$^{\gamma}$ shifts in proteins, such as myoglobin[10] and plastocyanin,[11] where it was found that pK_a values could be deduced from the experimentally observed shifts. The range of chemical shifts observed in proteins was, however, rather small, about 6 ppm, due most likely to the relatively small number of residues investigated. With the advent of multidimensional NMR and isotopic labeling, many more shifts (of C$^{\gamma}$) have now been reported, as have other (C$^{\delta_2}$, C$^{\epsilon_1}$) shifts, and the overall shift ranges in proteins for C$^{\gamma}$ and C$^{\delta_2}$ are now much larger, some 14 ppm,[12] a range making quantum chemical analysis of such shifts of interest, from the perspective of both understanding the origins of these shifts and, potentially, providing detailed insights into hydrogen bonding interactions of histidine residues in proteins. In earlier work, Wei et al.[13] reported the results of quantum chemical investigations of the ^{15}N NMR spectra of HisLeu and a series of histidine salts but noted[13] that there were few structures reported for the neutral histidine species (**2**,**3**) of importance in enzyme catalysis. Likewise, the results of quantum chemical calculations of ^{13}C NMR chemical shifts in the amino acid histidine have been reported,[14−16] but these results were compared with only one experiment (on an imidazolium form) or with theory alone. Clearly then, there is a need to investigate a much broader range of well-characterized histidine-containing peptides (whose chemical shifts should ideally cover the same large ranges as those observed in proteins), to obtain a better understanding of these shifts and to begin to make use of them in studies of structure and, potentially, electrostatics. In this work, we have therefore investigated the solid-state ^{13}C NMR spectra of eight histidine-containing dipeptides, comprising four N$^{\delta_1}$−H tautomers (**2**), three N$^{\epsilon_2}$−H tautomers (**3**), plus one imidazolium species (**1**). The structures of all eight species were determined crystallographically (six being new structures), and these structures were then used in an investigation of the ^{13}C NMR shieldings and electrostatics in each compound, using quantum chemistry. Unlike previous ^{13}C NMR studies of peptides and proteins,[17−20] tautomeric state and hydrogen bonding interactions make major contributions to the large chemical shift ranges seen experimentally. We also investigate the nature of these hydrogen bonding interactions using atoms-in-molecules (AIM) theory,[21] as a prelude to related studies in proteins.

Experimental Section

Crystallographic Aspects. We crystallized the eight His-containing species (**5**−**12**) whose structures are shown in Figure 1. All the compounds were obtained from Bachem (King of Prussia, PA). **5**, **6**, **8**, and **9** were crystallized from water while **7** and **10**−**12** were crystallized from aqueous EtOH. Diffraction data for all eight compounds were collected at 193 K on a Bruker SMART CCD system. Data reduction and integration were performed with the software package SAINT,[22] and absorption corrections were applied by using the program SADABS.[22] The positions of the non-hydrogen atoms were found by direct methods using the Bruker SHELXTL software package.[22] The same crystal batches were used in the NMR experiments. Crystal data and structure refinement information for all of the new structures are shown in Table 1. The structures of **5** and **10** have been reported previously[23,24] and are in good accord with the results we obtained (not shown). More detailed descriptions of the six novel structures (coordinates, geometries, hydrogen bonding patterns, and B factors) are provided in the Supporting Information, Tables S1−S42.

NMR Spectroscopy. Carbon-13 NMR spectra were obtained by using the cross-polarization magic-angle sample-spinning technique[25,26] with either full proton decoupling[27] or interrupted decoupling[28] (using a dipolar dephasing time of 150 μs), for selection of the nonprotonated

---

(10) Wilbur, D. J.; Allerhand, A. *J. Biol. Chem.* **1977**, *252*, 4968.
(11) Ugurbil, K.; Norton, R. S.; Allerhand, A.; Bersohn, R. *Biochemistry* **1977**, *16*, 886.
(12) Seavey, B. R.; Farr, E. A.; Westler, W. M.; Markley, J. L. *J. Biomol. NMR* **1991**, *1*, 217. BioMagResBank (BMRB), a NIH funded bioinformatics resource, Department of Biochemistry, University of Wisconsin-Madison, Madison, WI USA (URL: http://www.bmrb.wisc.edu).
(13) Wei, Y.; de Dios, A. C.; McDermott, A. E. *J. Am. Chem. Soc.* **1999**, *121*, 10389.
(14) Strohmeier, M.; Stueber, D.; Grant, D. M. *J. Phys. Chem. A* **2003**, *107*, 7629.
(15) Moon, S.; Christiansen, P. A.; DiLabio, G. A. *J. Chem. Phys.* **2004**, *120*, 9080.
(16) Chen, X.; Zhan, C. G. *THEOCHEM* **2004**, *682*, 73.
(17) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *261*, 535.
(18) Sun, H.; Sanders, L. K.; Oldfield, E. *J. Am. Chem. Soc.* **2002**, *124*, 5486.
(19) Sun, H.; Oldfield, E. *J. Am. Chem. Soc.* **2004**, *126*, 4726.
(20) (a) Xu, X. P.; Case, D. A. *Biopolymers* **2002**, *65*, 408. (b) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321.
(21) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, UK, 1990.
(22) (a) *SAINT*; Bruker AXS, Inc.: Madison, WI, 2001. (b) *SADABS*; Bruker AXS, Inc.: Madison, WI, 2001. (c) *SHELXTL*; Bruker AXS, Inc.: Madison, WI, 2001.
(23) Krause, J. A.; Baures, P. W.; Eggleston, D. S. *Acta Crystallogr., Sect. B* **1993**, *49*, 123.
(24) Steiner, T. *Acta Crystallogr., Sect. C* **1996**, *52*, 2554.
(25) Pines, A.; Gibby, M. G.; Waugh, J. S. *J. Chem. Phys.* **1973**, *59*, 569.
(26) Schaefer, J.; Stejskal, E. O. *J. Am. Chem. Soc.* **1976**, *98*, 1031.

Table 1. Crystallographic Data Summary for Histidine Dipeptides

	His-Met (6)	Gly-His (7)	Leu-His (8)	His-Asp (9)	His-Glu (11)	Ala-His (12)
formula	$C_{11}H_{18}N_4O_3S$	$C_8H_{16}N_4O_5$	$C_{12}H_{22}N_4O_4$	$C_{10}H_{20}N_4O_8$	$C_{11}H_{16}N_4O_5$	$C_{22}H_{42}N_8O_9$
mol wt	286.35	248.25	286.34	324.30	284.28	562.64
cryst syst	monoclinic	monoclinic	orthorhombic	monoclinic	monoclinic	orthorhombic
space group	$P2_1$	$P2_1$	$P2_12_12$	$P2_1$	$P2_1$	$P2_12_12$
a (Å)	6.5893(13)	6.0389(7)	12.240(3)	4.8451(18)	4.9316(7)	9.283(2)
b (Å)	5.4676(11)	7.2672(9)	21.081(5)	18.326(7)	15.472(2)	18.035(4)
c (Å)	18.392(3)	12.9258(15)	6.1137(14)	8.678(3)	8.6523(13)	8.549(2)
α (deg)	90.00	90.00	90.00	90.00	90.00	90.00
β (deg)	96.499(6)	99.908(2)	90.00	104.029(7)	99.102(4)	90.00
γ (deg)	90.00	90.00	90.00	90.00	90.00	90.00
V (Å^3)	658.4(2)	558.80(11)	1577.6(6)	747.5(5)	649.84(17)	1431.2(6)
Z	2	2	4	2	2	2
ρ_{calcd} (g cm^{-3})	1.444	1.475	1.206	1.441	1.453	1.306
T (K)	193(2)	193(2)	193(2)	193(2)	193(2)	193(2)
μ (mm^{-1})	0.257	0.123	0.091	0.125	0.116	0.102
R1 (all data)	0.0664	0.0343	0.0337	0.0481	0.0338	0.0928
wR2 (all data)	0.1338	0.0935	0.0877	0.1235	0.0666	0.0857
crystallization	H_2O	H_2O/EtOH	H_2O	H_2O	H_2O/EtOH	H_2O/EtOH

His aromatic carbon (C^γ). Spectra were typically recorded using a 5 s recycle time. The ^1H and ^{13}C 90° pulse widths both were 2.75 μs. Chemical shifts were referred to external glycine, setting the α carbon to 43.6 ppm downfield from tetramethylsilane (TMS). All spectra were obtained by using a Varian (Palo Alto, CA) Infinityplus 600 MHz (^1H) NMR spectrometer at an 8 kHz spinning frequency using a 3.2 mm Varian/Chemagnetics HXY probe.

Computational Aspects. To compute the ^{13}C NMR chemical shifts (or shieldings), we carried out four different sets of calculations. In each case, we used the Hartree−Fock method as incorporated in Gaussian 98,[29] using the locally dense basis set scheme used previously (6-311++G(2d,2p)/6-311G) with the denser basis set on the atoms of interest (and their nearest neighbors), combined with the gauge-including atomic orbitals (GIAO) method.[30] In the first set of calculations, we used the N-formyl-X-amide model approach described previously,[17] with torsion angles set to the X-ray values, as shown for example for HisLeu (**5**) in Figure 2A. Since the results obtained were only modest, we next used the monomers found in the X-ray crystal structure, as shown for example for **5** in Figure 2B. Third, we used cluster models, as shown again for **5** in Figure 2C, in which the effects of neighboring residues were included by incorporating methylimidazole, acetate, or methylammonium ions, as appropriate, to represent the histidine's lattice partners. The structures of each of these eight "supermolecule" clusters are shown in the Supporting Information (Figure S1). And finally, we carried out a geometry optimization at the hydrogen atom positions (using HF/3-21G) in each of the supermolecule clusters, to see to what extent improvements in the shielding predictions might be made. In addition, we also carried out DFT calculations with the cluster models, using the B3LYP functional[31] and a locally dense basis set scheme (6-311++G(d,p)/6-31G) with the denser basis set again on the atoms of interest (and their nearest neighbors). The hydrogen bond critical point (BCP) properties were evaluated by AIM2000[32] using the wave functions from the HF and B3LYP cluster calculations.

Results and Discussion

We chose to investigate the solid-state ^{13}C NMR shifts in eight histidine-containing dipeptides (**5**−**12**) to provide a basis for chemical shielding calculations of the origins of the C^γ, C^{δ_2} and C^{ϵ_1} shifts in peptides and in proteins. We determined the structures of all eight dipeptides using X-ray crystallography and used the same crystal batches for our NMR investigations. Of these eight molecules, the structures of six are novel and we show in Table 1 a summary of the crystallographic parameters for each, together with, in Figure 3, their crystallographic structures displayed by using the ORTEP program.[33] Among the eight structures investigated by NMR, four (**5**−**8**) exist in the N^{δ_1}−H (also known as the π) tautomeric state (**2**), **9** is in

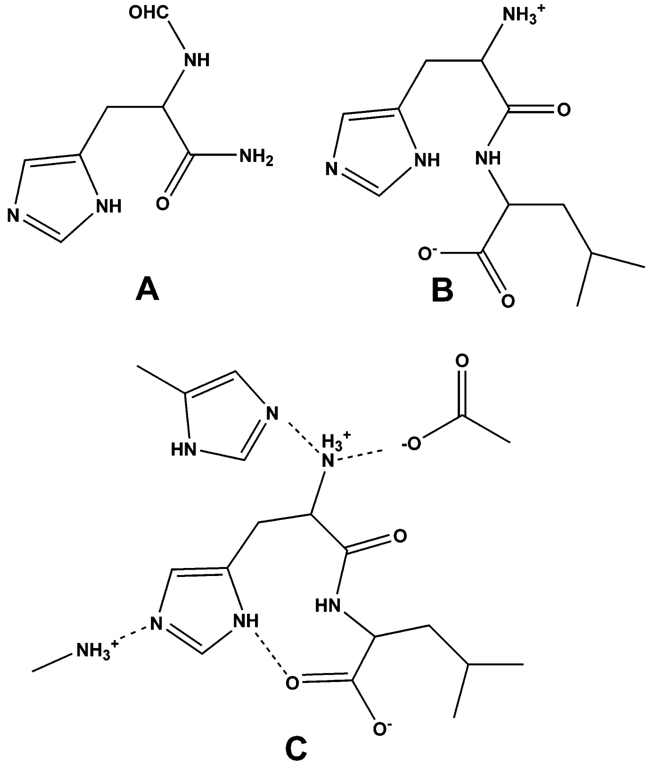
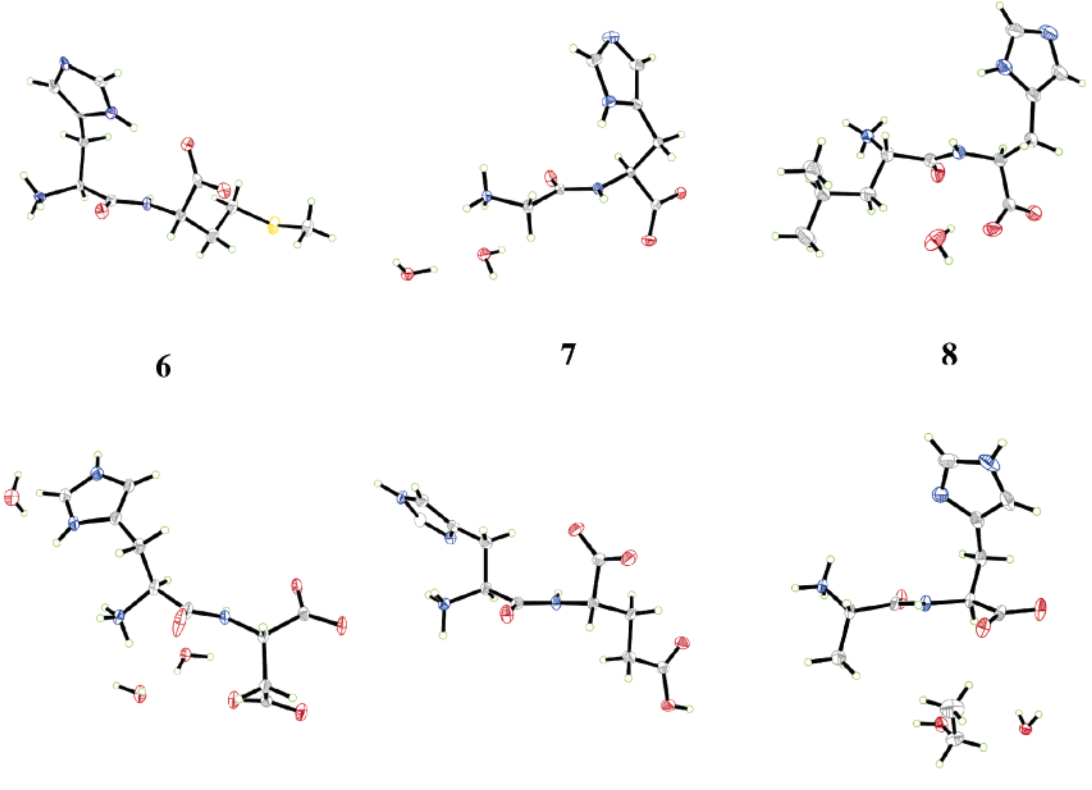


Figure 2. Models used for calculations: (A) N-formyl-histidine amide; (B) monomer from crystal structure of His-Leu (**5**); (C) monomer from crystal structure of His-Leu (**5**) with surrounding hydrogen bond partners included.

---

(27) Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G. *J. Chem. Phys.* **1995**, *103*, 6951.
(28) Opella, S. J.; Frey, M. H. *J. Am. Chem. Soc.* **1979**, *101*, 5855.
(29) Frisch, M. J. et al. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
(30) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789.
(31) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(32) Biegler-König, F. *AIM2000*, version 1.0; University of Applied Science: Bielefeld, Germany.
(33) Johnson, C. K. *ORTEP-II: A FORTRAN Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations*; Oak Ridge National Laboratory Report ORNL-5138, 1976.



6 **7** **8**

9 **11** **12**

Figure 3. Crystal structures of compounds **6**, **7**, **8**, **9**, **11**, and **12**.

the imidazolium form (**1**), while **10**−**12** are in the N^{ϵ_2}−H (also known as the τ) tautomeric state. In each system, there is a complex network of hydrogen bonds to both N^{δ_1} and N^{ϵ_2} (see Figure S1 for more details).

Next, we recorded the ^{13}C magic-angle sample-spinning (MAS) NMR spectra of each system under conditions of full (TPPM)[27] proton decoupling or using interrupted decoupling,[28] which in the aromatic region selects for the C^{γ} resonance. The most highly shielded aromatic peak was then assigned in each case to C^{δ_2} ($\delta_{\text{average}} = 120.9 \pm 4.6$ ppm from TMS), while the most deshielded aromatic methine carbon was assigned in each case to C^{ϵ_1} ($\delta_{\text{average}} = 135.3 \pm 2.5$ ppm). Some representative spectra, of **5** and **12**, are shown in Figure 4, and the experimentally observed shifts for each compound are presented in Table 2. Also shown in Table 2 are the solution NMR shifts for histidine and a protected histidine peptide, which are found to be very close to the average values found for the eight crystalline dipeptides.

On examination of the dipeptide shifts, Table 2, it can immediately be seen that there are large chemical shift ranges, particularly so for C^{γ} ($\Delta\delta = 12.7$ ppm) and for C^{δ_2} ($\Delta\delta = 13.8$ ppm). These chemical shift ranges are very close to those found in proteins,[12] as shown for example in Figure 5 for C^{γ}, strongly suggesting that the interactions which dominate ^{13}C shielding in these histidine peptides are the same as those which dominate shielding in proteins. It is also of interest to note that the actual values of the C^{γ} and C^{δ_2} shifts are highly correlated. This is shown graphically in Figure 6A where we find an $R^2 = 0.90$ (with a slope of -1.03). For C^{ϵ_1}, which has a much smaller shift range ($\Delta\delta = 7.7$ ppm), no correlations with the C^{γ}, C^{δ_2} shifts are seen, presumably since C^{ϵ_1} is not connected to either atom and has different factors which dominate its shielding. The C^{γ}−C^{δ_2} shift correlation seen in the peptides is also seen in proteins, as shown in Figure 6B (and Table S43), where we present C^{γ}, C^{δ_2} shifts for eight proteins (from the BioMagRes Bank;[12] PDB File nos.: 1RCF (recombinant oxidized flavodoxin); 1HOE (α-amylase inhibitor Hoe-467A); 1EY7 (Staphylococcal nuclease); 1HG6 (Microcin J25); 1QH7(xylanase); 5FX2 (flavodoxin); 1EIA (eiav capsid protein); and 1EHK (ba3-type cytochrome-c oxidase)).[34] However, the "raw" correlation (for all data points) in proteins is less clear than with the peptide data. This effect could be due in part to small differences in chemical shift referencing between the different protein studies, but in addition to this effect it appears that there is in fact a cluster of residues which all have similar C^{γ}, C^{δ_2} shifts falling slightly above the trend line. On inspection of the X-ray crystallographic structures, we found that all these residues correspond to solvent exposed histidines, and this is shown more graphically in Figure 6B in which histidine residues having solvent-exposed surface areas of >50 Å^2 (computed using the POPS program, ref 35) are shown in red. This type of behavior has been reported previously by us for C^{γ} of solvent-exposed Trp residues in proteins[19] and may be due to both solvation and enhanced mobility.[19] The centroid of the cluster of these solvent exposed residues is at ~131 (C^{γ}), 119 (C^{δ_2}) ppm, the same as that found for histidine or a protected histidine peptide, in solution (Table 2), confirming this idea.

---

(34) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(35) Fraternali, F.; Cavallo, L. *Nucleic Acids Res.* **2002**, *30*, 2950−2960.

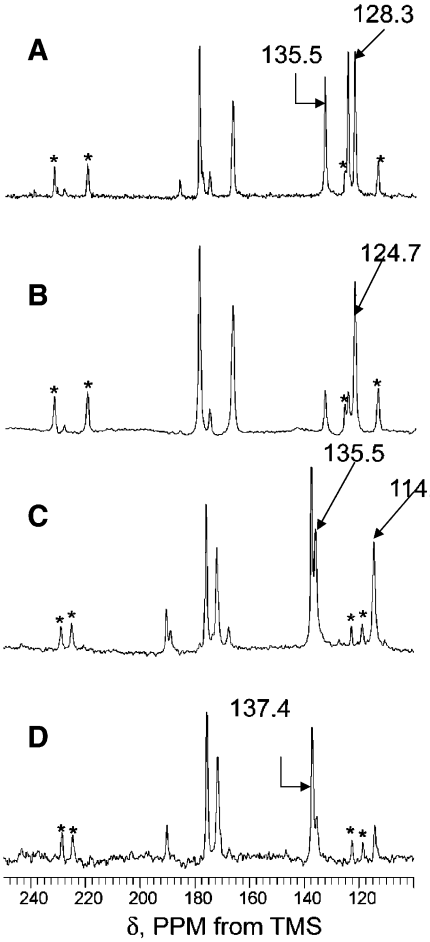


Figure 4. 125 MHz ^{13}C MAS NMR spectra of **5** and **12**. (A) **5**, fully proton decoupled and (B) with interrupted decoupling. (C) **12**, fully proton decoupled and (D) with interrupted decoupling. The His C^{γ} chemical shift range is 12.7 ppm. * = carbonyl group spinning-sideband.

Next, we investigated the use of quantum chemical methods to see to what extent the experimental shifts in the eight dipeptides could be predicted and questions as to the origins of these shifts, probed in more detail. We first used the N-formyl-histidine amide model approach employed previously[17] to compute, primarily, C^{α} and C^{β} shifts in peptides and proteins but found relatively poor accord with experiment. The *R*-values were not unreasonable (0.82, 0.94, and 0.80 for C^{γ}, C^{δ_2}, and C^{ϵ_1}, respectively, to be compared with ideal values of 1.0), but the slopes were poor: −2.17, −2.28, and 0.42, for C^{γ}, C^{δ_2}, and C^{ϵ_1} (Table 3), to be compared with ideal values of −1.00. Full results are given in the Supporting Information (Table S44). The predicted C^{γ} and C^{δ_2} shifts from this simple model were very highly correlated, as shown in Figure 6C ($R^2 = 0.98$) with a slope (−1.02) similar to that seen experimentally in the eight histidine dipeptides (−1.03), but as can be seen in Figure 6C the shifts of the N^{δ_1}−H and N^{ϵ_2}−H tautomers form rather pronounced clusters corresponding to the π and τ tautomeric forms and do not reflect the broad distribution of shifts seen experimentally, Figure 6A. In addition, the apparently good agreement with the experimental slope is illusory, since the computed values are both in error by factors of ~2, Table 3, and these errors cancel, in Figure 6C. Thus, while these results do indicate the dominance of tautomeric state on shielding, they also imply the omission of an essential ingredient in the calculations, for example, the presence of terminal NH_3^+ and CO_2^- groups or other electrostatic effects (in proteins). So, we next used single, isolated monomer molecules from the crystal structures, but again found no marked improvement (Table 3; all shielding values are given in Table S44). These results indicated the desirability of incorporating additional, intermolecular interactions in the calculations. To do this, we incorporated the effects of nearest-neighbor carboxylate groups with

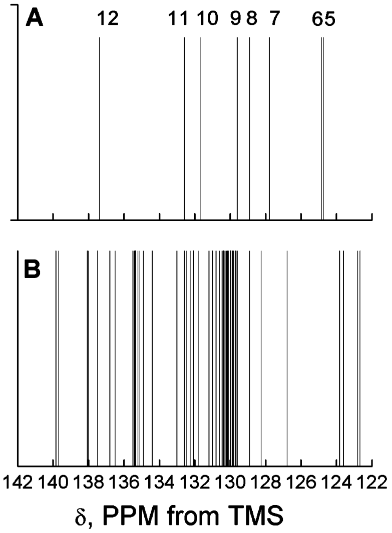


Figure 5. Schematic illustrating the His C^{γ} chemical shifts observed in: (A) model systems (**5**−**12**) and (B) in proteins. Protein chemical shift data from the BMRB database (http://www.bmrb.wisc.edu/).

Table 2. Experimental Chemical Shifts and Computed Chemical Shieldings for C^{γ}, C^{δ_2}, and C^{ϵ_1}[a]

	C^{γ} (ppm)					C^{δ_2}(ppm)					C^{ϵ_1} (ppm)				
		HF		B3LYP			HF		B3LYP			HF		B3LYP	
compound	δ^{expt}	σ^{calc}	δ^{pred}	σ^{calc}	δ^{pred}	δ^{expt}	σ^{calc}	δ^{pred}	σ^{calc}	δ^{pred}	δ^{expt}	σ^{calc}	δ^{pred}	σ^{calc}	δ^{pred}
5	124.7	58.7	123.4	48.1	123.2	128.3	63.2	127.3	54.3	127.2	135.5	43.4	136.6	43.7	136.5
6	124.9	57.1	124.9	46.8	124.3	124.9	69.8	121.3	60.3	122.2	135.0	41.7	137.7	41.7	137.6
7	127.8	52.4	129.3	42.5	128.0	124.0	66.5	124.3	57.4	124.6	133.6	47.6	133.9	47.6	134.2
8	129.6	51.9	129.7	41.6	128.7	120.1	69.2	121.8	59.6	122.8	133.1	51.5	131.4	51.4	131.9
9	128.9	55.3	126.5	41.7	128.6	120.6	68.9	122.1	60.5	122.1	132.8	48.0	133.6	51.2	132.1
10	131.7	48.1	133.3	36.3	133.2	117.4	73.4	118.0	66.1	117.4	136.4	47.7	133.8	47.7	134.1
11	132.6	51.6	130.0	38.8	131.1	117.6	71.9	119.4	64.4	118.8	140.5	36.7	141.0	36.9	140.5
12	137.4	44.0	137.1	32.8	136.2	114.5	82.2	109.9	74.1	110.8	135.5	46.8	134.4	47.1	134.5
average (**5-12**)	**129.7**					**120.9**					**135.3**				
His (pH = 7.0)[b]	131.0					117.7					136.7				
TFA·GGHA·OMe[b]	130.3					118.7					135.2				

[a] Calculated with exact crystal structures with surrounding hydrogen bond partner molecules included. The basis sets for the calculations are described in the Experimental Section of the text. The predicted shifts were from eqs 1−6. [b] His, protected His peptide shifts in solution, from ref 6.

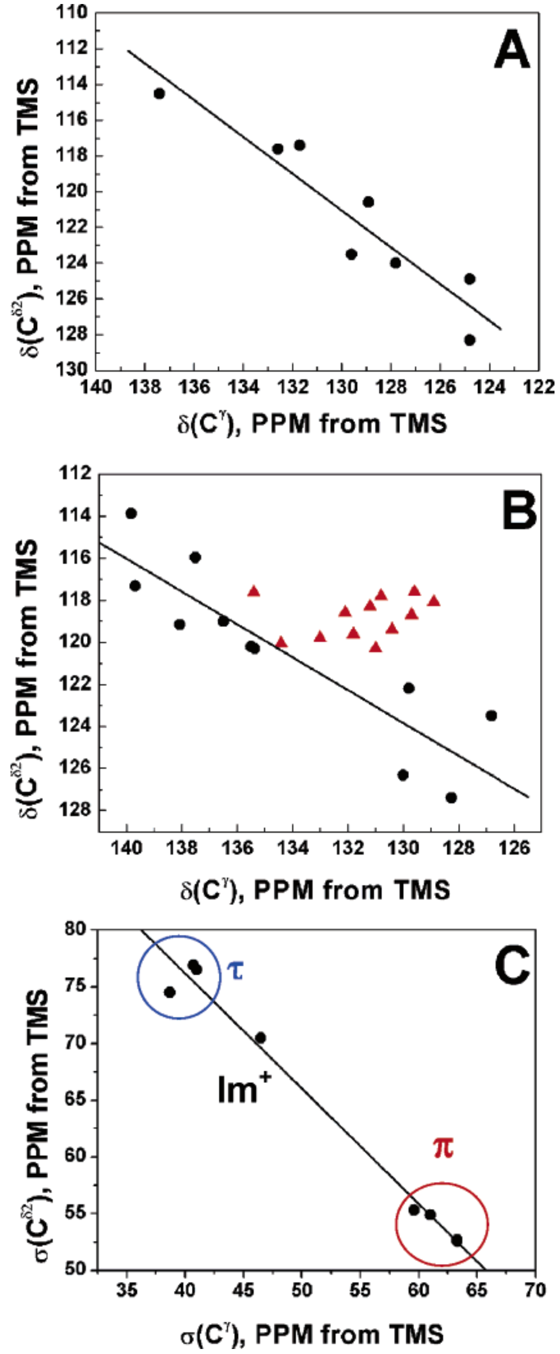


Figure 6. (A) experimental shift correlation between C^{γ} and C^{δ_2} for the eight dipeptides. (B) Experimental shift correlation between C^{γ} and C^{δ_2} for proteins. The red points are for solvent accessible histidines (exposed surface area > 50 Å^2). (C) Computed C^{γ}, C^{δ_2} isotropic chemical shieldings (σ_i) for the eight dipeptides using the N-formyl-histidine amide model. The π and τ tautomeric states stand for N^{δ_1}−H and N^{ϵ_2}−H, respectively.

Table 3. Statistical Results for HF Chemical Shift/Shielding Calculations

atom type	method	R	slope
C^{γ}	FHA[a]	0.82	−2.17
	EX[b]	0.91	−2.80
	HB[c]	0.94	−1.07
	HBMIN[d]	0.86	−0.85
C^{δ_2}	FHA[a]	0.94	−2.28
	EX[b]	0.91	−1.75
	HB[c]	0.88	−1.09
	HBMIN[d]	0.80	−0.67
C^{ϵ_1}	FHA[a]	0.80	0.42
	EX[b]	0.37	−1.09
	HB[c]	0.82	−1.54
	HBMIN[d]	0.72	−1.45

[a] N-formyl-histidine amide model. [b] Exact crystal structures. [c] Exact crystal structures with surrounding hydrogen bond partner molecules included. [d] Same as for footnote c but after geometry optimization of hydrogen bond H-atom positions.

acetate ions, histidine ammonium groups as methylammonium ions, and histidine imidazole rings as methylimidazoles (see for example Figure 2C for a representative cluster; all clusters investigated are shown in the Supporting Information, Figure S1). This protocol gave much better accord with experiment for C^{γ} and C^{δ_2}, as shown in the experimental shift versus theoretical shielding results presented in Figure 7A and 7B, with R values of 0.94 (C^{γ}) and 0.88 (C^{δ_2}) and slopes of −1.07 (C^{γ}) and −1.09 (C^{δ_2}), and greatly improved results ($R = 0.82$, slope $= -1.54$) for C^{ϵ_1} (Figure 7C and Tables 2,3). The predicted shifts can then be obtained from the regression lines using calculated shieldings:

$$\delta^{\mathrm{pred}}\,(C^{\gamma}) = (190.7 - \sigma^{\mathrm{calc}})/1.07 \qquad (1)$$

$$\delta^{\mathrm{pred}}\,(C^{\delta_2}) = (202.0 - \sigma^{\mathrm{calc}})/1.09 \qquad (2)$$

$$\delta^{\mathrm{pred}}\,(C^{\epsilon_1}) = (253.8 - \sigma^{\mathrm{calc}})/1.54 \qquad (3)$$

which then enables a comparison of all 24 predicted shifts (eqs 1−3) with those determined experimentally, Figure 7D, in which the R value is 0.97 and the rms error between the shift predictions and experiment is 1.9 ppm over the entire 26 ppm experimental shift range, corresponding to about a 7% error in shift prediction, on average. The effects of hydrogen atom geometry optimization resulted in little or no improvement to these results, Tables 3 and S44, at a considerable increase in computational time.

In addition to these results we also carried out a series of DFT calculations to investigate the effects of electron correlation on shielding. The B3LYP-calculated chemical shieldings (Table 2) were found to be highly correlated with the HF data with correlation coefficients R of 0.97, 0.99, and 0.98 for C^{γ}, C^{δ_2}, and C^{ϵ_1} shieldings, respectively. Such high correlations between B3LYP and HF data were also found in ^{13}C NMR shift predictions of Trp-containing peptides[19] as well as ^{31}P NMR shift predictions of phosphonates.[36] The theory-versus-experiment correlations for the B3LYP results are marginally improved over the HF predictions (an increase of 0.02−0.03 in R), with $R = 0.97$, 0.90, 0.85 for C^{γ}, C^{δ_2}, and C^{ϵ_1} shifts, respectively. However, the slopes (C^{γ}: −1.18; C^{δ_2}: −1.21; C^{ϵ_1}: −1.70) of the B3LYP theory-versus-experiment predictions are much worse than those of the HF predictions (C^{γ}: −1.07; C^{δ_2}: −1.09; C^{ϵ_1}: −1.54). Moreover, the B3LYP computed absolute shieldings of TMS (known experimentally to be 186 ppm)[37] in the B3LYP predictions (C^{γ}: 193.5; C^{δ_2}: 208.2; C^{ϵ_1}: 275.7) are again as expected worse than the results of the HF calculations (C^{γ}: 190.7; C^{δ_2}: 202.0; C^{ϵ_1}: 253.8). Clearly then, the B3LYP results have poorer slopes (and, hence, absolute shieldings), suggesting that electron correlation effects do not have any significant effects on shielding in these systems.

---

(36) Zhang, Y.; Oldfield, E. *J. Phys. Chem. B* **2004**, *108*, 19533.
(37) Jameson, A. K.; Jameson, C. J. *Chem. Phys. Lett.* **1987**, *134*, 461.

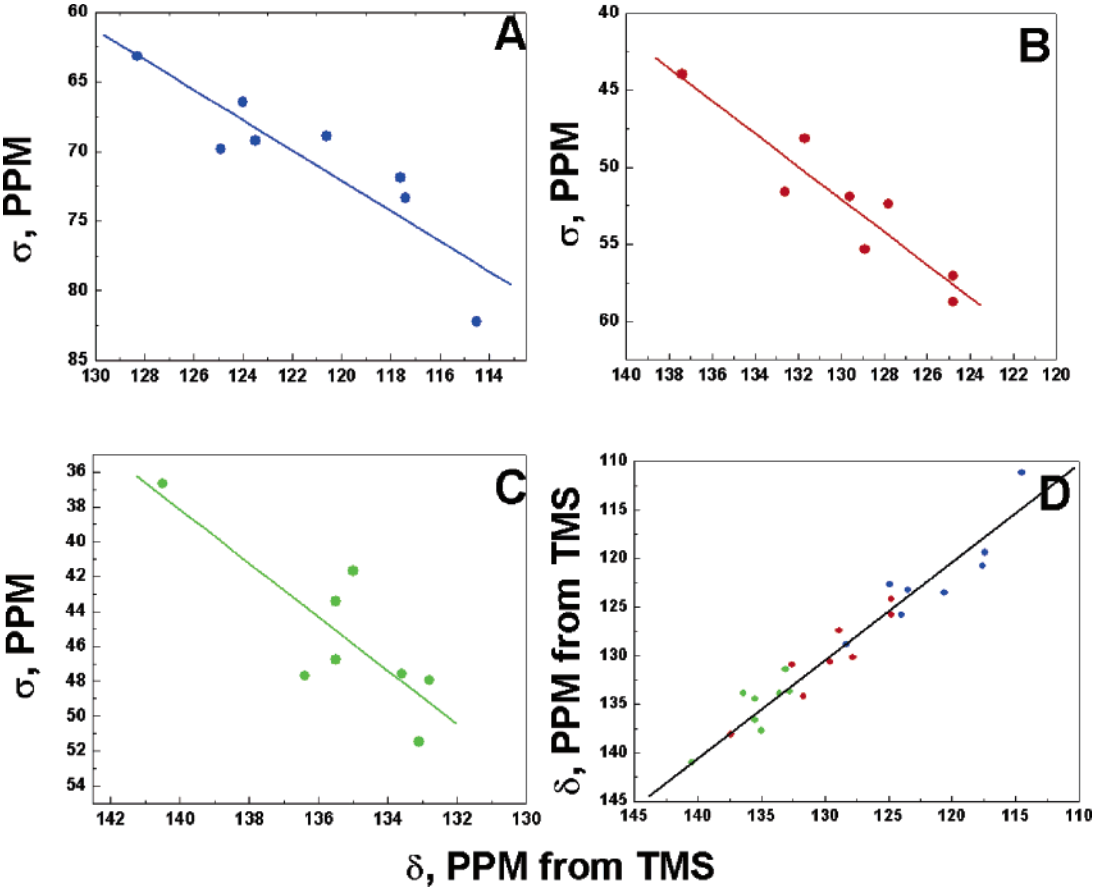


Figure 7. Correlations between experimental chemical shifts and those computed theoretically using the supermolecule approach and the HF method. (A) C^{γ}; (B) C^{δ_2}; (C) C^{ϵ_1}. (D) C^{γ}, C^{δ_2}, and C^{ϵ_1} combined chemical shift predictions (using eqs 1−3).

Using the correlation lines for the B3LYP results, the predicted ^{13}C shifts can be obtained from following equations:

$$\delta^{\text{pred}}\,(C^{\gamma}) = (193.5 - \sigma^{\text{calc}})/1.18 \tag{4}$$

$$\delta^{\text{pred}}\,(C^{\delta 2}) = (208.2 - \sigma^{\text{calc}})/1.21 \tag{5}$$

$$\delta^{\text{pred}}\,(C^{\epsilon 1}) = (275.7 - \sigma^{\text{calc}})/1.70 \tag{6}$$

which again enables a comparison of all 24 predicted shifts (eqs 4−6) with those determined experimentally. We find an overall R value of 0.98 and the rms error between theory and experiment of 1.6 ppm (about a 6% error). This represents a minor improvement over the HF predictions. It is clear then that the HF and DFT results have comparable R values (0.97, 0.98) and errors, but the HF results have better slopes and absolute shieldings than do the DFT predictions. Since such HF calculations were also used in previous investigations of ^{13}C NMR chemical shifts in proteins and the 20 common amino acids,[17−19] in addition to enabling the accurate predictions of torsion angles in peptides,[38] the HF results are used in the subsequent discussions of NMR properties, unless otherwise noted.

In these histidine dipeptides (and most likely in proteins), the major contributions to shielding are, therefore, the tautomeric state, together with intermolecular hydrogen bonding contacts in the "lattice", since the formyl-His-amide model gave only modest predictions while the monomer crystal structure results with hydrogen bond partners gave relatively good correlations with experiment (Figure 7D). The shift ranges seen in histidine are larger than those seen in tryptophan,[19] due to both tautomerism and hydrogen-bond interactions in histidine, the latter effects being attributable to the much more basic nature of imidazole versus indole (pK_a values of ∼7, 17, respectively). This then raises the question as to whether there are large shielding tensor magnitude and orientation changes associated with tautomerism and hydrogen-bond formation in these histidine dipeptides and, by inference, in proteins.

We show in Figure 8 (and Table S45) the computed shielding tensor element magnitudes (σ_{11}, σ_{22}, and σ_{33}) as a function of the isotropic shielding, for C^{γ} (Figure 8A), C^{δ_2} (Figure 8B), and C^{ϵ_1} (Figure 8C). Three representative tensor orientations (for **5**) are shown in Figure 8D−F, and full tensor orientation information is given in Tables S46 and S47. The most shielded element (σ_{33}) is perpendicular to the imidazole ring plane, as expected, with σ_{22} making the major contribution to σ_i (followed by σ_{33}) for each of the three atom types (C^{γ}, C^{δ_2}, C^{ϵ_1}), Table S45. As to the tensor orientations: we first investigated the orientation of σ_{11} relative to two molecule-fixed axes, the $C^{\gamma}-C^{\delta_2}$ bond vector and the $C^{\epsilon_1}-N^{\delta_1}$ bond vector, Table S46. These results showed that, for C^{γ} and C^{δ_2}, there are only small changes in tensor orientation between the two neutral forms (Table S46) but, for C^{ϵ_1}, which is attached to both nitrogen sites, there are major changes in the orientation of the in-plane tensor components in the two tautomers.

---

(38) Wi, S.; Sun, H.; Oldfield, E.; Hong, M. *J. Am. Chem. Soc.* **2005**, *127*, 6451.

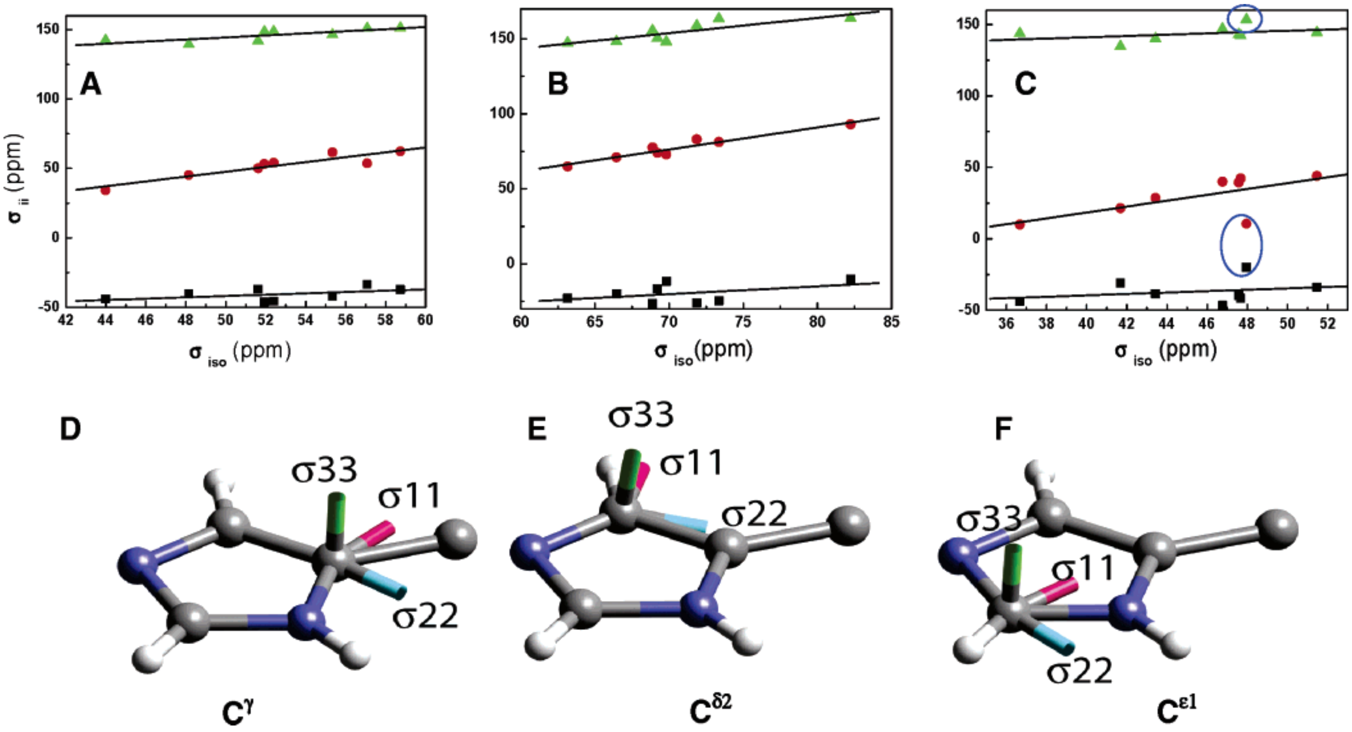


Figure 8. Computed shielding tensor magnitude and orientation results for C^{γ}, C^{δ_2}, and C^{ϵ_1} sites in histidine dipeptides (HF data). (A) Shielding tensor magnitudes as a function of the isotropic shielding for C^{γ} for compounds **5**−**12**. (B) Shielding tensor magnitudes for C^{δ_2} for compounds **5**−**12**. (C) Shielding tensor magnitudes for C^{ϵ_1} for compounds **5**−**12**. The circled blue points are for the imidazolium species, **9**. (D) Shielding tensor orientation of C^{γ} for compound **5**. (E) Shielding tensor orientation for C^{δ_2} for compound **5**. (F) Shielding tensor orientation for C^{ϵ_1} for compound **5**.

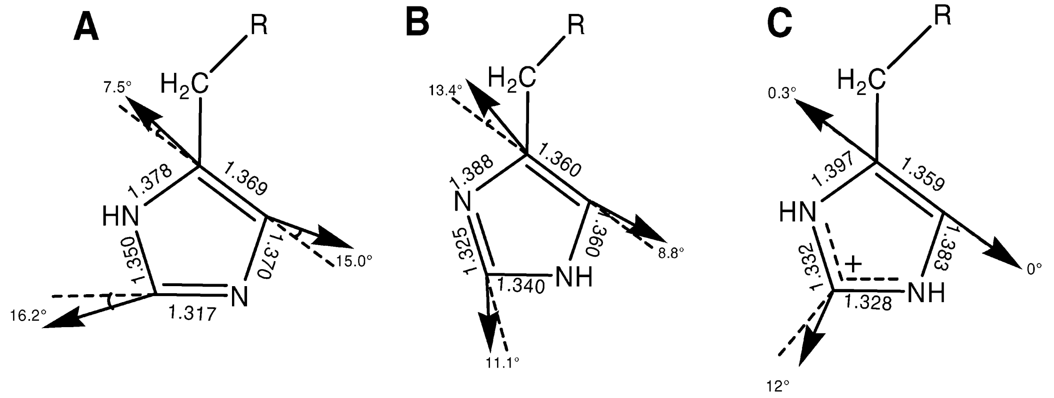


Figure 9. Schematic illustration of the shielding tensor element orientation for C^{γ}, C^{δ_2}, and C^{ϵ_1} in the two tautomeric forms of histidine and in the imidazolium form. (A) The N^{δ_1}−H tautomer **2** (average values shown for **5**−**8**). (B) The N^{ϵ_2}−H tautomer **3** (average values shown for **10**−**12**). (C) The imidazolium form **1** (**9**). The bond lengths shown are also average values found in this work. The angles shown are relative to the double bonds or in part C, the N^{δ_1}−C^{ϵ_1}−N^{ϵ_2} bisector (illustrated). The tensor element shown is σ_{22} in all cases except for C^{ϵ_1} in part C in which it is σ_{11}. See the text for details.

While on first inspection these tensor orientation results for C^{γ}, C^{δ_2}, and C^{ϵ_1} in the three different histidine forms may appear rather complex, they can in fact be greatly simplified and readily understood using structural and symmetry arguments. More specifically, in all three imidazole forms (**1**, **2**, **3**) the C^{γ}−C^{δ_2} bond is formally a double bond, as shown by the average bond lengths for **5**−**8**, **10**−**12**, and **9** in Figure 9. Based on symmetry, one component of the shielding tensor for C^{γ} and C^{δ_2} is expected to be oriented close to the C^{γ}−C^{δ_2} bond vector and as can be seen in Figures 8 and 9, this component is σ_{22}, as expected for a double bond. For C^{γ}, we find on average that the σ_{22}/C^{γ}−C^{δ_2} angle is ~9° while for C^{δ_2} we find that this angle is ~11°, Table S47, for all three types of histidine side chain. When the two tautomeric forms are compared, Figure 9A,B, it can be seen that there are even stronger similarities in tensor orientation in that the site adjacent the protonated nitrogen has on average an orientation of 7.5° (Figure 9A) or 8.8° (Figure 9B), while the site adjacent the deprotonated N has an orientation of 15° (Figure 9A) or 13.4° (Figure 9B). For the single imidazolium species **9**, the delocalized charge distribution among N^{δ_1}, C^{ϵ_1}, and N^{ϵ_2} leads to a close to C_{2v} symmetry and the tensor orientations of both C^{γ} and C^{δ_2} are essentially equivalent (0, 0.3°; Figure 9C) and are similar to the orientations found in His·HCl found previously.[39] For C^{ϵ_1}, the situation is slightly more complex, but all the same symmetry principles apply. For **5**−**8**, the relevant double bond is now between C^{ϵ_1} and N^{ϵ_2} (Figure 9A) and on average σ_{22} is oriented at ~16° from this bond vector. For **10**−**12**, the double bond is now between C^{ϵ_1} and N^{δ_1}, Figure

---

[39] Strohmeier, M.; Alderman, D. W.; Grant, D. M. *J. Magn. Res.* **2002**, *155*, 263.

Table 4. Hydrogen Bond BCP and Geometry Results for N^{δ_1}, N^{ϵ_2} Sites for Histidine Dipeptides and Some Other Molecules

			$\rho(\mathbf{r})$ (au)	$G(\mathbf{r})$ (au)	$V(\mathbf{r})$ (au)	$\nabla^2\rho(\mathbf{r})$ (au)	$G(\mathbf{r})/\rho(\mathbf{r})$ (au)	d_{NX} (Å)
Hydrogen Bond in Dipeptides[a]								
$\mathbf{N^{\delta_1}}$	**5**	$N^{\delta_1}-H\cdots O$	0.0298	0.0290	−0.0269	0.1248	0.9732	2.706
	6	$N^{\delta_1}-H\cdots O$	0.0351	0.0348	−0.0331	0.1464	0.9915	2.637
	7	$N^{\delta_1}-H\cdots O$	0.0262	0.0249	−0.0227	0.1080	0.9504	2.786
	8	$N^{\delta_1}-H\cdots O$	0.0289	0.0286	−0.0262	0.1240	0.9896	2.726
	9	$N^{\delta_1}-H\cdots O$	0.0320	0.0317	−0.0296	0.1356	0.9906	2.696
	10	$N^{\delta_1}\cdots H-O$	0.0278	0.0276	−0.0252	0.1196	0.9928	2.775
	11	$N^{\delta_1}\cdots H-N$	0.0304	0.0256	−0.0256	0.1024	0.8421	2.818
	12	$N^{\delta_1}\cdots H-O$	0.0296	0.0288	−0.0270	0.1228	0.9730	2.728
$\mathbf{N^{\epsilon_2}}$	**5**	$N^{\epsilon_2}\cdots H-N$	0.0315	0.0271	−0.0263	0.1116	0.8603	2.843
	6	$N^{\epsilon_2}\cdots H-N$	0.0323	0.0281	−0.0274	0.1148	0.8700	2.833
	7	$N^{\epsilon_2}\cdots H-N$	0.0347	0.0312	−0.0309	0.1256	0.8991	2.783
	8	$N^{\epsilon_2}\cdots H-N$	0.0357	0.0319	−0.0320	0.1268	0.8936	2.782
	9	$N^{\epsilon_2}-H\cdots O$	0.0388	0.0375	−0.0368	0.1532	0.9665	2.635
	10	$N^{\epsilon_2}-H\cdots O$	0.0258	0.0265	−0.0232	0.1188	1.0271	2.786
	11	$N^{\epsilon_2}-H\cdots O$	0.0221	0.0199	−0.0186	0.0848	0.9005	2.840
Closed-Shell Interactions[b]								
Hydrogen bond in $(H_2O)_2$			0.0198			0.0623	0.806	
Hydrogen bond in $(HF)_2$			0.0262			0.1198	1.027	
van der Waals bond in[b]								
Ne−HF			0.0099			0.0484	1.096	
Ar−HF			0.0077			0.0311	0.828	
Strong Hydrogen Bond in								
$[F\cdots H\cdots F]^{-}$[c]			0.1795	0.1113	−0.3370	−0.4572	0.6201	
$[NH_4]^{+}[F\cdots H\cdots F]^{-}$[d]			0.1912			−0.3735		
$[NH_4]^{+}[F\cdots H\cdots F]^{-}$[d]			0.1882			−0.7056		
$[H-O-H\cdots O-H]^{-}$[c]			0.1122	0.0695	−0.1291	−0.0397	0.6194	
$[H_3N\cdots H-NH_3]^{+}$[c]			0.0765	0.0401	−0.0667	−0.0542	0.5242	
cis-maleate monoanion[e]			0.1338	0.0851	−0.1754	−0.0210	0.6360	
Shared Interactions[b]								
CC bond in ethylene			0.3627			−1.1892	0.383	
CC bond in benzene			0.3268			−1.0134	0.293	

[a] Using the hydrogen-bonded supermolecule cluster model and HF wave functions. [b] Reference 21. [c] B3LYP/6-311++G(2d,2p) calculation on the geometry taken from ref 41 (optimized by MP2/aug-cc-pVTZ). [d] X-ray crystal data from ref 42. [e] B3LYP/6-311++G(2d,2p) calculation on the geometry taken from ref 41 (optimized by MP2/6-311++G(d,p)).

9B, and for these three species we find that σ_{22} is oriented at ~11° to the $C^{\epsilon_1}-N^{\delta_1}$ bond vector. The mean bond lengths for the two sets of tautomers (**5**−**8** and **10**−**12**) are shown in Figure 9 and strongly support this bonding picture in the dipeptides investigated here.

For the imidazolium species (**9**), the tensor orientation situation for C^{ϵ_1} is rather different, but we can still apply a symmetry principle. More specifically, **9** exists as two resonance forms with essentially identical (1.332, 1.328 Å) $C^{\epsilon_1}-N^{\delta_1}$ and $C^{\epsilon_1}-N^{\epsilon_2}$ bond lengths, as shown in Figure 9C. It therefore follows that one in-plane tensor element will lie along the $N^{\delta_1}-C^{\epsilon_1}-N^{\epsilon_2}$ bisector. As can be seen in Figure 9C and Table S47, we find that σ_{11} is oriented at ~12° from this bisector. In addition, the C^{ϵ_1} shielding tensor has close to axial symmetry, as illustrated by the circled points in Figure 8C. Overall then, the basic shielding tensor orientations for all three sites in all three basic structures can be readily understood based on simple symmetry arguments.

Of course, small deviations from these simple pictures are not unexpected since we have already demonstrated that intermolecular interactions contribute to shielding, but these effects appear to contribute more to the magnitudes of σ_{ii} (Figure 8) than to their orientation (Figure 9). To explore these intermolecular interactions in more detail, we next investigated their nature using Bader's atoms in molecules (AIM) theory,[21] an approach we used previously to investigate hydrogen bonding in asparagine·H_2O and in a series of other model compounds, as well as in the GB1 protein.[40]

We evaluated the charge density, $\rho(\mathbf{r})$; the Laplacian of the charge density, $\nabla^2\rho(\mathbf{r})$; the electronic kinetic energy density, $G(\mathbf{r})$; the electronic potential energy density, $V(\mathbf{r})$, and $G(\mathbf{r})/\rho(\mathbf{r})$, all at hydrogen bond critical points (BCPs) between $N^{\delta_1}(H^{\delta_1})$ or $N^{\epsilon_2}(H^{\epsilon_2})$ and a H-bond partner ligand (methylimidazole, $MeNH_3^+$, $CH_3CO_2^-$, H_2O, or EtOH). BCP results for the N^{δ_1} and N^{ϵ_2} sites from the HF calculations are shown in Table 4. The B3LYP results (Table S48) were found to be highly correlated with the HF results, having R values of 1.00, 0.98, 0.98, 0.92, 0.96 for $\rho(\mathbf{r})$, $G(\mathbf{r})$, $V(\mathbf{r})$, $\nabla^2\rho(\mathbf{r})$, $G(\mathbf{r})/\rho(\mathbf{r})$ and slopes all very close to 1.

In all cases, the Laplacian $\nabla^2\rho(\mathbf{r})$ is positive, as found in all closed-shell interactions, including hydrogen bonds and van der Waals bonds.[21,40] The ratio $G(\mathbf{r})/\rho(\mathbf{r})$[21] was also used to investigate the nature of these interactions. As shown in Table 4, the range of $G(\mathbf{r})/\rho(\mathbf{r})$ in hydrogen bonds in these His-containing dipeptides is 0.84−1.03, very similar to that found (0.81−1.03) with the hydrogen bonded dimer complexes investigated by Bader,[21] as well as the values found for protein backbone hydrogen bonds (0.77−0.93) we reported previously.[40] Clearly, these results indicate strong similarities in the nature of the hydrogen bonds in the histidine dipeptides, peptide backbones, and simple dimers and are quite distinct to the values found in stronger hydrogen bonds found in species such as $[FHF]^-$ or the *cis*-maleate monoanion,[41−42] where $\nabla^2\rho(\mathbf{r})$ is negative and $G(\mathbf{r})/\rho(\mathbf{r})$ is ca. 0.52−0.64, close to the values found in typical covalent bonds, Table 4.

---

(40) (a) Arnold, W. D.; Oldfield, E. *J. Am. Chem. Soc.* **2000**, *122*, 12835. (b) Arnold, W. D.; Sanders, L. K.; McMahon, M. T.; Volkov, A. V.; Wu, G.; Coppens, P.; Wilson, S. R.; Godbout, N.; Oldfield, E. *J. Am. Chem. Soc.* **2000**, *122*, 4708.

(41) Barich, D. H.; Nicholas, J. B.; Haw, J. F. *J. Phys. Chem. A* **2001**, *105*, 4708.

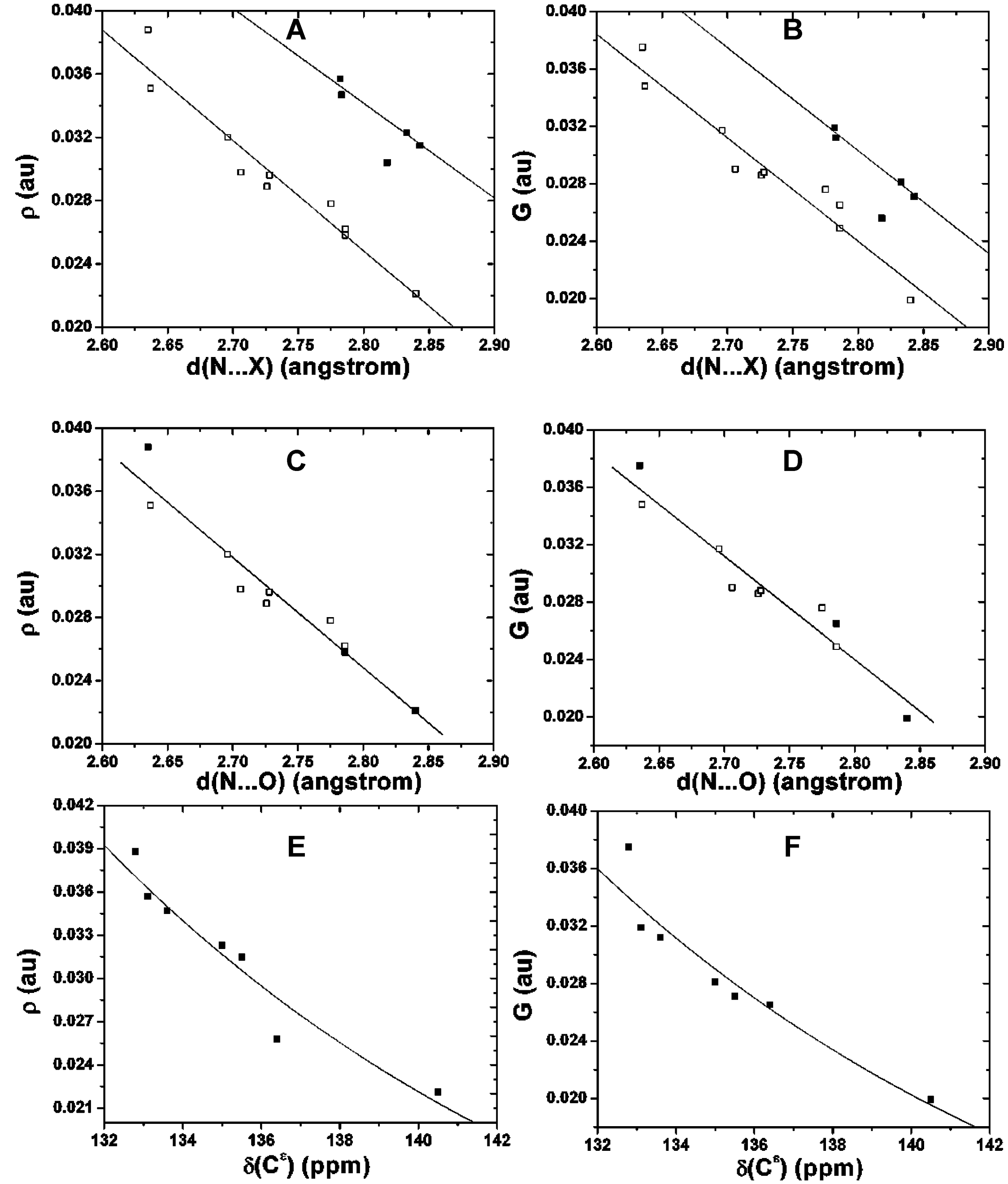


Figure 10. Correlations between BCP properties and geometric, spectroscopic observables. (A) Plot of $\rho(\mathbf{r})$ versus d_{NX} (□: X = O; ■: X = N). (B) Plot of $G(\mathbf{r})$ versus d_{NX} (□: X = O; ■: X = N). (C) Plot of $\rho(\mathbf{r})$ versus d_{NO} (□: $\mathrm{N}^{\delta 1}$; ■: $\mathrm{N}^{\epsilon 2}$). (D) Plot of $G(\mathbf{r})$ versus d_{NO} (□: N^{δ_1}; ■: N^{ϵ_2}). (E) Plot of N^{ϵ_2} $\rho(\mathbf{r})$ versus $\delta(\mathrm{C}^{\epsilon_1})$. (F) Plot of N^{ϵ_2} $G(\mathbf{r})$ versus $\delta(\mathrm{C}^{\epsilon_1})$. The lines in plots A−D are straight line fits to the data. In plots E and F, the fitting equations were $\rho(\mathbf{r}) = 501.4 \exp(-0.07163\ \delta)$ (E) and $G(\mathbf{r}) = 471.2 \exp(-0.07182\ \delta)$ (F). Data are from HF calculations.

We next investigated whether there might be some more quantitative correlations between each of the BCP properties and various structural and spectroscopic features, perhaps the most likely candidate being the proximity of the N-atoms in the histidine rings to their hydrogen bond partner atoms. We show these nearest neighbor distances, d_{NO} and d_{NN}, in Table 4 and, on inspection, there are clearly correlations between each of the BCP properties and the d_{NO}, d_{NN} heavy atom hydrogen bond distances. Graphs of $\rho(\mathbf{r})$ and $G(\mathbf{r})$ plotted versus d_{NX} are shown in Figure 10A and 10B, and other plots of $\mathrm{V}(\mathbf{r})$ and $\nabla^2\rho(\mathbf{r})$ versus d_{NX} are given in Figure S2A and S2B. Even better correlations are obtained when the H-bond distances are sorted according to whether the bond is N···H···O or N···H···N, an effect which actually combines results for the N^{δ_1} and N^{ϵ_2} sites. Results for $\rho(\mathbf{r})$ and $G(\mathbf{r})$ plotted versus d_{NO} are shown in Figure 10C and D, with similar results for $\mathrm{V}(\mathbf{r})$ and $\nabla^2\rho(\mathbf{r})$ given in Figure S2C and D. The N···H···O site results (a combination of $\mathrm{N}^{\delta_1}-\mathrm{H}^{\delta_1}\cdots\mathrm{O}$, $\mathrm{N}^{\delta_1}\cdots\mathrm{H}-\mathrm{O}$, and $\mathrm{N}^{\epsilon_2}-\mathrm{H}^{\epsilon_2}\cdots\mathrm{O}$ sites) have an average R^2 value of 0.93 for the d_{NO} correlations with $\rho(\mathbf{r})$, $G(\mathbf{r})$, $V(\mathbf{r})$, and $\nabla^2\rho(\mathbf{r})$. For the N···H···N data, we find good correlations with the N^{ϵ_2} site BCP results (average $R^2 = 0.98$) but not for the N^{δ_1} site, as shown in Figure S2E−H, an effect which may simply be attributable to the small range in d_{NN} and the small number ($N = 4$) of data points for this site. Clearly

---

(42) van Reeuwijk, S. J.; van Beek, K. G.; Feil, D. *J. Phys. Chem. A* **2000**, *104*, 10901.

then, all of the BCP results are correlated with hydrogen bond length. The charge density, as expected, is largest for the shortest hydrogen bond distance and, over this relatively small range of distances, appears to vary in a linear manner with hydrogen bond length. An exponential relationship is expected[40,43] but is not observed over the small range in d_{NX} in these dipeptides.

But are there correlations between the BCP properties and the ^{13}C NMR chemical shifts? To make these comparisons, we plotted the BCP properties $\rho(\mathbf{r})$, $G(\mathbf{r})$, $V(\mathbf{r})$, and $\nabla^2 \rho(\mathbf{r})$ versus the C^{γ}, C^{δ_2}, and C^{ϵ_1} chemical shifts using both the N^{δ_1} and N^{ϵ_2} BCP results. For C^{γ} and C^{δ_2}, there were no correlations, but for C^{ϵ_1}, there were clear correlations, but only with the N^{ϵ_2} BCP data set. These results are shown in Figure 10E and F and S2I and J for each of the four BCP properties and, on average, the correlation coefficient $R^2 = 0.92$. These results strongly suggest that C^{ϵ_1} is highly sensitive to the effects of hydrogen bonding, and indeed in our initial calculations (Figure 7C), shielding at C^{ϵ_1} was clearly the most sensitive to the nature of the fragments employed, with reasonable results only being obtained with the supermolecule calculations. On the other hand, the C^{γ}/C^{δ_2} shieldings were highly (anti)correlated, even in the single N-formyl-histidine amide model calculations. The fact that the shifts of C^{ϵ_1} are particularly sensitive to intermolecular interactions is perhaps not unexpected given its low pK_a value and the observation that H^{ϵ_1} readily exchanges under basic conditions;[44] that is, the $C^{\epsilon_1}-H^{\epsilon_1}$ bond is quite polar. The good shift correlation with only the N^{ϵ_2} BCP results is more surprising but appears to be related to the relatively larger covalence and the larger ranges of the BCP properties at this site, as described above.

Conclusions

The results we have described above are of interest for a number of reasons. First, we have obtained the X-ray crystallographic structures of six histidine-containing dipeptides containing $N^{\delta_1}-H$ and $N^{\epsilon_2}-H$ tautomers or an imidazolium side chain. Second, we have obtained the ^{13}C MAS NMR spectra of these plus two other His dipeptides. The chemical shift ranges are very large for C^{γ} (12.7 ppm) and for C^{δ_2} (13.8 ppm); plus, the C^{γ} and C^{δ_2} shifts are highly (anti)correlated ($R^2 = 0.90$; slope $= -1.03$). A similar chemical shift range and correlation is also seen in proteins, but only for solvent inaccessible histidine residues, with the solvent accessible (>50Å^2 exposed surface area) residues having essentially the same shifts seen with His or a histidine-containing peptide, in aqueous solution. Third, we have used quantum chemical methods to investigate the histidine dipeptide MAS NMR chemical shifts. The use of N-formyl-histidine amide or crystal monomer structures did not give good predictions of the experimental shifts; however, incorporation of near-neighbor residues in a fully quantum mechanical "supermolecule" calculation provided much improved predictions, with an overall rms error of 1.6−1.9 ppm over a 26 ppm chemical shift range; $R = 0.97-0.98$. The C^{ϵ_1} shifts appear to be dominated by intermolecular interactions with hydrogen-bond partner molecules. Changes in isotropic shielding were dominated by σ_{22}, and σ_{11}/σ_{22} tensor orientations were correlated with imidazole structure types (π, τ, imidazolium) via simple symmetry-based models. Fourth, we used AIM theory to probe the details of the interactions between imidazole and its hydrogen bond partners. In all cases, the interactions could be classified as closed shell interactions. The BCP properties $\rho(\mathbf{r})$, $V(\mathbf{r})$, $G(\mathbf{r})$, and $\nabla^2\rho(\mathbf{r})$ were all highly correlated with hydrogen bond lengths (d_{NO} and d_{NN}) and for C^{ϵ_1}, BCP properties were correlated with the C^{ϵ_1} chemical shift, supporting the dominance of intermolecular interaction effects on shielding at this site. Taken together, the results of these experimental and theoretical investigations indicate the need to incorporate lattice partners (or in a protein, neighboring residues) in order to reproduce imidazole C^{γ}, C^{δ_2}, and C^{ϵ_1} NMR chemical shifts. With the availability of high quality wave functions from such supermolecule clusters, the way is then open to evaluate the details of these hydrogen bond interactions. In addition, in the "reverse direction", when structures are not known, knowledge of C^{γ} and C^{δ_2} shifts enables in most cases good predictions of the tautomeric state, while knowledge of C^{ϵ_1} shifts enable predictions of BCP properties and from these more detailed insights into the nature of these hydrogen bonds. These results are of general interest in the context of NMR and quantum chemical studies of protein structure and function. In addition, it is to be expected that many imidazole-containing drugs will exhibit extremely large chemical shift ranges when bound to proteins and that ^{13}C shifts of these species can be expected to be useful probes of local structure, providing information not directly obtainable from protein crystallographic investigations, such as protonation, tautomer, and conformer states.

Acknowledgment. We thank Scott Wilson for his help with the crystallography. This work was supported by the United States Public Health Service (NIH Grant GM50694). The X-ray crystallographic facilities were supported by the National Science Foundation (Grant CHE 95-03145).

Supporting Information Available: Crystal structure files and more results (Tables S1−S48 and Figures S1 and S2). This material is available free of charge via the Internet at http://pubs.acs.org.

JA051528C

---

(43) Espinosa, E.; Molins, E.; Lecomte, C. *Chem. Phys. Lett.* **1998**, *285*, 170.

(44) Schramm, S.; Oldfield, E. *Biochemistry* **1983**, *22*, 2908.