J|A|C|S
ARTICLES
Published on Web 03/22/2002

Structural Dependencies of Interresidue Scalar Coupling $^{h3}J_{NC'}$ and Donor ^{1}H Chemical Shifts in the Hydrogen Bonding Regions of Proteins

Michael Barfield*

Department of Chemistry, University of Arizona, Tucson, Arizona 85721

Received December 7, 2001. Revised Manuscript Received January 29, 2002

Abstract: A study is presented of the structural dependencies for scalar *J*-coupling and the amide donor ^{1}H chemical shifts in the hydrogen bonding regions of proteins. An analysis of the interactions between the donor hydrogen and acceptor oxygen orbitals in an N−H···O═C moiety suggests that there are three major structural factors for $^{15}N-^{13}C$ coupling across hydrogen bonds: (1) the H···O′ internuclear separation $r_{HO'}$, (2) the H···O′═C′ angle θ_2, and (3) indirect contributions involving the oxygen loan pair electrons should lead to a dependence on the H···O′═C′−N′ dihedral angle ρ. Density functional theory (DFT) and finite perturbation theory (FPT) were used to obtain the Fermi contact (FC) contributions to interresidue coupling in formamide dimers with systematic variation of these structural parameters. The computed $^{h3}J_{NC'}$ exhibit good correlations with $\cos^2\theta_2$ combined with an exponential dependence on $r_{HO'}$. The correlation is further improved by including a dependence on the dihedral angle ρ. For each of the 34 H-bonds having observable interresidue coupling in the immunoglobulin binding domain of streptococcal protein G, a formamide dimer was generated from the crystallographic structure with energy-optimized donor H-atom positions. The computed coupling constants are in reasonable agreement with the experimental, and there are excellent correlations with the simple equations involving θ_2 and r_{HO} if α-helix and β-sheet regions are treated separately. This dichotomy is removed by introducing the dependence on the dihedral angle ρ. Justification for the use of formamide dimers is provided by almost identical interresidue coupling contants for larger sequences extracted from the X-ray structure. The amide donor ^{1}H chemical shifts, which were based on DFT and GIAO (gauge including atomic orbital) methods, are in poorer agreement with the experimental data but exhibit excellent correlation with $r_{HO'}$, θ_2, and ρ.

I. Introduction

Recent experimental and computational studies of nuclear spin−spin coupling constants across hydrogen bonds[1−8] in nucleic acids have provided new structural constraints for conformational studies of nucleic acids. Subsequently, two groups independently observed $^{15}N-^{13}C$ coupling across H-bonds in uniformly enriched human ubiquitin.[9,10] For as many as 31 backbone trans H-bonds, the observed scalar couplings ranged in magnitude from 0.25 to 0.92 Hz. Because the low (1.8 Å)-resolution crystal structure for this protein gave poor correlations with the structural data, Cornilescu et al.[11] investigated the immunoglobulin binding domain of protein G, for which a higher-resolution crystal structure (1.1 Å) was available. They reported interresidue coupling constants for 34 bond pairs, determined that these have a negative sign, and on empirical grounds proposed a simple exponential dependence on the N···O′ distance r_{NO}. Subsequently, it was demonstrated that trans H-bond coupling[12] $^{h3}J_{NC'}$ could be observed in larger proteins,[13] as well as in guanosine quartets.[14] Trans H-bond $^{15}N-^{13}C$ coupling constants have been used to examine pressure effects,[15] to investigate ligand-induced changes in the geometry around

* Address correspondence to: Michael Barfield, Department of Chemistry, University of Arizona, Tucson, AZ 85721, U.S.A. Telephone: (520) 621-6348. Fax: (520) 621-8407. E-mail: Barfield@u.arizona.edu.

(1) For a review of scalar coupling across hydrogen bonds, see: Grzesiek, S.; Cordier, F.; Dingley A. J. *Methods Enzymol*. **2001**, *338*, 111−133.
(2) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc*. **1998**, *120*, 8293−8297.
(3) Pervushin, K.; Ono, A.; Fernández, C.; Szyperski, T.; Kainosho, M.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A*. **1998**, *95*, 14147−14151.
(4) Dingley, A. J.; Masse, J. E.; Peterson, R. D., Barfield, M.; Feigon, J.; Grzesiek, S. *J. Am. Chem. Soc*. **1999**, *121*, 6019−6027.
(5) Benedict, H.; Shenderovich, I. G.; Malkina, O. L.; Malkin, V. G.; Denisov, G. S.; Golubev, N. S.; Limbach, H.-H. *J. Am. Chem. Soc*. **2000**, *122*, 1979−1988.
(6) Del Bene, J. E., Perera, S. A.; Bartlett, R. J. *J. Am. Chem. Soc*. **2000**, *122*, 3560−3561.
(7) Pecul, M.; Leszczynski, J.; Sadlej, J. *J. Chem. Phys*. **2000**, *112*, 7930−7938; Pecul, M.; Leszczynski, J.; Sadlej, J. *J. Phys. Chem*. **2000**, *104*, 8105−8113.
(8) Barfield, M.; Dingley, A. J.; Feigon, J.; Grzesiek, S. *J. Am. Chem. Soc*. **2001**, *123*, 4014−4022.
(9) Cordier, F.; Grzesiek, S. *J. Am. Chem. Soc*. **1999**, *121*, 1601−1602.
(10) Cornilescu, G.; Hu, J.-S.; Bax, A. *J. Am. Chem. Soc*. **1999**, *121*, 2949−2950.
(11) Cornilescu, G.; Ramirez, B. E.; Frank, M. K.; Clore G. M.; Gronenborn, A. M.; Bax, *A. J. Am. Chem. Soc*. **1999**, *121*, 6275−6279.
(12) Conventions for describing scalar coupling between nuclei A and B over H-bonds include $^{hx}J_{AB}$ and $^{xh}J_{AB}$ where *x* designates the sum of bonds and H-bonds separating the coupled nuclei. Here we follow the Wüthrich convention and use the former. See Pervushin, K.; Fernández, C.; Riek, R.; Ono, A.; Kainosho, M.; Wüthrich, K. *J. Biomol. NMR* **2000**, *16*, 39−46.
(13) Wang, Y.-X.; Jacob, J.; Cordier, F.; Wingfield, P.; Stahl, S. J.; Lee Huang, S.; Torchia, D.; Grzesiek, S.; Bax, A. *J. Biomolecular NMR* **1999**, *14*, 181−184. Lohr, F.; Mayhew, S. G.; Ruterjans, H. *J. Am. Chem. Soc.* **2000**, *122*, 9289−9295.
(14) Dingley, A. J.; Masse, J. E.; Feigon, J.; Grzesiek, S. *J. Biomol. NMR* **2000**, *16*, 279−289.
(15) Li, H.: Yamada, H.; Akasaka, K.; Gronenborn, A. M. *J. Biomol. NMR*, **2000**, *18*, 207−216.

10.1021/ja012674v CCC: $22.00

the H-bonds of proteins,[16] as a measure of the populations of closed H-bonds,[17] and as a method for analyzing the influence of crystallographic resolution in H-bond length comparisons.[18] The very good correlations between trans H-bond coupling constants and donor ^{1}H chemical shifts observed in DNA triplexes[4] were also noted experimentally between $^{h3}J_{NC'}$ and the amide ^{1}H chemical shifts in ubiquitin,[9,10] and computationally in model *N*-methylacetamide dimers.[19]

Of particular interest is an understanding of the factors controlling this type of coupling so that these quantities could be used for structural studies in proteins. Scheurer and Brüschweiler[19] used DFT methods for an *N*-methylacetamide dimer model to compute and plot $^{h3}J_{NC'}$ as a function of H···O′ separation and the ∠H···O′═C′ angle. Subsequently, this work was extended by Bagno[20] using formamide dimers as model compounds to investigate the dependence of $^{h3}J_{NC'}$ on additional structural factors including the effects of nonplanarity. In comparison with the experimental ubiquitin results,[9,10] the systematic over-estimation of $^{h3}J_{NC'}$ was attributed to motional processes and possible inadequacies of the use of model compounds. No simple functional dependence relating the interresidue coupling to bond lengths and angles was proposed.

The present work extends previous studies of interresidue coupling in proteins to include the following: (1) an analysis of the angular dependence of the interactions between the donor hydrogen and orbitals of the acceptor oxygen. (2) The relevant angles and distances are varied systematically and used to develop explicit expressions for the dependence of $^{h3}J_{NC'}$ on the important structural factors (θ_2, r_{OH}, and ρ). (3) All 34 values of $^{h3}J_{NC'}$ in protein G and all 31 values for ubiquitin are computed using formamide dimers extracted from the crystallographic data. The explicit expressions for the $^{h3}J_{NC'}$ on three structural parameters provide excellent correlations of all DFT/FPT results as well as the computed DFT/GIAO results for the amide ^{1}H chemical shifts in protein G.

II. Computational Methods

The formamide molecule was fully optimized using the Gaussian 98 codes[21,22] at the B3PW91/6-31G** level of density functional theory.[23,24] The B3PW91 method[25−27] makes use of Becke's three-parameter hybrid exchange functional[25] and the gradient correlated Perdew−Wang 1991 correlation functional.[26] No further discussion of the monomer is presented here since there exist extensive theoretical studies of NMR parameters at a variety of computational levels.[7,28] The optimized formamide structure was used to generate dimers in which scalar coupling could be investigated systematically as a function of intermolecular distances and angles.

A. Scalar *J*-Coupling Based on DFT/FPT Methods. The Fermi contact term is usually the dominant contributor to scalar coupling. This term is exceedingly sensitive to the inclusion of electron correlation effects.[29−31] These effects can very effectively be introduced into density functional methods, making them well suited for such calculations, especially in larger molecules which are not easily accommodated by many-body techniques.[29−34] Recent studies[4,8,35−37] of nuclear spin−spin coupling, which combined DFT and finite perturbation theory (FPT) methods,[38] gave excellent conformity with the experimental data. Fermi contact (FC) contributions to the scalar coupling constants for the optimized structures were obtained at the unrestricted UB3PW91/6-311G** triple-split level with polarization functions on hydrogen and heavier atoms. Calculated DFT/FPT results are based on the FC output of the FIELD option of Gaussian98.[21,39] Four mechanisms are generally considered to be important for nuclear spin−spin coupling.[29−31] On the basis of computations including noncontact terms, Scheurer and Brüschweiler[19] concluded that the Fermi contact term contributed 96% of the total $^{h3}J_{NC'}$ in acetamide dimers. Only the Fermi contact (FC) contributions are reported here.

B. Amide ^{1}H Chemical Shifts Using GIAO Methods. Magnetic shielding results were based on the GIAO (gauge including atomic orbitals) formulation[40,41] using DFT at the B3PW91/6-311G** level of Gaussian98.[21] The DFT functionals as implemented in these codes do not include a specific magnetic field dependence.[42] All ^{1}H chemical shifts reported here are *isotropic* values, which are indirectly referenced to tetramethylsilane (TMS). The CH_4 molecular structure (optimized at the B3PW91/6-31G** level) led to a 31.65 ppm ^{1}H isotropic magnetic shielding (B3PW91/6-311G**). The magnetic shielding of TMS was inferred from the experimental shift of gas-phase methane (0.13 ppm).[43,44] Computations were performed using Silicon Graphics IRIS Origin 2000 and RISC 6000 IBM590 workstations.

(16) Cordier, F.; Wang, C.; Grzesiek, S.; Nicholson, L. K. *J. Mol. Biol*. **2000**, *304*, 497−505.
(17) Jaravine, V. A.; Alexandrescu, A. T.; Grzesiek, S. *Protein Sci.* **2001**, *10*, 943−950.
(18) Alexandrescu, A. T.; Snyder, D. R.; Abildgaard, F. *Protein Sci*. **2001**, *10*, 1856−1868.
(19) Scheurer, C.; Brüschweiler, R. *J. Am. Chem. Soc*. **1999**, *121*, 8661−8662. Czernek, J.; Brüschweiler, R. *J. Am. Chem. Soc*. **2001**, *123*, 11079−11080.
(20) Bagno, A. *Chem. Eur. J*. **2000**, *6*, 2925−2930.
(21) Frisch, M. J., Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G., Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G,; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W. M.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
(22) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley-Interscience: New York, 1986.
(23) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, 864−871. Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133−1138.
(24) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford, New York, 1989. March, N. H. *Electron Density Theory of Atoms and Molecules*; Academic Press: San Diego, 1992.
(25) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648−5652.
(26) Perdew, J. P.; Wang, Y. *Phys. Rev. B.* **1992**, *45*, 13244−13249.
(27) For a comparison of a molecular structure computed via B3LYP and B3PW91, see Ashvar, C. S.; Devlin, F. J.; Stephens, P. J. *J. Am. Chem. Soc.* **1999**, *121*, 2836−2849.
(28) Vaara, J.; Kaski, J.; Jokisaari, J.; Diehl, P. *J. Phys. Chem*. **1997**, *101*, 5069−5081.
(29) Helgaker, T.; Jaszuński, M.; Ruud, K. *Chem. Rev*. **1999**, *99*, 293−352.
(30) Contreras, R. H.; Peralta, J. E.; Giribet, C. G.; Ruiz de Azúa, M. C.; Facelli, J. C. A*nn. Rep. NMR Spectrosc.* **2000**, *41*, 57−166.
(31) Theoretical aspects of nuclear spin−spin coupling have been reviewed annually starting in 1972: Fukui H.; Baba, T. In *Nuclear Magnetic Resonance; Specialist Periodical Reports* 30; The Chemical Society London: London, 2001.
(32) Malkin, V. G.; Malkina, O. L.; Salahub, D. R. *Chem. Phys. Lett.* **1994**, *221*, 91−99. Malkina, O. L.; Salahub, D. R.; Malkin, V. G. *J. Chem. Phys*. **1996**, *105*, 8793−8800.
(33) Dickson, R. M.; Ziegler, T. *J. Phys. Chem*. **1996**, *100*, 5286−5290.
(34) Bacskay, G. B.; Bytheway, I.; Hush, N. S. *J. Am. Chem. Soc*. **1996**, *118*, 3753−3756; Hush, N. S. *J. Am. Chem. Soc*. **1997**, *119*, 1717−1719.
(35) Onak, T.; Jaballas, J.; Barfield, M. *J. Am. Chem. Soc*. **1999**, *121*, 2850−2856.
(36) Fierman, M.; Nelson, A.; Khan S. I.; Barfield, M.; O'Leary, D. J. *Org. Lett.* **2000**, *2*, 2077−2080.
(37) Barfield, M.; Bergset, J. M.; O'Leary, D. J. *Magn. Reson. Chem*. **2001**, *39*, S115−S125.
(38) Pople, J. A.; McIver, J. W., Jr.; Ostlund, N. S. *J. Chem. Phys.* **1968**, *49*, 2960−2964, 2965−2970.
(39) All scalar-coupling computations were performed with $\lambda = 0.0025$ in eq 3 of ref 35, and the very-tight SCF convergence option of G98.
(40) For general reviews of theory of magnetic shielding see, for example: Jameson, C. J.; De Dios, A. C. Nuclear Magnetic Resonance. In *Specialist Periodical Reports*; The Chemical Society: London, 2001; No. 30 and previous chapters in this series.
(41) Ditchfield, R. *Mol. Phys*. **1974**, *27*, 789−807.
(42) Becke, A. D.*Can. J. Chem.* **1996**, *74*, 995−997.

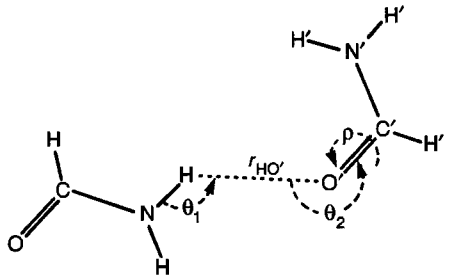

Figure 1. Diagram of formamide dimer depicting the designations of the internuclear distances and angles used in this study. The H-bond distance $r_{HO'}$, the internal angles $\theta_1 = \angle N-H\cdots O'$, and $\theta_2 = \angle H\cdots O'=C'$. The dihedral angle $\rho = \angle H\cdots O'=C'-N'$ is measured about the O′═C′ bond.

III. Structural Dependence of $^{15}NH\cdots O^{13}C$ Scalar Coupling in Formamide Dimers

Calculated results for scalar coupling based on ab initio or DFT methods are often in good agreement with the experimental data. However, these methods (as typically performed) give no information on the structural/electronic interactions leading to the computed results. The following section presents analyses of the structural dependencies of the interactions between the donor hydrogen and all valence orbitals on the acceptor oxygen in an N−H···O═C moiety. These equations provide a basis for analyzing the DFT/FPT data for formamide dimers where the associated structural parameters are varied systematically.

A. Structural Dependencies Of H···O′ Interactions In An H···O′═C′−N′ Moiety. Early studies of scalar coupling made use of semiempirical valence-bond[45−48] (VB) and molecular orbital (MO) methods.[49] Perturbation theory was used to obtain expressions relating coupling constants to VB exchange integrals[46−48] or MO resonance integrals.[49−52] In semiempirical MO theory the latter are assumed to be proportional to overlap integrals between orbitals on different centers. The exchange integrals of simple VB theory contain other terms, but the overlap integrals also play a major role.[53]

A variety of angles and distances have been used to characterize the geometry of the N−H···O′═C′ H-bond.[11,19,20,54,55] Consider the situation presented by trans H-bonding $^{15}N-^{13}C$ coupling in the formamide dimer depicted in Figure 1. This figure defines the structural parameters used here, for example, the H···O′ interatomic distance $r_{H-O'}$, the two internal angles, $\theta_1 \equiv \angle N-H\cdots O'$ and $\theta_2 \equiv \angle H\cdots O'=C'$, and the dihedral angle, $\rho \equiv \angle H\cdots O'=C'-N'$, measured about the O′═C′ bond.

Depicted in Figure 2 for an N−H···O′═C′−N′ moiety are those orbitals which are relevant to this discussion. For simplicity, the classical H-bonding geometry[54,55] is assumed with trigonal hybridization at the N, O, and C atoms. One of the hybrid-type orbitals (HTOs) n on N is directed to the 1s atomic orbital (AO) h of the donor hydrogen atom. In addition to the three trigonal sp^2 HTOs on the acceptor oxygen atom O′, there is a $2p_\pi$ AO (not depicted) perpendicular to the plane. The HTO o_1' forms the O′═C′ σ-bond with c_1' on C′ while o_2' and o_3' designate the oxygen lone pair orbitals. The most important Fermi contact contributions to scalar coupling are expected to arise from the interactions between bonds having nonvanishing density at the coupled nuclei,[45−52] for example, the donor n−h bond and acceptor $o_1'-c_1'$ σ-bond in Figure 2. It seems likely that $^{h3}J_{NC'}$ would be dominated by the interaction between the donor hydrogen AO h and the acceptor oxygen o_1' orbitals since the integrals defining the electronic interactions typically decrease exponentially with the distance between atoms. It will also be of interest to examine the possible role of interactions between the donor hydrogen and the acceptor oxygen lone pairs and $2p_\pi$ orbitals.

To examine the structural dependence of interactions involving the HTOs in Figure 2, it is convenient to introduce spherical polar coordinates θ and φ. A general hybrid orbital t can then be constructed[56] from the 2s (s) and the 2p atomic orbitals (p_x, p_y, and p_σ)

$$t = as + (1 - a^2)^{1/2}[p_x \cos\varphi \sin\theta + p_y \sin\varphi \sin\theta + p_\sigma \cos\theta], \quad (1)$$

where a^2 is the s-character, and θ and φ can be associated with internal and dihedral angles, respectively. With the assumption of trigonal HTOs on O′ each HTO in eq 1 has $\theta = 120°$ and $a^2 = {}^1/_3$. In an axis system with the *z*-axis lying along the C′═O′ bond, the orbitals o_1', o_2', o_3' and p_π' on O′ can be related to the dihedral angle ρ defined in Figure 1,

$$o_1' = \{s' + 2^{1/2}p_\sigma'\}/3^{1/2}, \quad (2a)$$

$$o_2' = \{s' + 2^{1/2}[3^{1/2}(p_x' \cos\rho + p_y' \sin\rho) - p_\sigma']/2\}/3^{1/2}, \quad (2b)$$

$$o_3' = \{s' - 2^{1/2}[3^{1/2}(p_x' \cos\rho + p_y' \sin\rho) + p_\sigma']/2\}/3^{1/2}, \quad (2c)$$

$$p_\pi' = (p_x' \sin\rho - p_y' \cos\rho) \quad (2d)$$

where s′ denotes the 2s atomic orbital on O′, and p_x', p_y', p_σ' are the three 2p orbitals in this axis system.

Because of their proximity, the most important overlap integrals should involve the donor hydrogen atom and the acceptor oxygen orbitals. These are most conveniently evaluated in a coordinate system having the *z*-axis along the H···O′ line depicted in Figure 1. On rotation by an angle θ_2 (∠H···O′═C′), overlap integrals between the donor hydrogen h and the acceptor oxygen orbitals in eqs 2a−d assume simple trigono-

(43) Emsley, J. W.; Feeney, J.; Sutcliffe, L. H. *High-Resolution Nuclear Magnetic Resonance*; Pergamon: Elmsford, NY, 1966; Vol. 2. The experimental gas phase data were converted as described by Kutzelnigg et al.[44]
(44) Kutzelnigg, W.; Fleischer, U.; Schindler, M. In *NMR Basic Principles and Progress*; Diehl, P., Fluck, E., Kosfeld, R., Eds.; Springer: Berlin, 1990; Vol. 23, pp 165−262.
(45) Karplus, M.; Anderson, D. H. *J. Chem. Phys.* **1959**, *30*, 6−10; Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11−15.
(46) Barfield, M.; Grant, D. M. *Adv. Magn. Reson.* **1965**, *1*, 149−193.
(47) Barfield, M. *J. Chem. Phys.* **1967**, *46*, 811−812.
(48) Barfield, M.; Karplus, M. *J. Am. Chem. Soc.* **1969**, *91*, 1−10.
(49) Pople, J. A.; Santry, D. P. *Mol. Phys.* **1964**, *8*, 1−18. Pople, J. A.; Santry, D. P. *Mol. Phys.* **1965**, *9*, 311−318.
(50) Murrell, J. N.; Gil, V. M. S. *Theor. Chim. Acta* **1966**, *4*, 114−122.
(51) Gil, V. M. S.; Teixeira-Dias, J. J. C. *Mol. Phys.* **1968**, *15*, 47−55; Gil, V. M. S.; Formosinho-Simões, S. J. S. *Mol. Phys.* **1968**, *15*, 639−643; Gil, V. M. S.; Geraldes, C. F. G. C., *Rev. Port. Quím.* **1970**, *12*, 32−35.
(52) Barfield, M.; Smith, W. B. *J. Am. Chem. Soc.* **1992**, *114*, 1574−1581.
(53) Slater, J. C. *Quantum Theory of Molecules and Solids*; McGraw-Hill: New York, 1963; Vol. 1, p 51.
(54) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc.* **1983**, *105*, 5761−5766.
(55) Mitchell, J. B. O.; Price, S. L. *Chem. Phys. Lett.* **1989**, *154*, 267−72.
(56) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960; Chapter 4.

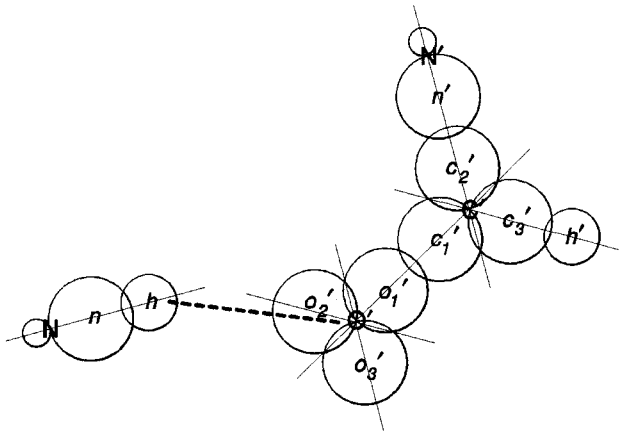

Figure 2. Schematic diagram showing representative hybrid and atomic orbitals in the formamide dimer depicted in Figure 1. For simplicity, each second-row atom has a set of three (trigonal) hybrid-type orbitals (HTOs). Only one of these HTOs is depicted for N and N′. Of particular interest here are the 1s atomic orbital h of the donor hydrogen and the HTOs o_1'−o_3' on the acceptor oxygen. The $2p_\pi$ atomic orbitals on O′ and C′ perpendicular to the plane are not depicted.

metric dependencies on ρ and θ_2,

$$S(\mathrm{ho}_1') = [S_{\mathrm{hs}'} + (2)^{1/2}\cos\theta_2 S_{\mathrm{h}\sigma'}]/3^{1/2} \quad (3a)$$

$$S(\mathrm{ho}_2') = [S_{\mathrm{hs}'} + (2)^{1/2}(3^{1/2}\cos\rho\sin\theta_2 - 2^{-1}\cos\theta_2)S_{\mathrm{h}\sigma'}]/3^{1/2} \quad (3b)$$

$$S(\mathrm{ho}_3') = [S_{\mathrm{hs}'} - (2)^{1/2}(3^{1/2}\cos\rho\sin\theta_2 + 2^{-1}\cos\theta_2)S_{\mathrm{h}\sigma'}]/3^{1/2} \quad (3c)$$

$$S(\mathrm{hp}_\pi') = \sin\rho\sin\theta_2 S_{\mathrm{h}\sigma'} \quad (3d)$$

since (by symmetry) the only nonvanishing overlap integrals are $S_{\mathrm{hs}'}$ and $S_{\mathrm{h}\sigma'}$ associated with the AOs on H and O′. These equations have the expected form. For example, $S(\mathrm{ho}_1')$ depends only on $\cos\theta_2$, and all overlap integrals are independent of the dihedral angle ρ in the linear N−H···O′═C′ arrangement ($\theta_2 = 180°$). In this arrangement the overlap integrals $S(\mathrm{ho}_2')$ and $S(\mathrm{ho}_3')$ are identical and substantially larger in magnitude than $S(\mathrm{ho}_1')$ since o_1' points away from the donor hydrogen. However, for $\theta_2 = \rho = 90°$ the overlap integrals with all three hybrid orbitals are equal, and the overlap with the $2p_\pi$ AO on oxygen assumes the maximum value.

The Pople−Santry MO formulation[49−52] implies a quadratic dependence of the coupling constant on $S(\mathrm{ho}_1')$ in eq 3a. This should lead to the dominant contribution to $^{h3}J_{\mathrm{NC}'}$ since it involves a direct (electron-mediated) mechanism between the n−h and o_1'−c_1' bond pairs in Figure 2. Therefore, the most important structural features for $^{h3}J_{\mathrm{NC}'}$ will contain a $\cos^2\theta_2$ dependence on the H···O′═C′ angle, and an exponential dependence on $r_{\mathrm{HO}'}$,[57] arising from the square of the overlap integrals ($S_{\mathrm{hs}'}$ and $S_{\mathrm{h}\sigma'}$ in eq 3a) between AOs on the donor hydrogen and acceptor oxygen.

Since the oxygen lone pairs ($o_2'$and o_3') and $2p_\pi'$ orbitals in Figure 2 have little or no density at the coupled nuclei, they could lead to indirect (electron-mediated) mechanisms.[48] These usually lead to smaller coupling contributions than direct mechanisms since they arise in higher order perturbation theory.[48,50,51] Consider the situation for trans-H-bond coupling in Figure 2. In many H-bonding situations the donor hydrogen will interact very effectively with the lone pairs $o_2'$and o_3', and the latter by angularly independent interactions with o_1', and c_1'−c_3'. As a consequence, indirect contributions to $^{h3}J_{\mathrm{NC}'}$ arising from the lone pairs on oxygen should depend on both ρ and θ_2 in eqs 3b and 3c. The form of the interactions implies that this dependence will be important for bent H···C′═O′ arrangements. Since all of the overlap integrals in eq 3 depend on $S_{\mathrm{hs}'}$ and $S_{\mathrm{h}\sigma'}$, the direct and indirect contributions to $^{h3}J_{\mathrm{NC}'}$ should have identical exponential dependencies on $r_{\mathrm{HO}'}$. Indirect contributions from the $2p_{\pi'}$ orbital on oxygen could also be implicated but would be difficult to distinguish from the lone pair contributions because of their similar dependence on ρ, θ_2, and $r_{\mathrm{OH}'}$.

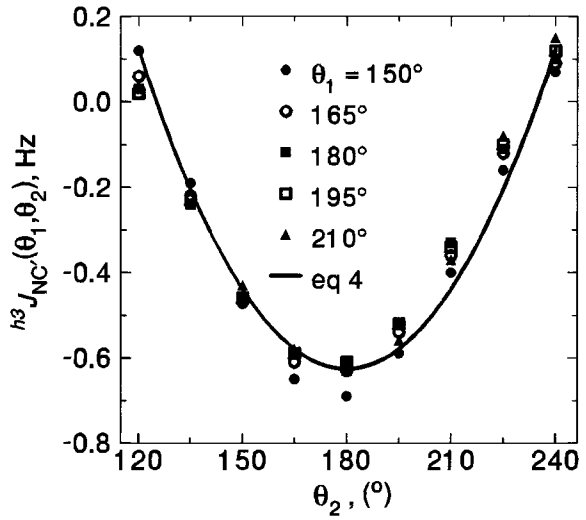

Figure 3. DFT/FPT results (symbols denote the computed points) for $^{h3}J_{\mathrm{NC}'}$ in formamide dimers plotted versus θ_2 (∠H···O′═C′) in the range 120−240° for θ_1 (∠N−H···O′) in the range 150−210°. The dimers were constrained to planar arrangements with $r_{\mathrm{HO}'} = 2.0$ Å. The solid line is a plot of eq 4.

B. Relationships of DFT/FPT $^{h3}J_{\mathrm{NC}'}$ to Structural Parameters. The following three sections present results for the dependence of $^{h3}J_{\mathrm{NC}'}$ on systematic variations of θ_1, θ_2, $r_{\mathrm{HO}'}$, and ρ. In a study of the formamide dimer, Bagno presented 3D plots showing the dependence of $^{h3}J_{\mathrm{NC}'}$ on the first three of these quantities, and an angle equivalent to the N−H···O′═C′ dihedral angle about the H···O′ H-bond.[20] He concluded that the latter is not a very important factor. Several other angles and distances were investigated during the course of this study but are not discussed here since their inclusion negligibly improved the overall correlation.

1. Dependence on θ_1 and θ_2. The DFT/FPT method was used to compute the Fermi contact contributions to $^{h3}J_{\mathrm{NC}'}$ in formamide dimers as a function of the θ_1 (∠N−H···O) and θ_2 (∠H···O═C) internal angles. The dimers were constrained to the plane. In Figure 3 the interresidue coupling constants are plotted versus θ_2 in the range 120°−240° at 15° intervals of θ_1 with $r_{\mathrm{HO}'}$ fixed at 2.0 Å. The sense of the measurement of these angles is depicted in Figure 1, for example, $\theta_1 > 180°$ range or $\theta_1 < 180°$ if the C−N bond and the H···O′ H-bond are cis and trans, respectively. Similarly, $\theta_2 > 180°$ or $\theta_2 < 180°$ if the N−H and O′═C′ bonds are cis and trans, respectively.

Figure 3 shows that $^{h3}J_{\mathrm{NC}'}$ is only slightly dependent on θ_1 in the range 150−210°. This is also apparent from the surface

(57) Mulliken, R. S.; Rieke, C. A.; Orloff, D.; Orloff, H. *J. Chem. Phys.* **1949**, *17*, 1248−1267.

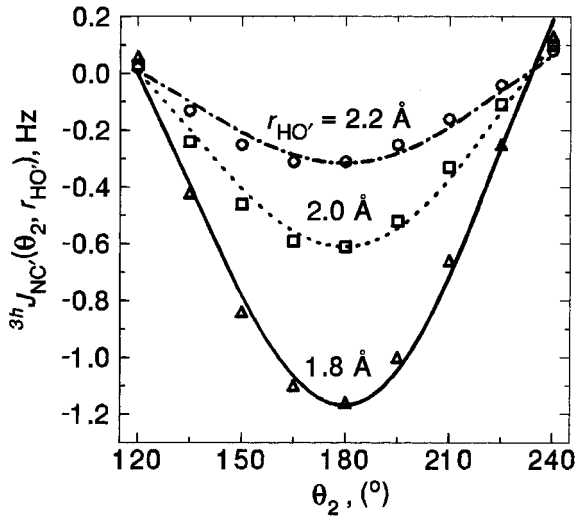

Figure 4. DFT/FPT results for $^{h3}J_{NC'}$ in formamide dimers plotted versus θ_2 in the range 120−240° with $r_{HO'} = 1.8$ (triangles), 2.0 (squares), and 2.2 Å (circles). The dimers were constrained to the plane with θ_1 fixed at 180°. The lines are plots of the results from eq 6 for the three values of $r_{HO'}$.

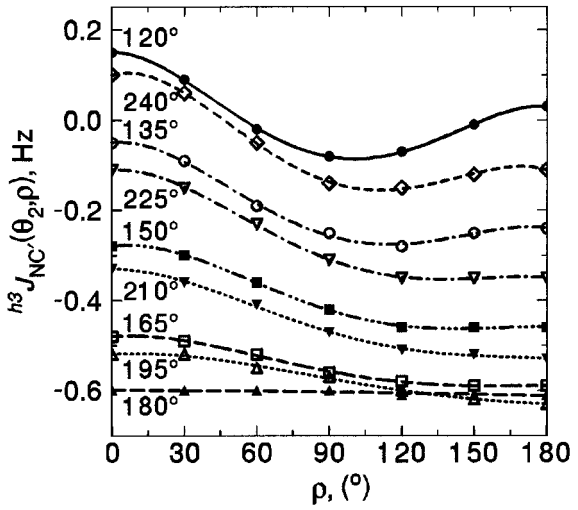

Figure 5. DFT/FPT results for $^{h3}J_{NC'}$ in formamide dimers plotted versus ρ (∠H···O′═C′−N′) in the range 0−180° at 15° increments of θ_2. For this set of calculations $r_{HO'}$ and θ_1 were fixed at 2.0 Å and 180°, respectively. The symbols represent the DFT/FPT results, and the lines represent polynomial fits of these data.

plot (Figure 3) in ref 20. Moreover, θ_1 values in the 34 H-bonds of protein G (as described in the next section) encompass a comparable range (146−207°) to those plotted in the Figure 3. Therefore, in the next two sections θ_1 is constrained at the 180° angle as other structural parameters are varied.

In Figure 3 there are substantial dependencies of $^{h3}J_{NC'}$ on the H···O′═C′ angle θ_2: $^{h3}J_{NC'}$ ranges from −0.9 Hz (θ_2 = 180°) to small positive values for substantially bent arrangements (θ_2 < 120° or > 235°). The DFT/FPT results in Figure 3 appear to follow a simple trigonometric dependence on the internal angle θ_2, and vanish for θ_2 near 130/235° for all values of θ_1. A quadratic dependence of the interresidue coupling on $S(ho_1')$ in eq 3a implies that $^{h3}J_{NC'}$ would vanish near $\theta_2 = 135/225°$ (assuming that $S_{ho'} = S_{hs'}$). The 63 computed values of $^{h3}J_{NC'}$ for the formamide dimers at $r_{OH} = 2.0$ Å were fit to a $\cos^2\theta_2$ dependence

$$^{h3}J_{NC'}(\theta_2) = -1.01\cos^2\theta_2 + 0.34 \text{ Hz}. \quad (4)$$

The standard deviation is 0.06 Hz and correlation coefficient r^2 is 0.957. The inclusion of a quadratic dependence of the overlap integral in eq 3a also implies a cos θ_2 term. However, this does not improve the correlation.

2. Dependence on $r_{HO'}$ and θ_2. The DFT/FPT results for the Fermi contact contributions to $^{h3}J_{NC'}$ in the formamide dimers were obtained at 15° intervals of θ_2 and three values of $r_{HO'}$. The dimers were constrained to the plane with $\theta_1 = 180°$. The computed data are plotted versus θ_2 in Figure 4 for $r_{HO'} = 1.8$ Å (triangles), 2.0 Å (squares), and 2.2 Å (circles). The computed values are strongly dependent on both quantities. The Pople−Santry formulation[49−52] implies a quadratic dependence on the overlap integrals $S_{ho'}$ and $S_{hs'}$ in eq 3a. Since the latter are dominated by an exponential dependence on internuclear separation,[57] the computed quantities were fit to a form combining the $\cos^2\theta_2$ and the exponential dependencies,

$$^{h3}J_{NC'}(\theta_2,r_{HO'}) = -1.45\cos^2\theta_2 \exp[-a(r_{HO'} - r_{HO'}{}^0)] + 0.16 \text{ Hz}. \quad (5)$$

Equation 5 has a standard deviation 0.12 Hz and correlation coefficient $r^2 = 0.898$. A value of 3.2 Å^{-1} was adopted for the exponent a in eq 5 as this occurs as the optimum value in several of these analyses including those for the protein G related data in the next section. Additionally, the term $r_{HO'}{}^0$ was placed in the exponential as a convenient way to avoid extremely large coefficients. In eq 5 and all subsequent equations $r_{HO'}{}^0$ is assigned the value 1.760 Å. This is the smallest H−O′ distance found for protein G as described in the next section. Cornilescu et al.[11] introduced an expression for $^{h3}J_{NC'}$ depending exponentially on the donor nitrogen−acceptor oxygen distance $r_{NO'}$ rather than $r_{HO'}$. Since θ_1 is constrained at 180° in this and the subsequent section, $r_{NO'} = r_{HO'} + r_{NH'}$, this would only have the effect of changing the coefficients in eq 5.

In the previous section it was noted that indirect contributions involving the lone pairs on oxygen could lead to a dependence of $^{h3}J_{NC'}$ on the H···O′═C′−N′ dihedral angle ρ in Figure 1. The overlap integrals between the donor hydrogen orbital h and the lone pair orbitals $o_{2'}$ and $o_{3'}$, depend on cos ρ, sin θ_2 and exponentially on $r_{HO'}$. Introducing the dependence on these parameters as described in the next section, leads to the result

$$^{h3}J_{NC'}(\theta_2,\rho,r_{HO'}) = \{-1.35\cos^2\theta_2 + [0.57\cos^2\rho + 0.14\cos\rho]\sin^2\theta_2\} \exp[-3.2(r_{HO'} - r_{HO'}{}^0)] + 0.01 \text{ Hz}. \quad (6)$$

In comparison with eq 5, the standard deviation is reduced to 0.04 Hz, and the correlation coefficient r^2 is 0.990. Equation 6 presents a substantial improvement over eq 5 and clearly shows the importance of including the dihedral angle ρ. The lines through the computed points in Figure 4 are plots of $^{h3}J_{NC'}(\theta_2,\rho,r_{HO'})$ from eq 6.

3. Dependence on θ_2 and the H···O═C−N Dihedral Angle ρ. The DFT/FPT results for the Fermi contact contributions to $^{h3}J_{NC'}$ in formamide dimer were obtained at 15° intervals of θ_2 and 30° intervals of the dihedral angle ρ. The N−H···O′ angle θ_1 and the interresidue separation $r_{HO'}$ were constrained at 180° and 2.0 Å, respectively. The symbols plotted in Figure 5

Table 1. Structural and NMR Data for 34 Formamide Dimers with Geometries (optimized donor hydrogen positions) Extracted from the Crystallographic Data for Protein G

no.	H^N–C′	$r_{HO'}$ (Å)	$r_{NO'}$ (Å)	θ_1 (deg)[a]	θ_2 (deg)[b]	ρ (deg)[c]	$^{h3}J_{NC'}$ (Hz) exptl[d]	$^{h3}J_{NC'}$ (Hz) DFT/FPT[e]	δ_H (ppm) exptl[f]	δ_H (ppm) DFT/GIAO[g]
1	Y3−T18	1.765	2.777	183.6	175.2	137.0	−0.51	−1.32	9.09	8.74
2	K4−K50	2.029	3.015	197.0	153.2	14.5	−0.42	−0.29	9.13	6.97
3	L5−T16	1.904	2.901	193.5	149.3	103.7	−0.70	−0.60	8.66	7.30
4	L7−G14	1.797	2.812	182.1	152.0	90.0	−0.68	−0.88	8.80	8.01
5	N8−V54	1.847	2.859	176.2	170.1	45.9	−0.70	−0.91	9.01	7.86
6	G9−L12	2.018	2.913	145.8	157.1	31.5	−0.33	−0.52	7.94	6.85
7	G1−L7	2.053	3.027	201.2	143.9	179.2	−0.24	−0.35	8.34	7.04
8	T16−L5	2.037	2.979	206.7	144.2	179.5	−0.38	−0.37	8.84	7.17
9	T18−Y3	2.016	2.989	201.9	150.6	146.4	−0.41	−0.47	8.95	7.16
10	A20−M1	1.954	2.932	199.2	153.9	164.7	−0.51	−0.59	9.13	7.61
11[h]	A26−D22	2.231	3.237	189.2	149.5	27.4	−0.18	−0.10	7.28	6.15
12	E27−A23	1.781	2.788	188.8	204.6	58.8	−0.54	−0.89	8.37	8.28
13	K28−A24	2.058	3.063	189.5	208.1	62.0	−0.13	−0.32	7.21	6.58
14	V29−T25	2.161	3.168	188.4	212.5	52.5	−0.21	−0.16	7.24	6.20
15	F30−A26	1.900	2.905	189.8	199.7	40.9	−0.64	−0.61	8.62	7.59
16	K31−E27	1.768	2.769	168.7	202.7	66.0	−0.72	−0.96	9.05	8.33
17	Q32−K28	1.930	2.937	170.7	210.5	66.5	−0.19	−0.46	7.52	7.10
18	Y33−V29	1.949	2.957	189.4	211.8	51.3	−0.27	−0.36	8.09	7.10
19	A34−F30	1.917	2.923	170.5	211.3	63.7	−0.49	−0.48	9.23	7.13
20	N35−K31	1.916	2.927	186.7	210.7	62.3	−0.31	−0.47	8.42	7.29
21	D36−Q32	1.760	2.769	187.7	201.8	62.0	−0.60	−1.03	8.83	8.50
22[h]	N37−Y33	2.103	3.086	197.7	210.6	60.3	−0.19	−0.20	7.41	6.21
23[i]	G38−N35	2.065	3.061	166.5	249.0	81.6	<0.10	0.00	7.82	5.78
24[i]	V39−A34	1.843	2.846	169.1	146.9	109.1	−0.34	−0.72	8.15	7.75
25	E42−T55	1.929	2.881	204.9	208.8	159.9	−0.43	−0.63	8.27	7.59
26	T44−T53	1.952	2.925	200.7	148.7	169.2	−0.53	−0.53	9.41	7.61
27	D46−T51	1.902	2.903	191.9	153.4	114.1	−0.36	−0.69	7.60	7.51
28[i]	T49−D46	2.382	3.359	161.7	250.9	80.3	<0.10	0.01	7.03	4.95
29	T51−D46	2.312	3.273	202.2	144.0	152.0	−0.22	−0.15	7.43	5.97
30	F52−K4	1.808	2.811	190.6	197.1	42.4	−0.70	−0.94	10.40	8.24
31	T53−T44	1.967	2.926	203.5	136.9	133.7	−0.61	−0.36	9.19	7.30
32	V54−I6	2.000	2.969	201.0	193.7	9.2	−0.39	−0.49	8.29	6.99
33	T55−E42	2.024	2.998	199.8	154.4	169.4	−0.51	−0.50	8.36	7.15
34	E56−N8	2.059	3.043	197.4	206.2	47.4	−0.33	−0.34	7.91	6.63

[a] ∠N−H···O. [b] ∠H···O═C. [c] ∠H···O═CN. [d] Experimental data from ref 11. [e] Coupling constants based on DFT/FPT at the UB3PW91/6-311G** level. [f] Experimental ^{1}H chemical shifts from B. E. Ramiriz, private communication, 2001. [g] Isotropic ^{1}H chemical shifts at the B3PW91/6-311G** level referenced to TMS at 31.78 ppm. [h] Hydrogen bonds no. 11−22 are in the α-helix region. All others are β-sheet or irregular. [i] Irregular regions.

represent the computed $^{h3}J_{NC'}$ as a function of ρ for values of θ_2 in the range 120−240°. The computed values become increasingly dependent on ρ with increasing H···O′═C′−N′ nonlinearity. On the basis of eqs 3b and 3c, a plausible form for the angular part of the indirect coupling contribution is $\{[A \cos^2 \rho + B \cos \rho + C]\sin^2 \theta_2\}$, where the coefficients *A*, *B*, and *C* are determined empirically. The exponential dependence should be identical for both the direct and indirect contributions since all overlap integrals in eqs 3a−d depend on $r_{HO'}$. Combining the indirect dependence on ρ and θ_2 with the direct dependence on θ_2 leads to the result,

$$^{h3}J_{NC'}(\theta_2,\rho) = \{-1.43 \cos^2 \theta_2 + [0.43 \cos^2 \rho + 0.35 \cos \rho + 0.15] \sin^2 \theta_2\} \times \exp[-3.2(r_{HO'} - r_{HO'}{}^0)] + 0.05 \text{ Hz.} \quad (7)$$

The standard deviation is 0.04 Hz and $r^2 = 0.976$ in eq 7. The exponential term is just a constant (computations were performed with $r_{HO'}$ fixed at 2.0 Å), included in eq 7 to be consistent with all other expressions for $^{h3}J_{NC'}$. The indirect term, which contains the polynomial in cos ρ, is generally smaller than the direct term and vanishes for $\theta_2 = 180°$. However, this term becomes increasingly important as θ_2 approaches 90°. The cos ρ term in eq 7 is necessary to account for the $^{h3}J_{NC'}$ sign changes near $\rho = 90°$in Figure 5.

IV. Models for Protein Structures Based on X-ray Crystallographic Data

A. Formamide Dimers. For each of the 34 H-bonds having observable coupling in the B1 immunoglobulin binding domain of protein G,[11] a formamide dimer was generated using the 1.1 Å crystallographic structure (PDB code 1IGD).[58,59] The crystallographic techniques do not usually lead to accurate hydrogen atom positions. This is a particular concern in this study because of the expected sensitivity of interresidue coupling constants on the position of the donor hydrogen atom. Therefore, a hybrid model was adopted for the formamide dimers. The relative positions of the donor O═C−N and acceptor O′═C′−N′ atoms were held at the X-ray positions, and hydrogen atoms were added (r_{NH} = 1.100 Å, ∠O═C−H = 114.0°, ∠C−N−H = 120.5°). The positions of the hydrogen atoms on the donor nitrogen were then fully optimized at the B3PW91/6-31G** level. The relevant structural data ($r_{NO'}$, $r_{HO'}$, θ_1, θ_2, and ρ) in the H-bonding regions are given columns 3−7 of Table 1. These 34 formamide dimers were used to compute the FC contributions to the coupling constants $^{h3}J_{NC'}$ and the isotropic amide ^{1}H chemical shifts δ_H. These are entered and compared with the experimental NMR data in the last four columns of Table 1. The first column lists the H-bonds in the order in which they

(58) Derrick, J. P.; Wigley, D. B. *J. Mol. Biol.* **1994**, *243*, 906−918.
(59) Gallegher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. *Biochemistry* **1994**, *33*, 4721−4729.

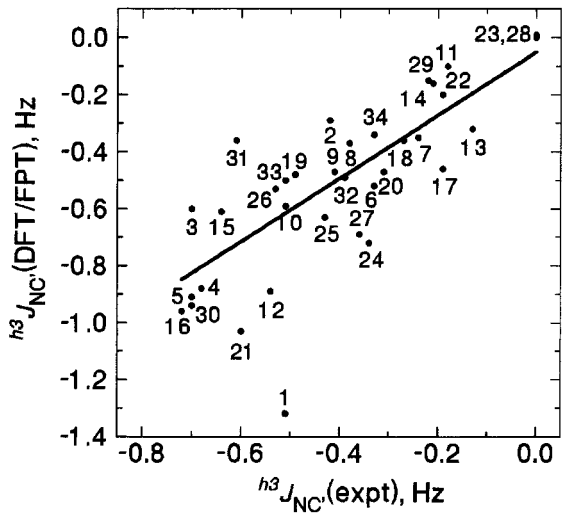

Figure 6. The DFT/FPT data for $^{h3}J_{NC'}$ in the 34 formamide dimers extracted from protein G plotted versus the experimental data.[11] The numbering of the H-bonds is given in Table 1. The solid line is the linear regression result $^{h3}J_{NC'}$(DFT/FPT) from eq 8.

occur in the protein. The second column specifies the residue numbers of the donor and acceptor H-bonds. Twelve of the H-bonds are in the α-helix region (11−22), the three connecting different regions are designated *irregular* (23, 24, and 28), and the remaining 19 occur in β-sheet regions.

Plotted in Figure 6 are the calculated $^{h3}J_{NC'}$ versus the experimental data[11] for the protein G-based dimers. The H-bonds are numbered as they occur in Table 1. The coupling constants in two H-bonding bonding regions were too small to be observed (entries 23 and 28 in Table 1). For inclusion in the analyses these were assigned values 0.0 Hz. The solid line in Figure 6 is a plot of the linear regression result,

$$^{h3}J_{NC'}(\text{DFT/FPT}) = 1.12\ ^{h3}J_{NC'}(\text{expt}) - 0.05\ \text{Hz}, \quad (8)$$

with standard deviation 0.16 Hz and correlation coefficient $r^2 = 0.673$. Not included in the regression analysis was the data for the extreme outlier (entry no. 1 in Table 1). This H-bond is in the β-sheet region of protein G and has one of the smallest values of both $r_{OH'}$ and $r_{NO'}$ and an essentially linear N−H···O′═C′ arrangement. On the basis of the analyses of section III it *should* exhibit the coupling of maximum magnitude. There are four other comparably short H-bonds in protein-G (H-bonds no. 4, 12, 16 and 21 in Table 1). Although these are nonlinear by at least 20°, all have somewhat larger magnitudes of $^{h3}J_{NC'}$ than H-bond no. 1.

This level of correspondence between the calculated results and experimental data in eq 8 is not unreasonable: Probably, the greatest disparities arise from the following: (1) uncertainties in the atomic positions even at 1.1 Å resolution could lead to significant errors in $^{h3}J_{NC'}$(DFT/FPT). At this resolution the approximate 0.05−0.1 Å rms error in the atomic coordinates implies from Figure 4 an approximate 0.1−0.2 Hz error in the computed coupling constants. (2) Perhaps of even greater importance are the effects of conformational averaging in solution. Better agreement between the calculated and experimental values in the α-helix region than for the β-sheet regions (standard deviations 0.16 and 0.23 Hz, respectively) could be evidence of greater structural rigidity in the α-helix region than for the β-sheet regions. (3) Another source of disparity is the experimental errors of the measurements. An indication of this is provided by the approximate 0.05 Hz average deviation between the two experimental data sets for ubiquitin.[9,10] (4) The question of the adequacy of the simple formamide dimer model to describe the complex problem presented by the proteins, will be addressed here (in part) by computations for larger segments based on the protein G structure.

The procedures described above for investigating scalar couplings in protein G were repeated for ubiquitin using formamide dimers extracted from the 1.8 Å X-ray structure.[60] The several DFT/FPT values for $^{h3}J_{NC'}$ reported by Bagno,[20] are extended to all 31 backbone H-bonds for which experimental data were reported.[9,10] The relevant structural data and inter-residue coupling constants for ubiquitin are included in the Supporting Information. Linear regression analysis of the calculated $^{h3}J_{NC'}$ versus the average of the two sets of experimental data, leads to the result,

$$^{h3}J_{NC'}(\text{DFT/FPT}) = 1.41^{h3}J_{NC'}(\text{expt}) + 0.09\ \text{Hz}, \quad (9)$$

where the standard deviation is substantially larger (0.23 Hz), and the correlation coefficient is smaller ($r^2 = 0.500$) than for protein G in eq 8. The poorer correspondence for the ubiquitin data certainly reflects the 0.1−0.2 Å precision in the 1.8 Å crystallographic structure.[60] From Figure 4 this implies errors of 0.1−0.3 Hz in the computed values of $^{h3}J_{NC'}$.

The structural constraints, which were imposed in constructing the model formamide dimers in section III.B, are not implicit in the dimers based on crystallographic data and optimized donor H-atom positions. Therefore, it was of particular interest to explore the applicability of the structural correlations. The DFT/FPT data for the FC contribution to $^{h3}J_{NC'}$ in the 34 formamide dimers were fit empirically to the simplest possible equation which combines both angular (θ_2) and exponential dependencies on H···O′ distances,

$$^{h3}J_{NC'}(\theta_2,r_{HO'}) = -1.29\cos^2\theta_2 \exp[-3.2(r_{HO'} - r_{HO'}{}^0)] + 0.04\ \text{Hz}, \quad (10)$$

where $r_{HO'}{}^0 = 1.760$ Å is the minimum computed H−O′ distance in Table 1. Equation 10 provides a good fit to the computed data (the standard deviation is 0.07 Hz and correlation coefficient $r^2 = 0.955$). In Figure 7a the DFT/FPT results from Table 1 are plotted (points) versus the results from eq 10. Each H-bond is labeled α, β, or i to denote the α-helix, β-sheet, and irregular regions, respectively. The data in Table 1 and Figure 7a fall into the following two groups: (1) those for which $\rho < 80°$ including all data in the α-helix regions of protein G ($\rho = 27-66°$), six β-sheet entries for which ρ is in the range 9−47°, and (2) those for which $\rho \geq 80°$ which includes data from both β-sheet and irregular regions. Linear regression analysis for the twelve α-helix H-bonds in Table 1 leads to

$$^{h3}J_{NC'}(\theta_2,r_{HO'},\alpha\text{-helix}) = -1.30\cos^2\theta_2 \exp[-3.2(r_{HO'} - r_{HO'}{}^0)] + 0.11\ \text{Hz}, \quad (11a)$$

with a standard deviation 0.02 Hz and $r^2 = 0.996$. The inclusion

(60) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.

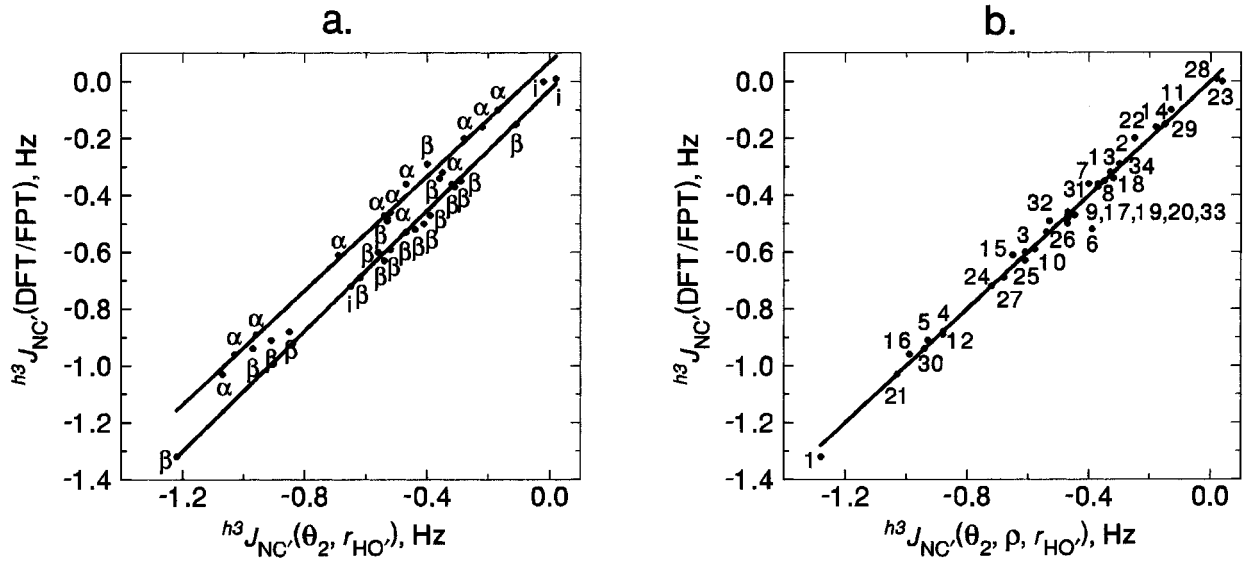

Figure 7. (a) The DFT/FPT results for $^{h3}J_{NC'}$ in the 34 formamide dimers extracted from protein G plotted versus $^{h3}J_{NC'}(\theta_2,r_{HO'})$ from eq 10. The data fall into two distinct groups depending on the dihedral angle ρ. The upper set comprises all α-helix entries and several β-sheet H-bonds. The lower data set consists entirely of β-sheet and the three irregular (designated i) H-bonds having $\rho > 80°$. The solid lines correspond to the linear regression results from eqs 11a and 11b. (b) The DFT/FPT results for $^{h3}J_{NC'}$ in the 34 formamide dimers extracted from protein G plotted versus $^{h3}J_{NC'}(\theta_2,\rho,r_{HO'})$ from eq 12. The H-bonds are labeled as they occur in Table 1. In this case the extreme outlier is the H-bond no. 6 which has the smallest N−H···O′ angle θ_1 (146°) in Table 1.

of the other H-bonds with $\rho < 80°$ in Table 1 gives a slightly poorer correlation. An analysis of the 17 H-bonds with $\rho \geq 80°$ in β-sheet regions leads to comparable results for the region, which is almost entirely β-sheet

$$^{h3}J_{NC'}(\theta_2,r_{HO'},\beta\text{-sheet}) = -1.37\cos^2\theta_2\exp[-3.2(r_{HO'} - r_{HO'}{}^0)] + 0.01\ \text{Hz}, \quad (11b)$$

with a standard deviation 0.03 Hz and $r^2 = 0.994$. The solid lines in Figure 7a are plots of the results from eqs 11a and 11b. Since these two equations present excellent correlations and depend on only two structural parameters, they probably provide the most practical form for connecting interresidue coupling $^{h3}J_{NC'}$ to structural data.

From the analysis of the indirect interactions involving the lone pairs on oxygen in section III, the separation of $^{h3}J_{NC'}$ into two groups in Figure 7a must arise from the dependence on the dihedral angle ρ. Those species having dihedral angles less than about 80° (including all α-helix species) in Figure 5 will have more positive $^{h3}J_{NC'}$ than those with larger dihedral angles. This dependence can be introduced via inclusion of the trigonometric form introduced in section III.B,

$$^{h3}J_{NC'}(\theta_2,\rho,r_{HO'}) = \{-1.31\cos^2\theta_2 + [0.62\cos^2\rho + 0.92\cos\rho + 0.14]\sin^2\theta_2\} \times \exp[-3.2(r_{HO'} - r_{HO'}{}^0)]\ \text{Hz}, \quad (12)$$

with standard deviation 0.03 Hz and $r^2 = 0.989$. The constant term is not included in eq 12 since it is less than 0.01 Hz. The DFT/FPT data for the 34 dimers are plotted in Figure 7b versus the $^{h3}J_{NC'}(\theta_2,\rho,r_{HO'})$ from eq 12. The H-bonds are numbered as they occur in Table 1. Note that the extreme outlier (by about 0.1 Hz) occurs for H-bond no. 6, which (by far) has the smallest value of the N−H···O′ angle θ_1 (146°, the next smallest value is 161°) in Table 1. It is fortunate that large deviations of θ_1 from 180° are unusual, thereby avoiding complexity of additional terms in the structural analyses. Equation 12 provides an excellent representation of the structural dependence of interresidue coupling in the protein G model compounds.

The 34 *experimental* $^{15}N-^{13}C'$ coupling constants in Table 1 were also fit empirically to an equation which uses the same functional form as eq 12,

$$^{h3}J_{NC'}(\theta_2,\rho,r_{HO'}) = \{-0.70\cos^2\theta_2 + 0.74\cos\rho\sin^2\theta_2\}\exp[-3.2(r_{HO'} - r_{HO'}{}^0)] - 0.10\ \text{Hz}. \quad (13)$$

The standard deviation is 0.12 Hz and correlation coefficient $r^2 = 0.656$ which is slightly better than the calculated results versus experimental data in eq 8. The analysis of interresidue coupling in protein G[11] was based on averaged structural data from three crystallographic structures. The empirical expression proposed for $^{h3}J_{NC'}$ included only an exponential dependence on the N···O′ distance $r_{NO'}$. The use of their averaged structural data[11] and the dihedral angles from Table 1 leads to additional improvement in the dependence on three structural parameters including $r_{NO'}$

$$^{h3}J_{NC'}(\theta_2,\rho,r_{NO'}) = \{-0.84\cos^2\theta_2 + [-0.66\cos^2\rho + 0.17\cos\rho - 0.35]\sin^2\theta_2\} \times \exp[-3.2(r_{NO'} - r_{NO'}{}^0)] + 0.06\ \text{Hz}, \quad (14)$$

where the standard deviation is 0.09 Hz and $r^2 = 0.838$. The quantity $r_{NO'}{}^0$ was taken to be 2.88 Å, which is the shortest distance reported in the averaged data.[11] Perhaps, the better agreement for eq 14 arises from use of structural data averaged over several crystallographic structures as this could better simulate the motional averaging in solution. Since eqs 13 and 14 are based on the experimental data, they should be of use in studies of other proteins. Although, the dependence on the dihedral angle ρ could probably be neglected in eqs 12−14,

Table 2. Structural and NMR Data with Geometries Extracted from 27 Segments (Optimized Donor Hydrogen Positions) Extracted from the Crystallographic Data for Protein G

no.	H^N–C′	$r_{HO'}$ (Å)	$r_{NO'}$ (Å)	θ_1 (deg)[a]	θ_2 (deg)[b]	ρ (deg)[c]	$^{h3}J_{NC'}$ (Hz) exptl[d]	$^{h3}J_{NC'}$ (Hz) DFT/FPT[e]	δ_H (ppm) exptl[f]	δ_H (ppm) DFT/GIAO[g]
1	Y3−T18	1.761	2.777	175.7	174.5	157.8	−0.51	−1.38	9.09	10.04
2	K4−K50	2.022	3.015	164.6	153.8	11.4	−0.42	−0.38	9.13	8.26
3	L5−T16	1.893	2.901	167.8	150.8	102.6	−0.70	−0.62	8.66	8.17
4	L7−G14	1.794	2.812	178.9	152.5	87.7	−0.68	−0.89	8.80	9.23
5	N8−V54	1.844	2.859	172.4	170.2	52.0	−0.70	−1.00	9.01	8.78
6	G9−L12	2.073	2.913	138.7	154.2	32.2	−0.33	−0.35	7.94	7.35
7	G14−L7	2.055	3.027	158.5	144.6	179.9	−0.24	−0.36	8.34	8.12
8	T16−L5	2.047	2.979	150.6	143.7	177.5	−0.38	−0.39	8.84	8.15
9	T18−Y3	2.007	2.966	155.9	150.4	148.7	−0.41	−0.53	8.95	8.35
10	A20−M1	1.974	2.932	155.3	151.5	163.3	−0.51	−0.62	9.13	9.06
11[h]	A26−D22	2.252	3.237	163.4	146.9	29.8	−0.18	−0.14	7.28	7.11
12	E27−A23	1.806	2.788	162.0	151.5	60.2	−0.54	−0.83	8.37	8.60
13	K28−A24	2.077	3.063	163.3	148.1	61.9	−0.13	−0.32	7.21	7.11
14	V29−T25	2.159	3.168	175.4	147.9	54.5	−0.21	−0.18	7.24	5.81
15	F30−A26	1.910	2.902	165.2	157.0	45.0	−0.64	−0.61	8.62	8.14
16	K31−E27	1.773	2.769	169.3	157.0	69.1	−0.72	−0.98	9.05	7.82
17	Q32−K28	1.935	2.937	171.2	149.3	69.1	−0.19	−0.46	7.52	6.50
18	Y33−V29	1.973	2.957	162.3	144.4	53.4	−0.27	−0.37	8.09	7.49
19	A34−F30	1.933	2.923	163.9	145.5	65.0	−0.49	−0.40	9.23	7.57
20	N35−K31	1.937	2.927	164.4	146.0	63.0	−0.31	−0.46	8.42	7.19
21	D36−Q32	1.766	2.760	165.0	155.3	63.6	−0.60	−1.01	8.83	9.17
22[h]	N37−Y33	2.259	3.086	137.9	141.4	58.1	−0.19	−0.13	7.41	5.67
23[i]	G38−N35	2.078	3.061	163.0	109.1	80.8	<0.10	−0.01	7.82	6.69
24[i]	V39−A34	1.827	2.846	172.3	147.2	103.5	−0.34	−0.76	8.15	9.76
25	E42−T55	1.927	2.881	155.1	149.9	158.1	−0.43	−0.71	8.27	8.57
26	T44−T53	1.973	2.925	154.9	149.1	175.8	−0.53	−0.54	9.41	8.70
27	D46−T51	1.911	2.903	164.1	151.4	111.5	−0.36	−0.71	7.60	8.34
28[i]	T49−D46	2.404	3.359	157.0	107.4	79.5	<0.10	0.00	7.03	5.06
29	T51−D46	2.301	3.273	159.2	144.8	152.4	−0.22	−0.19	7.43	6.86
30	F52−K4	1.797	2.811	174.2	163.9	48.1	−0.70	−0.96	10.40	8.84
31	T53−T44	1.970	2.926	155.4	137.3	134.7	−0.61	−0.43	9.19	8.44
32	V54−I6	1.985	2.969	161.7	166.7	15.4	−0.39	−0.59	8.29	8.41
33	T55−E42	2.027	2.998	158.9	153.3	166.2	−0.51	−0.54	8.36	8.25
34	E56−N8	2.078	3.043	158.0	157.9	51.3	−0.33	−0.30	7.91	6.87

[a] ∠N−H···O. [b] ∠H···O═C. [c] ∠H···O═CN. [d] Experimental data from ref 11. [e] Coupling constants based on DFT/FPT at the UB3PW91/6-311G** level. [f] Experimental ^{1}H chemical shifts from B. E. Ramiriz, private communication, 2001. [g] Isotropic ^{1}H chemical shifts at the B3PW91/6-311G** level referenced to TMS at 31.78 ppm. [h] Hydrogen bonds no. 11−22 are in the α-helix region. All others are β-sheet or irregular. [i] Irregular regions.

the dependence on $\cos^2\theta_2$ clearly cannot since $^{h3}J_{NC'}$ will vanish independently of the H-bond distance if θ_2 becomes substantially nonlinear.

B. Protein Sequences. The formamide dimers provide a convenient model for exploring the conformational and distance dependencies of the NMR parameters. Since this model omits the electronic influence of nearby atoms, does it provide an adequate representation of the H-bonding regions of proteins? A partial answer is provided here by an investigation of NMR parameters in sequences of residues extracted from the 1IGD crystallographic structure for protein G.[58,59] These sequences are substantially larger than the formamide dimers since amino acid residues are included on at least one side and (in as many as 27 cases) on both sides of the donor and acceptor residues. To reduce the computational demands, the side chains were removed giving hydrogen-bonding segments containing 34−42 atoms. The positions of all donor hydrogen atoms were fully optimized at the B3PW91/6-31G** level with all other atoms constrained to the crystallographic positions. Entered in the first four columns of Table 2 are the relevant structural data ($r_{NO'}$, $r_{HO'}$, θ_1, θ_2, and ρ) in the H-bonding regions. The 27 sequences were used to compute the FC contributions to the scalar couplings $^{h3}J_{NC'}$ and the amide proton isotropic chemical shifts δ_H. These are compared with the experimental NMR data in the last four columns of Table 2. The DFT/FPT data for the FC contribution to the coupling in the 27 segments containing 34 distinct coupling constants leads to an expression which differs by only 0.03 Hz from the formamide dimer data in eq 10. This provides justification for the simpler model but it is disappointing that there was no improvement since the computations for the sequences were more time-consuming. Not surprisingly, there are some substantial differences in calculated amide ^{1}H chemical shifts as discussed in the next section.

V. Structural Dependence of Donor ^{1}H Isotropic Shifts

Entered in the last column of Table 1 are the calculated DFT/GIAO results for the amide ^{1}H chemical shifts δ_H. These are isotropic values for the formamide dimers and are referenced indirectly to TMS. These are plotted in Figure 8 versus experimental data for protein G.[61] The H-bonds are labeled by their order of occurrence in Table 1. The linear regression results are given by the solid line in Figure 8,

$$\delta_H(\text{DFT/FPT}) = 0.76\,\delta_H(\text{expt}) + 0.82\ \text{ppm}, \qquad (15)$$

with standard deviation 0.58 ppm and $r^2 = 0.512$. Hydrogen bond no. 28 is the most significant outlier and was not included in this fit. This H-bonding situation is a most unusual arrangement with substantially larger values of $r_{HO'}$ and θ_2 (the donor N−H bond and the acceptor O═C bond are almost at right angles) than any other entries in Table 1. The poor correlation

(61) Ramirez, B. E. 2001. Private communication.

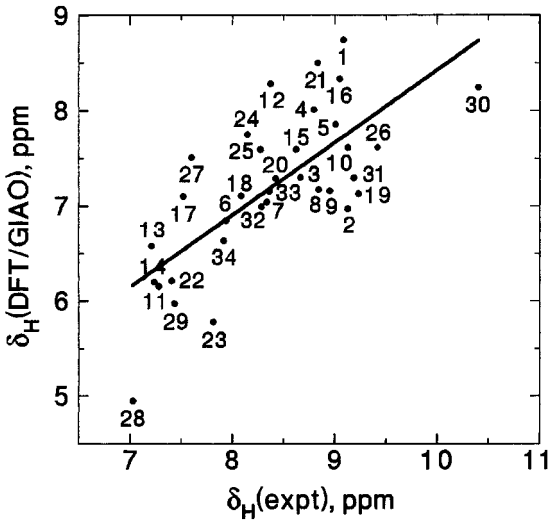

Figure 8. The DFT/GIAO results for isotropic shift of the donor proton δ_H in the 34 formamide dimers extracted from protein G plotted versus the experimental data.[61] The H-bond numbering is given in Table 1. The solid line is the linear regression result for δ_H(DFT/GIAO) from eq 15.

between calculated results for the dimer pairs and the experimental data for protein G is not surprising. Although errors in chemical shift measurements (as percentages) are substantially smaller than for the small interresidue coupling constants, they are known to be quite sensitive to the nature and conformations of remote substituents.[40,62] These are totally absent in the simple dimer model. Moreover, the effects of vibrational averaging can be substantial.[63,64] The DFT/GIAO results for amide ^{1}H shifts in the protein sequences are in worse agreement (standard deviation 0.82 ppm) with the experimental data than for the formamide dimers. Clearly, adequate prediction of the chemical shifts in proteins will require calculations with the side chains included.

Good correlations between trans-H bond coupling constants and the amide ^{1}H chemical shift have been noted both experimentally[2,9] and computationally.[4,8,19] This suggests common structural dependencies for these very different NMR parameters. Therefore, it was of interest to analyze the DFT/GIAO isotropic ^{1}H chemical shifts for the formamide dimers in terms of the structural parameters (θ_2, $r_{HO'}$, and ρ) which were found to be important for $^{h3}J_{NC'}$. The simplest correlation of δ_H in Table 1 with θ_2 and $r_{HO'}$ leads to

$$\delta_H(\theta_2,r_{HO'}) = 3.59\cos^2\theta_2\exp[-a(r_{HO'} - r_{HO'}{}^0)] + 5.32 \text{ ppm}, \quad (16)$$

(SD = 0.30 ppm, $r^2 = 0.875$) where $r_{HO'}{}^0 = 1.760$ Å, and $a = 2.0$ is the optimized value appearing in the subsequent analysis which includes terms for the dependence on the dihedral angle ρ. The δ_H(DFT/GIAO) data from Table 1 are plotted in Figure 9a versus the results from eq 16. The H-bonds are labeled α, β, and i to denote α-helix, β-sheet, and irregular, respectively. In comparison with $^{h3}J_{NC'}$ in Figure 7a there is no obvious separation of the amide ^{1}H shifts into essentially α-helix and β-sheet regions in Figure 9a. However, the inclusion of a conformational dependence on ρ gives a similar improvement,

$$\delta_H(\theta_2,\rho,r_{HO'}) = \{4.81\cos^2\theta_2 + [3.01\cos^2\rho - 0.84\cos\rho + 1.75]\sin^2\theta_2\} \times \exp[-2.0(r_{HO'} - r_{HO'}{}^0)] + 4.06 \text{ ppm}. \quad (17)$$

with standard deviation 0.09 ppm and $r^2 = 0.987$. The δ_H(DFT/GIAO) data from Table 1 are plotted (solid line) in Figure 9b versus the results from eq 17. The H-bonds are labeled as they occur in Table 1. Equations 12 and 17 provide excellent correlations of the computed $^{h3}J_{NC'}$ and δ_H, respectively, with the three structural parameters. However, as might be expected on comparing Figures 7a and 9a, there are some differences in the dependence on ρ. The coefficients of the $\cos^2\theta_2$ and the $\cos^2\rho$ terms are the same in eq 17 and opposite in eq 12. This difference could account for the somewhat poorer correlation between the computed interresidue coupling constants and δ_H found here for the formamide dimers ($r^2 = 0.907$) compared with the DFT results noted recently for DNA triplets[8] ($r^2 \geq 0.999$).

VI. Conclusions

Methods of density functional theory at the B3PW91/6-311G** level were used to investigate the structural dependence

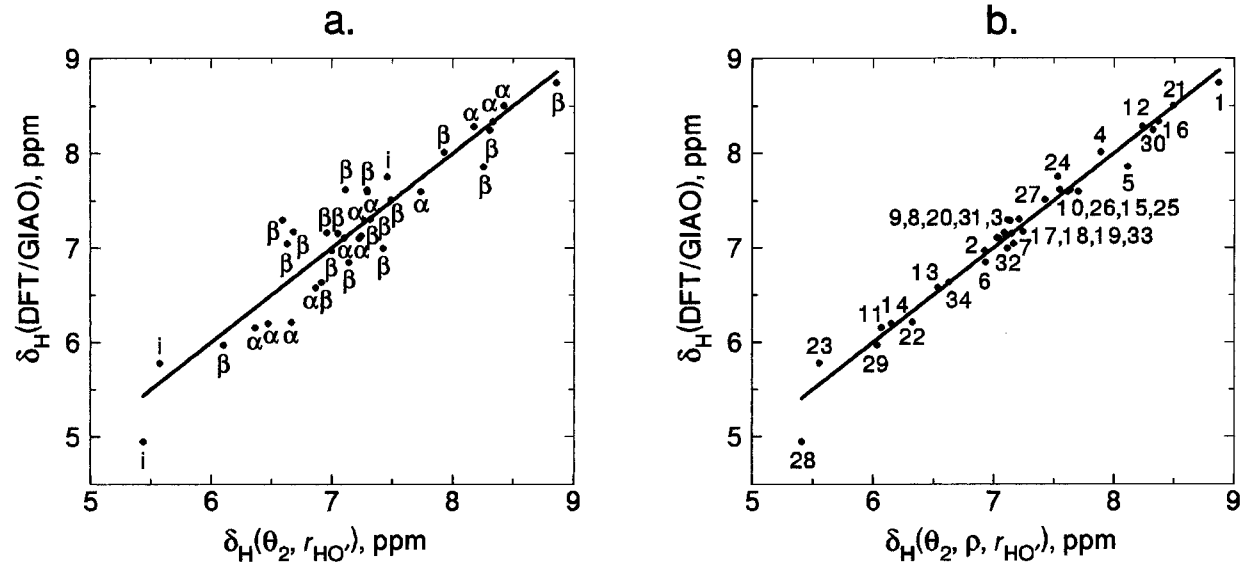

Figure 9. (a) The DFT/GIAO results for δ_H in the 34 formamide dimers extracted from protein G plotted versus $\delta_H(\theta_2,r_{HO'})$ from eq 16. The H-bonds are labeled α-helix, β-sheet, and irregular (i). (b) The δ_H(DFT/GIAO) from Table 1 are plotted versus the $\delta_H(\theta_2,\rho,r_{HO'})$ from eq 17.

of coupling constants and chemical shifts in the H-bonding regions of proteins. An analysis of the overlap integrals between the donor hydrogen and the four orbitals on the acceptor oxygen, and a systematic variation of structural parameters in formamide dimer indicated that $^{h3}J_{NC'}$ should primarily depend on the following quantities. (1) An exponential dependence on the H···O′ internuclear separation, $r_{HO'}$, (2) A $\cos^2\theta_2$ dependence on the H···O′═C′ angle, θ_2, and (3) a smaller contribution associated with the H···O′═C′−N′ dihedral angle, ρ. On the basis of the 1.1 Å crystallographic structural data for the B1 domain of streptococcal protein G, a set of 34 formamide dimers was generated, and the positions of donor H-atoms were optimized at the B3PW91/6-31G** level. The procedure was repeated for ubiquitin. The dimers were used to compute the NMR parameters. The computed $^{h3}J_{NC'}$ for protein G are in better agreement with the experimental data than those for the ubiquitin results since the latter were based on a lower-resolution (1.8 Å) crystallographic structure. The DFT/FPT results in these dimers correlate quite well with θ_2 and $r_{OH'}$ $\{^{h3}J_{NC'}(\theta_2,r_{OH'}) = -1.3\cos^2\theta_2\exp[-3.2(r_{OH'} - 1.76)]$ Hz$\}$. The correlation is improved by recognizing the importance of the dihedral angle $\rho \equiv \angle$H···O′═C′−N′, which (to a large extent) distinguishes α-helix from β-sheet arrangements. To investigate the limitations of the use of formamide dimers to represent the H-bonding regions of proteins, the calculations of NMR parameters were repeated for sequences extracted from the protein G crystallographic structure. The computed $^{h3}J_{NC'}$ were almost identical to those obtained for the dimers, providing some justification for the model. The recognition of the dependence of $^{h3}J_{NC'}$ on the internal angle θ_2 and the dihedral angle ρ in addition to the H-bond distance r_{HO} (or $r_{NO'}$) is important in the interpretation of external effects such as pressure.[15] From Figure 4 it can be seen that for $r_{HO} = 2$ Å, a 15° change in θ_2 could lead to ±0.2 Hz change in $^{h3}J_{NC'}$. The donor ^{1}H chemical shifts δ_H based on DFT and the GIAO (gauge including atomic orbital) methods also exhibit excellent correlation with these structural features and with computed coupling constants.

Acknowledgment. Discussions with Prof. Stephan Grzesiek on this subject are gratefully acknowledged.

Supporting Information Available: A table (Table 3) with structural data and computed and experimental values of $^{h3}J_{NC'}$ for 31 hydrogen bonds in ubiquitin (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA012674V

(62) Barfield, M. *J. Am. Chem. Soc.* **1995**, *117*, 2862−2876; Barfield, M.; Fagerness, P. *J. Am. Chem. Soc.* **1997**, *119*, 8699−8711; Barfield, M. *Magn. Reson Chem.* **1998**, *36*, S93−S103.
(63) Ruud, K.; Åstrand, P.-O.; Taylor, P. R. *J. Am. Chem. Soc.* **2001**, *123*, 4826−4833.
(64) Jordan, M. J. T.; Toh, J. S.-S.; Del Bene, J. *Chem. Phys. Lett.* **2001**, *346*, 288−292.