

Chemical Shifts in Proteins: A Shielding Trajectory Analysis of the Fluorine Nuclear Magnetic Resonance Spectrum of the *Escherichia coli* Galactose Binding Protein Using a Multipole Shielding Polarizability–Local Reaction Field–Molecular Dynamics Approach[†]

John G. Pearson,^{‡,§} Eric Oldfield,^{*,§} Frederick S. Lee,^{||,⊥} and Arieh Warshel^{||}

Contribution from the Department of Chemistry, University of Illinois at Urbana–Champaign, 505 South Mathews Avenue, Urbana, Illinois 61801 and the Department of Chemistry, University of Southern California, Los Angeles, California 90089-1062

Received November 23, 1992

Abstract: We report the first successful calculation of the fluorine nuclear magnetic resonance spectrum of a protein, the galactose binding protein from *Escherichia coli*, labeled with [5-¹⁹F]tryptophan. Our results indicate that the experimental ¹⁹F chemical shifts are dominated by weak, or long-range, electrical interactions, which can be calculated by using the responses of the shielding tensor elements to the uniform field components ($\partial\sigma_{\alpha\beta}/\partial E_x$) and the nonuniform or gradient terms ($\partial\sigma_{\alpha\beta}/\partial V_{ij}$), together with the average values of the fields (E_x) and field gradients (V_{ij}) obtained from molecular dynamics (MD) trajectories. A series of “shielding trajectories”, $\Delta\sigma(E_x, V_{xx}, V_{yy}, V_{zz}) f(\tau)$, are obtained, and the mean values, $\Delta\bar{\sigma}$, correlate well with the actual shift pattern and overall shift range observed experimentally (Luck, L. A.; Falke, J. J. *Biochemistry* 1991, 30, 4248–4256). The computed ¹⁹F NMR shielding of the pentapeptide Gly-Gly-[5-F]Trp-Gly-Gly in H₂O is close to, but somewhat more shielded than, that of the denatured protein. Almost all computed ¹⁹F chemical shifts are upfield of the field-free value, in accord with the results of *ab initio* calculations. The chemical shifts calculated are sensitive to the charge chosen for F in the LRF–MD trajectories, and best agreement with experiment is obtained with $q^F = -0.25$. Calculations on the *Salmonella typhimurium* galactose binding protein yield extremely similar chemical shift spectra, consistent with the ~7% difference in amino acid composition. The uniform field components make the largest contributions to the shielding patterns observed, presumably because the gradient terms fall off more rapidly with distance. The exposed residue, Trp 284, has the largest amplitude of fluctuation associated with its shielding trajectory, possibly due to the rapid and random movement of neighboring water molecules. Trp 284 is highly shielded due to interaction with water, although other buried residues (e.g. Trp 133) may also be highly shielded, due to electric field effects within the protein. Our results imply that van der Waals interactions do not play a major role for fluorine shielding nonequivalence in proteins, since the experimental results can be reproduced by using solely the computed field and field gradient terms. The ability to predict protein NMR shielding patterns and ranges offers promise for structural analysis and also provides a way of validating different methods of computing protein electrostatics. In this respect, it is instructive to note that the charge fields from ionized surface groups are found to be largely shielded by the dielectric response of the solvent and the protein.

Introduction

Folding a protein (or nucleic acid) into its native conformation causes a large range of chemical shift nonequivalencies to be generated^{1–4} and permits the application of multidimensional nuclear magnetic resonance (NMR) techniques for structural analysis.^{5,6} Such folding-induced nonequivalencies have been known for about a quarter of a century, but they have never been fully explained. Recent empirical approaches to the analysis of ¹H chemical shifts^{7–9} appear to have considerable utility in refining

protein NMR structures, but for ¹³C NMR, structure–shift correlations are only just beginning to be recognized,¹⁰ and in the case of e.g. ¹⁵N or ¹⁹F, there has been very little progress in the analysis of experimental chemical shifts. This is a somewhat surprising situation given that the high Z element chemical shift ranges due to folding are extremely large compared to that of ¹H, up to ~10 ppm for ¹³C and ¹⁷O, ~30 ppm for ¹⁵N, and ~20 ppm for ¹⁹F, and should be amenable to calculation. However, full *ab initio* calculations on complete proteins are not yet possible. Thus, one needs to develop new approaches in order to simplify shielding calculations while retaining the “essential physics”, and one such approach is described in this paper.

Our ideas have developed from recent work on CO-labeled heme proteins, where we observed ¹³C and ¹⁷O chemical shifts for over a dozen proteins, and we were also able to determine ¹⁷O nuclear quadrupole coupling constants for the bound CO.¹¹ When we compared each of these three NMR parameters with the CO

[†] This work was supported by the United States Public Health Service (Grant HL-19481).

[‡] University of Illinois at Urbana–Champaign.

[§] National Institutes of Health Postdoctoral Fellow (Grant GM-14545).

^{||} University of Southern California.

[⊥] Present address: Department of Physiology, UCLA School of Medicine, Center for the Health Sciences, 10833 Le Conte Avenue, Los Angeles, CA 90024-1751.

(1) McDonald, C. C.; Phillips, W. D. *J. Am. Chem. Soc.* 1969, 91, 1513.

(2) Allerhand, A.; Childers, R. F.; Oldfield, E. *Biochemistry* 1973, 12, 1335.

(3) Glushka, J.; Lee, M.; Coffin, S.; Cowburn, D. *J. Am. Chem. Soc.* 1989, 111, 7716.

(4) Luck, L. A.; Falke, J. J. *Biochemistry* 1991, 30, 4248.

(5) Bax, A.; Ikura, M.; Kay, L. E.; Torchia, D. A.; Tschudin, R. *J. Magn. Reson.* 1990, 86, 304.

(6) Clore, G. M.; Gronenborn, A. M. *Prog. Nucl. Magn. Reson. Spectrosc.* 1991, 23, 43–92.

(7) Williamson, M. P.; Asakura, T.; Nakamura, E.; Demura, M. *J. Biomol. NMR* 1992, 2, 83. Williamson, M. P.; Asakura, T. *J. Magn. Reson.* 1991, 94, 557.

(8) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* 1991, 113, 9436.

(9) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* 1992, 31, 1647.

(10) Spera, S.; Bax, A. *J. Am. Chem. Soc.* 1991, 113, 5490.

(11) Park, K. D.; Guo, K.; Adebodun, F.; Chiu, M. L.; Sligar, S. G.; Oldfield, E. *Biochemistry*, 1991, 30, 2333–2347.

infrared vibrational frequency, ν_{CO} , we found excellent correlations between $\delta_i(^{13}\text{C})$, $\delta_i(^{17}\text{O})$, $e^2qQ/h(^{17}\text{O})$, and ν_{CO} , and we explained our results in terms of a weak electrical interaction model¹² in which changes in the vibrational frequency of CO, the ^{13}C and ^{17}O chemical shifts, as well as the ^{17}O nuclear quadrupole coupling constants were all interpreted as due to changes in polarization in CO due to large electric fields from the protein.¹² This model is based on the demonstration^{13–15} that the primary electronic structure change upon weak interaction is electrical polarization. We then used an empirical version of this basic model to interpret over 130 CO vibrational frequencies in heme proteins as being due to weak electrical interaction,¹⁶ and this led to a molecular model of the four major "conformational substates" seen in heme proteins as due to the electrical influence of the two ring-flip isomers of the H^{B1} and H^{B2} forms of the distal histidine residue, which, depending on orientation, can generate fields of up to $\approx 4 \times 10^7 \text{ V cm}^{-1}$ at CO. Fields of this magnitude are expected to cause significant changes in the shielding tensors for *all nuclei* in proteins, and in this paper, we focus on ^{19}F shifts, which because of the large polarizability of the C–F bond are expected to be highly susceptible to such weak or long-range electrical interactions.

^{19}F NMR studies of proteins began in the late 1960s and early 1970s, with use of ^{19}F probes such as CF_3CO and CH_2F (see e.g. refs 17 and 18 and references cited therein). Then, a number of groups used various bacterial and animal systems to produce proteins containing ^{19}F -labeled aromatic amino acids. For example, Kimber et al. labeled dihydrofolate reductase with [6- ^{19}F]Trp,^{19,20} Hull and Sykes investigated *m*-F Tyr-labeled alkaline phosphatase,^{21,22} Browne and Otvos studied [4- ^{19}F]Trp-labeled alkaline phosphatase,²³ Chaiken et al. investigated semisynthetic ribonuclease S labeled with [4-F]Phe²⁴ and [4-F]-His,²⁵ Ho et al. investigated [5-F]Trp J-protein from *Escherichia coli*,²⁶ as well as [4-, 5-, and 6-F]Trp-labeled lactate dehydrogenase,^{27–29} Gerig et al. investigated [4-F]Phe labeled hemoglobins and carbonic anhydrase,^{30–32} and most recently, Luck and Falke have produced an elegant series of papers on [3-F]-Phe- and [5-F]Trp-labeled D-galactose and D-glucose chemosensory receptors and the transmembrane aspartate transducer.^{33–35}

In about one-half of these studies, no *rigorous* specific assignments were possible, while in the others, there was no detailed analysis of the observed shifts in terms of molecular structure. Indeed, the only system with known assignments and known (solid-state) structure is the galactose binding protein (GBP),^{4,33} which we have thus chosen for our initial studies of chemical shielding.

Theoretical Background

The maximum range of ^{19}F NMR chemical shifts observed in proteins due to folding is about 17 ppm (found in a 1:1 [4- ^{19}F]Trp hen egg white lysozyme–chitotriose inhibitor complex, ref 36), an order of magnitude larger than the total shielding expected from ring current and CONH susceptibility anisotropy. While there are several possible empirical approaches which might be taken to analyzing these shifts, such as solvent exposure, there are three main (potentially) quantitative approaches to determining chemical shielding, as follows: (i) full *ab initio* methods; (ii) weak electrical interaction models, described either by using *ab initio* techniques or by a calculable series of shielding polarizability and hyperpolarizability tensors, multiplied by the fields and field gradients;^{37–41} and (iii) "van der Waals" interaction models,^{42,43} in which solely the $\langle E^2 \rangle$ or fluctuating London dispersion forces are considered.

We discuss elsewhere⁴⁴ the *ab initio* approach i using a gauge-including atomic orbital technique,^{45,46} but focus in this paper on ii above, the weak electrical interaction model. Since ii accounts for the experimental observations, we also briefly discuss below the likely applicability of iii, the van der Waals model.

The effect of an electric field on a molecular electronic property, P , such as chemical shielding, was first suggested for small molecules in fluids by Stephen³⁷ and Buckingham³⁸ as being expressible as a power series in the uniform field, F :

$$P_{\alpha\beta} = P_{\alpha\beta}^{(0)} + P_{\alpha\beta,\gamma}^{(1)}F_{\gamma} + P_{\alpha\beta,\gamma\delta}^{(2)}F_{\gamma}F_{\delta} + \dots \quad (1)$$

The intermolecular shielding observed at a particular site is here regarded as a result of the uniform electric field arising from its neighbors. In proteins, these fields would be far from uniform, hence eq 1 should include additional terms that take into account this nonuniformity, as suggested by Buckingham and Lawley (for the H atom, ref 47) and Batchelor:⁴⁸

$$P_{\alpha\beta} = P_{\alpha\beta}^{(0)} + P_{\alpha\beta,\gamma}^{(1)}F_{\gamma} + P_{\alpha\beta,\gamma\delta}^{(1)}F_{\gamma\delta} + P_{\alpha\beta,\gamma\delta\epsilon}^{(2)}F_{\gamma}F_{\delta}F_{\epsilon} + P_{\alpha\beta,\gamma\delta\epsilon\zeta}^{(2)}F_{\gamma\delta}F_{\epsilon\zeta} + \dots \quad (2)$$

where $F_{\gamma\delta}$ is an element of the field gradient tensor. If this expansion converges, then knowledge of the coefficients describing the response tensors listed in eq 2 and the field and the field gradient tensor at the nucleus in question allows the electrical contribution to shielding to be calculated. The response tensors in eq 2 can be obtained by calculating corrections to the energy of the system arising from several perturbations: the external

- (12) Augspurger, J. D.; Dykstra, C. E.; Oldfield, E. *J. Am. Chem. Soc.* **1991**, *113*, 2447.
- (13) Dykstra, C. E. *J. Am. Chem. Soc.* **1989**, *111*, 6168.
- (14) Dykstra, C. E. *J. Phys. Chem.* **1987**, *91*, 6216.
- (15) Gutowsky, H. S.; Germann, T. C.; Augspurger, J. D.; Dykstra, C. E. *J. Chem. Phys.* **1992**, *96*, 5808.
- (16) Oldfield, E.; Guo, K.; Augspurger, J. D.; Dykstra, C. E. *J. Am. Chem. Soc.* **1991**, *113*, 7537.
- (17) Gerig, J. T. *Methods Enzymol.* **1989**, *177*, 3.
- (18) Phillips, L.; Separovic, F.; Cornell, B. A.; Barden, J. A.; dos Remedios, C. G. *Eur. Biophys. J.* **1991**, *19*, 147.
- (19) Kimber, B. J.; Feeney, J.; Roberts, G. C. K.; Birdsall, B.; Griffiths, D. V.; Burgen, A. S. V.; Sykes, B. D. *Nature* **1978**, *271*, 184.
- (20) Kimber, B. J.; Griffiths, D. V.; Birdsall, B.; King, R. W.; Scudder, P.; Feeney, J.; Roberts, G. C. K.; Burgen, A. S. V. *Biochemistry*, **1977**, *16*, 3492.
- (21) Hull, W. E.; Sykes, B. D. *Biochemistry*, **1976**, *15*, 1535.
- (22) Hull, W. E.; Sykes, B. D. *J. Mol. Biol.* **1975**, *98*, 121.
- (23) Browne, D. T.; Otvos, J. D. *Biochem. Biophys. Res. Commun.* **1976**, *68*, 907.
- (24) Chaiken, I. M.; Freedman, M. H.; Lyerla, J. R., Jr.; Cohen, J. S. *J. Biol. Chem.* **1973**, *248*, 884.
- (25) Taylor, H. C.; Richardson, D. C.; Richardson, J. S.; Wlodawer, A.; Komoriya, A.; Chaiken, I. M. *J. Mol. Biol.* **1981**, *149*, 313.
- (26) Robertson, D. E.; Kroon, P. A.; Ho, C. *Biochemistry*, **1977**, *16*, 1443.
- (27) Rule, G. S.; Pratt, E. A.; Simplaceanu, V.; Ho, C. *Biochemistry* **1987**, *26*, 549.
- (28) Peersen, O. B.; Pratt, E. A.; Truong, H.-T.; Ho, C.; Rule, G. S. *Biochemistry*, **1990**, *29*, 3256.
- (29) Ho, C.; Pratt, E. A.; Rule, G. S. *Biochim. Biophys. Acta* **1989**, *988*, 173.
- (30) Gamcsik, M. P.; Gerig, J. T.; Gregory, D. H. *Biochim. Biophys. Acta* **1987**, *912*, 303.
- (31) Gerig, J. T.; Klinkenborg, J. C.; Nieman, R. A. *Biochemistry* **1983**, *22*, 2076.
- (32) Gamcsik, M. P.; Gerig, J. T.; Swenson, R. B. *Biochim. Biophys. Acta* **1986**, *874*, 372.
- (33) Luck, L. A.; Falke, J. J. *Biochemistry* **1991**, *30*, 4257.

- (34) Luck, L. A.; Falke, J. J. *Biochemistry* **1991**, *30*, 6484.
- (35) Falke, J. J.; Luck, L. A.; Scherrer, J. *Biophys. J.* **1992**, *62*, 1.
- (36) Oldfield, E.; Montez, B.; Patterson, J.; Harrell, S.; Lian, S.; Le, H.; Fang, B. Unpublished results.
- (37) Stephen, M. *J. Mol. Phys.* **1957**, *1*, 223.
- (38) Buckingham, A. D. *Can. J. Chem.* **1960**, *38*, 300.
- (39) Day, B.; Buckingham, A. D. *Mol. Phys.* **1976**, *32*, 343. Žaucer, M.; Ažman, A. Z. *Naturforsch.* **1979**, *34a*, 1279. Volodicheva, M. I.; Rebane, T. K. *Teor. Eksp. Khim.* **1984**, *19*, 357.
- (40) Raynes, W. T.; Ratcliffe, R. *Mol. Phys.* **1979**, *37*, 571.
- (41) Augspurger, J.; Pearson, J.; Oldfield, E.; Dykstra, C. E.; Park, K. D.; Schwartz, D. J. *Magn. Reson.* **1992**, *100*, 342.
- (42) Bothner-By, A. A. *J. Mol. Spectrosc.* **1960**, *5*, 52.
- (43) Raynes, W. T.; Buckingham, A. D.; Bernstein, H. J. *J. Chem. Phys.* **1962**, *36*, 3481.
- (44) de Dios, A.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491.
- (45) Ditchfield, R. *Mol. Phys.* **1974**, *8*, 397.
- (46) Hinton, J. F.; Guthrie, P.; Pulay, P.; Wolinski, K. *J. Am. Chem. Soc.* **1992**, *114*, 1604.
- (47) Buckingham, A. D.; Lawley, K. P. *Mol. Phys.* **1960**, *3*, 219.
- (48) Batchelor, J. G. *J. Am. Chem. Soc.* **1975**, *97*, 3410.

magnetic field, the external electric field, and the nuclear magnetic moment, and over the past 30 years there has been considerable work aimed at determining these coefficients.³⁸⁻⁴¹ Most recently, Dykstra and colleagues have developed an open-ended way of differentiating electronic wave functions and energies, derivative Hartree-Fock (DHF) theory,⁴⁹ and have reported a series of so-called (multipole) shielding polarizability tensors.^{12,41,50}

The first step in implementing the weak electrical interaction model is to determine the shielding polarizabilities,^{12,50} that is, the responses of the shielding tensor elements, σ_{ii} , to external electric fields, E_γ , and field gradients, $V_{\gamma\delta}$, which yield the following:

$$A_{\alpha\beta,\gamma} = \frac{\partial \sigma_{\alpha\beta}}{\partial E_\gamma}, \quad \text{the dipole shielding polarizability,} \\ \text{a third-rank tensor}$$

$$A_{\alpha\beta,\gamma\delta} = \frac{\partial \sigma_{\alpha\beta}}{\partial V_{\gamma\delta}}, \quad \text{the quadrupole shielding polarizability,} \\ \text{a fourth-rank tensor}$$

etc. A series of analogous hyperpolarizabilities can be determined as

$$B_{\alpha\beta,\gamma,\gamma} = \frac{\partial \sigma_{\alpha\beta}}{\partial E_\gamma^2}, \quad \text{the dipole shielding hyperpolarizability,} \\ \text{a fourth-rank tensor}$$

the above term being related to the Buckingham "B term",³⁸ which we believe makes a relatively small contribution to shielding.^{48,51}

In general, we believe that it is instructive to break down the total shielding (σ_t) in a protein (or other macromolecular system) into three component parts:

$$\sigma_t = \sigma_s + \sigma_l + \sigma_o \quad (3)$$

σ_s is the short-range or *electronic* structural contribution to shielding and would play a major role for e.g. C α or ¹⁵NH and would include ϕ, ψ, χ torsion angle effects, strong hydrogen bonding, etc. σ_l is the contribution from the long-range or relatively weak electrical interaction with the surrounding protein and solvent molecules, which we believe is dominant for ¹⁹F. σ_o contains the other, classical interactions, such as ring current effects or other long-range magnetic anisotropies.

Of course, the breakdown shown in eq 3 is artificial, but we find it convenient since it has permitted us to compute many ¹³C and ¹⁵N shieldings in proteins, where σ_s dominates, followed by σ_l .⁴⁴ The exact partitioning of σ_t into our three component parts is somewhat dependent upon the nucleus of interest. For example, peptide hydrogen bonding could be included in σ_l in some cases, if suitable representations are available. At present, however, we use eq 3 in the following operational manner: σ_s contains what in our laboratory has to be calculated using *ab initio* methods; σ_l can be computed either using charge-field perturbation at the SCF level or using the methods we outline below; σ_o are the other interactions, which for the heavier elements are relatively small and can be neglected.

In this paper, we are interested primarily in determining σ_l . More specifically, we will compute the shielding changes due to electric field effects:

$$P_{\alpha\beta} - P_{\alpha\beta}^{(0)} = P_{\alpha\beta,\gamma}^{(1)} F_\gamma + P_{\alpha\beta,\gamma\delta}^{(1)} F_{\gamma\delta} \quad (4)$$

where $P_{\alpha\beta} \equiv \sigma_{\alpha\beta}$, the $\alpha\beta$ element of the ¹⁹F chemical shielding tensor, $P_{\alpha\beta,\gamma}^{(1)} \equiv A_{\alpha\beta,\gamma}$, and $P_{\alpha\beta,\gamma\delta}^{(1)} \equiv A_{\alpha\beta,\gamma\delta}$. The values of the coefficients comprising the A tensor are discussed elsewhere.⁴¹

(49) Dykstra, C. E.; Jasien, P. G. *Chem. Phys. Lett.* **1984**, *109*, 388.

(50) Augspurger, J. D.; Dykstra, C. E. *J. Phys. Chem.* **1991**, *95*, 9230.

(51) de Dios, A.; Oldfield, E.; Augspurger, J.; Dykstra, C. Unpublished results.

Fortunately, symmetry and other considerations indicate that only a limited number of tensor elements need to be incorporated into a shielding model for aromatic C-F bonds, and they are as follows:

$$\bar{A}_x = 1/3(A_{xx,x} + A_{yy,x} + A_{zz,x}) \quad (5)$$

$$\bar{A}_{xx} = 1/3(A_{xx,xx} + A_{yy,xx} + A_{zz,xx}) \quad (6)$$

$$\bar{A}_{yy} = 1/3(A_{xx,yy} + A_{yy,yy} + A_{zz,yy}) \quad (7)$$

$$\bar{A}_{zz} = 1/3(A_{xx,zz} + A_{yy,zz} + A_{zz,zz}) \quad (8)$$

where the x axis lies along the C-F bond, the y axis is coplanar with the aromatic ring, and the z axis is normal to the plane of the aromatic ring.

The next step is to determine the values of the fields (E_x) and field gradients (V_{xx} , V_{yy} , V_{zz}). The large number of conformational substates experienced by a protein suggests the desirability of incorporating molecular dynamics into the calculation, in order to obtain a more realistic estimate of average values of E_x and V_{ii} 's. We used a modified version of the ENZYMIK program⁵² to obtain the MD trajectories of the field and its gradient terms at the sites of interest. The program uses spherical boundary conditions that follow the surface-constrained all-atom solvent (SCAAS) approach.⁵³ This is accomplished by dividing the solvated protein system into five regions, as depicted in Figure 1 (see ref 53 for detailed accounts of the meaning of these regions

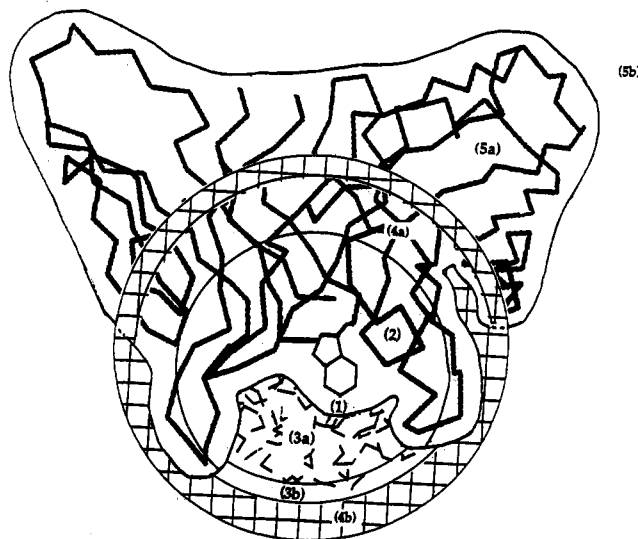


Figure 1. An illustration of the five regions of the protein system in the surface constrained all-atom model which is used in the MD calculations. Region 1 contains the fluorotryptophan of interest. Region 2 contains the unconstrained protein groups within a radius R_2 from the center. Region 3a contains the unconstrained water molecules, and region 3b contains the surface water molecules (which are subjected to polarization constraints). Region 4a contains the harmonically constrained protein atoms between the radii R_2 and R_{4a} , and region 4b contains the grid of Langevin dipoles that complete regions 1, 2, 3, and 4a to a sphere of a radius R_{4b} . Region 5 (which includes region 5a of the protein and region 5b) is modeled by considering the entire region as a dielectric continuum.

and the SCAAS approach). The SCAAS approach focuses on obtaining a reliable treatment of the long-range electrostatic forces, which are evaluated in the current version of ENZYMIK

(52) Warshel, A.; Creighton, S. *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989; pp 120-138.

(53) Warshel, A.; Sussman, A. F.; King, G. *Biochemistry*, **1986**, *25*, 8368. King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647. Lee, F. S.; Chu, Z. T.; Bolger, M.; Warshel, A. *Protein Eng.* **1992**, *5*, 215. Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.*, in press.

by the local reaction field (LRF) method.⁵⁴ The LRF method involves the partition of the system into electroneutral groups of atoms, such as water molecules, methyl groups of alanine side chains, carboxyl groups of acidic side chains, etc. The electrostatic potential at the α th charge of the i th group is divided into short- and long-range components. The short-range potential, Φ_s , is simply the sum of the electrostatic contributions from the groups inside a cutoff radius, R_{cut} . The long-range potential, Φ_l , is approximated in the LRF method by the first four terms in a Taylor series about the center of each group.⁵⁴ The long-range electrostatic energy of the i th group is then given by

$$U_i^l = Q_i(\Phi_l)_{R_i} - \mu_i(\xi_l)_{R_i} - \frac{1}{3} \Theta_i(F_l)_{R_i} - \frac{1}{15} \sum_{i'''} \Omega_{i'''}(F_{l'''}^{\prime\prime\prime})_{R_i} \quad (9)$$

where Q_i , μ_i , Θ_i and Ω_i are, respectively, the monopole, dipole, quadrupole, and octopole moments about the center of the i th group (see refs 55 and 56 for the expressions of these electric multipole tensors), while $(\xi_l)_{R_i}$ and $(F_l)_{R_i}$ are, respectively, the long-range contributions to the electric field and field gradient at R_i , and $(F_{l'''}^{\prime\prime\prime})_{R_i}$ are the tensor components corresponding to the long-range contribution to the gradient of the field gradient at R_i .

The total energy of the system is given by

$$U_{\text{total}} = \frac{1}{2} \sum_i \sum_{\alpha} q_i^{\alpha} [\Phi_s(r_i^{\alpha}) + \Phi_l(r_i^{\alpha})] + U_{\text{ind}} + U_{\text{vdW}} + U_{\text{bonding}} \quad (10)$$

where U_{bonding} is the bonding interaction between the different fragments of the system described by a standard force field,⁵⁷ U_{vdW} is the nonelectrostatic van der Waals interaction evaluated within the given R_{cut} , and U_{ind} is the contribution from the induced dipoles of the system, which can be treated as a part of the electrostatic potential (see ref 54 for more details). In the current model, water atoms as well as atoms in region 1 (the site of interest) are not polarizable. For ground-state properties, the use of polarizable water gives results similar to those obtained with a larger permanent dipole but no induced dipole.

Computational Procedures

All computations were performed on an IBM RISC/6000 cluster (International Business Machines Corporation, Austin, TX). Except as noted below, default values of the polarizable MOLARIS force field in ENZYME for all the nonbonded and "bonded" (i.e., bond-stretching, angle-bending, and torsional) interactions were used. [5-F]Trp was obtained by "mutating" the corresponding hydrogen of each Trp residue into a fluorine atom and relaxing the subsequent structure using low-temperature MD (5 K for 2 ps). Since F van der Waals (vdW) parameters for F-Trp are unknown, we used typical O values, and q^F was typically -0.25. In general, there is a very strong correlation between the partial charges and vdW parameters used. Here, we try to find the best F partial charge with a given set of vdW parameters, rather than trying to optimize both parameters independently.

For computational ease, the field and field gradient at the five sites were evaluated separately. For each site, the corresponding protein system was composed of several regions, as depicted in Figure 1. Region 1 contains the atom of interest. Region 2 contains the unconstrained protein groups found within a radius

R_2 from the center. Region 3 contains the explicit water molecules found within a radius R_3 from the center such that region 3a contains the unconstrained water molecules and region 3b contains the surface water molecules (which are subjected to polarization constraints). Region 4a contains the harmonically constrained protein atoms between the radii R_2 and R_{4a} , and region 4b contains the grid of Langevin dipoles that complete regions 1, 2, 3, and 4a to a sphere of a radius R_{4b} . Region 5 (which includes region 5a of the protein) is modeled by considering the entire region as a dielectric continuum. In the current model, no counterions are included in the solvent because the effects of ionic strength are expected to be very small, due to the high dielectric constant of water.⁵³

In the present study, region 1 contains the F atom of interest, whose initial position was designated to be the center of the system. The radii R_2 , R_3 , R_{4a} , and R_{4b} were all 16 Å so that no Langevin dipoles were included, and the system was reduced to a 16-Å sphere of explicit atoms. The field of region 5 was not included in the simulations. For our initial explorations, such cutoffs seemed reasonable. The F atom in region 1 interacts with all atoms in regions 2 and 3, which contribute to the field and field gradient at fluorine. All nonbonded interactions with fluorine (which include the 1-4 pairs) were evaluated *explicitly* (i.e., no expansion was used). Nonbonded interactions between atoms in regions 2 and 3 were subject to the cutoff radius, R_{cut} . A R_{cut} of 7 Å was used in all cases. The Coulombic interactions beyond R_{cut} were approximated by the LRF expansion, while the van der Waals and permanent dipole-induced dipole interactions beyond R_{cut} (which are expected to be very small) were neglected in the simulations. The reliability of the LRF method in reproducing Coulomb effects beyond $R_{\text{cut}} = 7$ Å has been demonstrated previously in evaluations of hydration energies of ions and pK_a 's of acidic groups in a solvated protein.⁵⁴ We reemphasize here that this cutoff R_{cut} is different from those used to define the different regions of the system and is used only for nonbonded interactions between *all* groups in region 2 + 3, which could be separated by as much as 32 Å in the explicit 16-Å sphere. Nonbonded interactions between region 1 and region 2 + 3 are evaluated without the use of R_{cut} .

All MD trajectories were generated at 300 K with time steps of 2 fs. The long-range potential and its derivatives (see eq 9) at each group in regions 2 and 3 were updated once in every 40 time steps. The field and its gradient at the fluorine atom (which should not be confused with the long-range field and field gradient terms of eq 9 for the long-range potential) were evaluated at each step and stored once every 10 time steps. Typically, 11 000 time steps were calculated for each trajectory, with the first 2 ps discarded, yielding a 20-ps trajectory.

Partial atomic charges for 5-fluorotryptophan, the buried Ca^{2+} /Glu-Glu-Asp cluster, and glucose were obtained from a Mulliken population analysis obtained by using the Gaussian-88 program.⁵⁸ The Ca^{2+} -cluster geometry used in the Gaussian-88 calculation of the charge field was obtained from the Brookhaven Protein Data Bank,⁵⁹ although this cluster geometry was allowed to vary during MD and could have a small effect on the observed shieldings. 3-21G* basis sets were used for the partial charge calculations.

(58) *Gaussian 88*; Frisch, M. J.; Head-Gordon, M.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1988.

(59) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures. *J. Mol. Biol.* 1977, 112, 535. Abola, E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. Protein Data Bank. In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; pp 107.

(54) Lee, F. S.; Warshel, A. *J. Chem. Phys.* 1992, 97, 3100.

(55) Buckingham, A. D. *Q. Rev. Chem. Soc.* 1959, 13, 189.

(56) Buckingham, A. D. *Perspectives in Quantum Chemistry and Biochemistry*; B. Pullman Ed.; Wiley: New York, 1978; Vol. II, pp 1-67.

(57) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991.

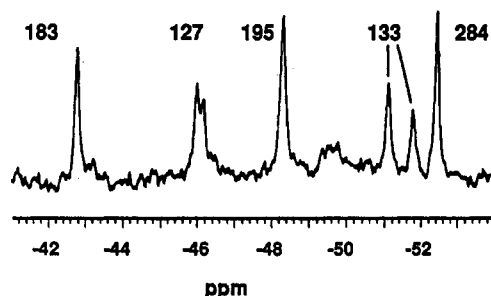


Figure 2. ^{19}F solution NMR spectrum (11.7 T; 470 MHz) of *E. coli* galactose binding protein in the presence of Ca^{2+} and glucose. Reprinted with permission from ref 4. Copyright 1991 American Chemical Society.

Results and Discussion

We show in Figure 2 the 470-MHz ^{19}F solution state (100 mM KCl, 10 mM Tris, pH 7.1, 10% D_2O , 0.5 mM CaCl_2 , 25 $^\circ\text{C}$) NMR spectrum of the *E. coli* galactose binding protein (GBP), labeled to 65% with $[5(\eta_3)\text{-F}]\text{Trp}$, obtained by Luck and Falke.⁴ GBP has five Trp residues, at positions 127, 133, 183, 195, and 284,⁶⁰ and each resonance in the spectrum shown in Figure 2 has been specifically assigned by using proteins prepared *via* site-directed mutagenesis.⁴ The resonances of Trp 127 and 133 are split into two peaks, possibly due to the presence of a nearby proline at position 231 in the amino acid sequence, in van der Waals contact with Trp 133. The crystal structures of GBP from *E. coli*⁶¹ and *S. typhimurium*⁶² have both been determined by means of X-ray crystallography, and examination of both structures (Brookhaven Protein Data Bank Files 2GBP and 3GBP, ref 59) shows the Trp conformations to be extremely similar, even though there are 21 amino acid modifications between the two proteins. GBP binds one Ca^{2+} and one galactose (or glucose) per molecule, and both cofactors have marked effects on the ^{19}F NMR spectrum.⁴ Trp 127 is close to Ca^{2+} , while Trp 183 is close to the galactose/glucose binding site. Since both X-ray structures have used the Ca^{2+} + glucose bound form of the protein, we use the spectrum of the fully ligated protein, Figure 2, as an initial test case for computation of ^{19}F NMR chemical shifts using the shielding polarizability tensor approach.

As described above, we require accurate estimates of the shielding polarizability tensor elements, and a good description of the dynamically averaged electrostatic potential surface, in order to compute the weak electrical interaction contributions to shielding in GBP. In a previous paper,⁴¹ Augspurger et al. reported that the dipole shielding polarizability, A_x , for fluorobenzene was 1885 ppm/au field, computed analytically by derivative Hartree-Fock theory using a TZ/P basis set,⁴¹ while that of 1,4-difluorobenzene was 1955 ppm/au and that for 1-fluoro-4-lithiobenzene was 1985 ppm/au. Thus, it is reasonable to suppose that a value of ~ 1900 ppm/au should be applicable to the F site in $[5\text{-F}]\text{Trp}$. This dipole shielding polarizability value is much larger than that considered previously,^{63,64} but within a factor of about 3 of that deduced from experiments on β,β -difluorostyrene,⁶⁵ in which the F sites are not aromatic. Results for $A_{\alpha\beta,\gamma\delta}$ are reported elsewhere,^{41,51} but for convenience are reproduced in full in Table I. The dominant term, as might be expected, is $A_{xx} = -4860$ ppm/au EFG. As previously,⁴¹ we use the following coordinate system (which relates to the typical experimental ^{19}F shift tensor element nomenclature of e.g. Mehring, ref 66).

Table I. Multipole Shielding Polarizability Tensor Elements Used in Calculation of ^{19}F Shielding in the *E. coli* Galactose Binding Protein^a

| element | value (ppm/au) | element | value (ppm/au) |
|-------------|----------------|-------------|----------------|
| $A_{xx,x}$ | 1483.2 | $A_{yy,zz}$ | -652.9 |
| $A_{yy,x}$ | 3576.5 | $A_{zz,xx}$ | -1232.1 |
| $A_{zz,x}$ | 593.6 | $A_{zz,yy}$ | -796.6 |
| $A_{xx,xx}$ | -4151.1 | $A_{zz,zz}$ | -283.8 |
| $A_{xx,yy}$ | -716.6 | A_x | 1884.4 |
| $A_{xx,zz}$ | -488.4 | A_{xx} | -4859.8 |
| $A_{yy,xx}$ | -9196.1 | A_{yy} | -1410.2 |
| $A_{yy,yy}$ | -2717.4 | A_{zz} | -475.0 |

^a From refs 41 and 51.

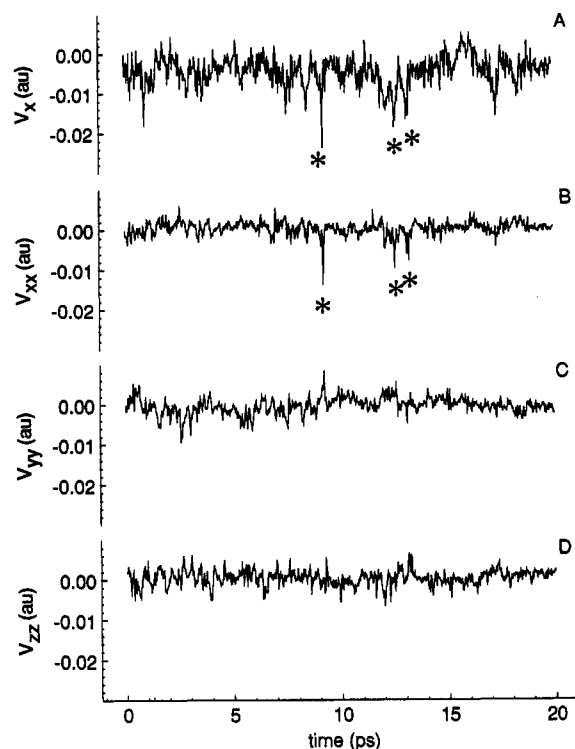
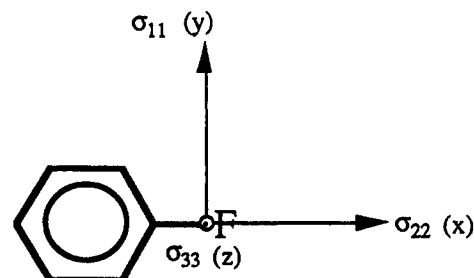


Figure 3. Field and field gradient trajectories for $[5\text{-}^{19}\text{F}]\text{Trp}$ 284 in GBP: A, $V_x f(\tau)$; B, $V_{xx} f(\tau)$; C, $V_{yy} f(\tau)$; D, $V_{zz} f(\tau)$.



Since there are no "ideal" simple experimental molecular systems which can be used to test the accuracy of the shielding polarizability coefficients (which may have errors on the order of 10–40%, depending on the order of derivative considered), we have examined elsewhere the agreement between a series of calculations of ^{19}F shielding in various model fluorobenzene- $(\text{HF})_n$ clusters using full *ab initio* techniques and those obtained by using the dipole and quadrupole shielding polarizabilities. There is generally good agreement between the two approaches,⁵¹ especially at distances ≥ 5 Å. Thus, at least in model calculations, the DHF derived shielding polarizability coefficients enable extremely rapid estimates of the effects of the potential surface on ^{19}F shielding in a fluoroaromatic species,⁵¹ as determined by comparisons with full *ab initio* calculations.

We now return to the question of determining the fields and

- (60) Benner-Luger, D.; Boos, W. *MGG, Mol. Gen. Genet.* **1988**, *214*, 579.
 (61) Vyas, N. K.; Vyas, M. N.; Quijcho, F. A. *Science* **1988**, *242*, 1290.
 (62) Mowbray, S. L.; Smith, R. D.; Cole, L. B. *Receptor* **1990**, *1*, 41.
 (63) D. H. Gregory. Ph. D. Thesis, University of California, Santa Barbara, University Microfilms International, Ann Arbor, MI, 9034327, 1990.
 (64) Gregory, D. H.; Gerig, J. T. *Biopolymers* **1991**, *31*, 845.
 (65) Reynolds, W. F.; Gibb, V. G.; Plavac, N. *Can. J. Chem.* **1980**, *58*, 839.
 (66) Mehring, M. In *NMR Basic Principles and Progress*; Diehl, P., Fluck, E., Kosfeld, R., Eds.; Springer-Verlag: Berlin, 1976; p 181.

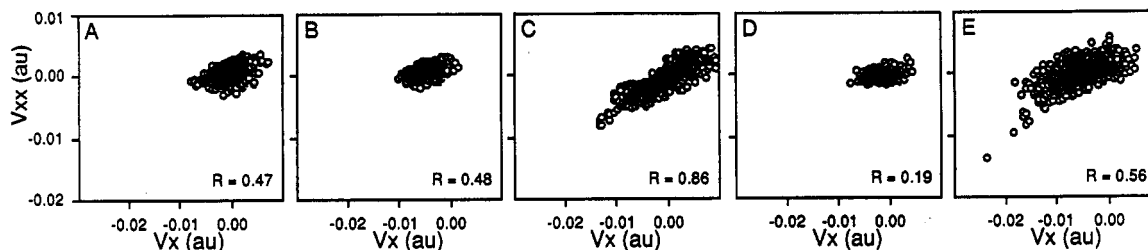


Figure 4. Plots of uniform field along the C–F bond axis *versus* field gradient component along the C–F bond axis for various [5-¹⁹F]Trp residues in GBP: A, Trp 127; B, Trp 133; C, Trp 183; D, Trp 195; and E, Trp 284. The uncorrelated results are more likely to give accurate shieldings since no single interaction dominates the potential.

field gradients in GBP. We use the LRF approach as implemented in the ENZYMIK suite of programs, and we show in Figure 3A the component of the electric field of the protein (and water) projected along the C–F bond axis during a single 20-ps 300 K LRF–MD trajectory, for Trp 284. Also shown are the time-dependent components of the electric field gradient tensor (V_{xx} , V_{yy} , V_{zz} , Figure 3B–D), using the convention that V_{xx} is the component along the C–F bond axis, V_{yy} is the in-plane component, and V_{zz} is perpendicular to the indole plane, as shown above. Weak harmonic constraints of 0.1 kcal mol^{−1} Å^{−2} were applied to all protein heavy atoms in the MD trajectory.

As shown by the asterisk in Figure 3, there are a number of correlations between E_x and V_{xx} , as a function of time. Both Trp 183 and Trp 284 show this type of behavior, as illustrated in Figure 4, while for Trp 127, Trp 133, and Trp 195 it is much less apparent. We believe these correlations, although weak, are due to the dominance of one or two specific short-range interactions, which tend to dominate the field/field gradient values at the sites in question. For Trp 183, it is likely that the adjacent Arg 158 is a major contributor to the electrostatic field experienced, while for Trp 284 it is likely that nearby oriented H₂O molecules are important. For Trp 127, Trp 133, and Trp 195, there is essentially no correlation between E_x and V_{xx} indicating that a much larger number of interactions contribute to the potential at these sites, any specific interactions tending to be “blurred” out.

Once we have established the values for $\langle E_x \rangle$, $\langle V_{xx} \rangle$, $\langle V_{yy} \rangle$, and $\langle V_{zz} \rangle$, we use the coefficients \bar{A}_x , \bar{A}_{xx} , \bar{A}_{yy} , and \bar{A}_{zz} listed in Table I, together with the following equations,

$$\sigma(E_x) = \bar{A}_x \langle E_x \rangle \quad (11)$$

$$\sigma(V_{xx}, V_{yy}, V_{zz}) = \bar{A}_{xx} \langle V_{xx} \rangle + \bar{A}_{yy} \langle V_{yy} \rangle + \bar{A}_{zz} \langle V_{zz} \rangle \quad (12)$$

to obtain the uniform field shifts, $\sigma(E_x)$, and the gradient shifts, $\sigma(V_{ii})$. The total shielding due to weak electrical interaction is thus given by

$$\Delta\sigma(\Phi) = \bar{A}_x \langle E_x \rangle + \bar{A}_{xx} \langle V_{xx} \rangle + \bar{A}_{yy} \langle V_{yy} \rangle + \bar{A}_{zz} \langle V_{zz} \rangle \quad (13)$$

where the $\langle \rangle$ indicate values obtained from one or more MD trajectories (see below). As may be deduced from Figures 3 and 4, there is a clear anticorrelation between the uniform field shift and the field gradient shift for Trp 183 and Trp 284, as shown for Trp 284 in Figure 5. This is as expected on the basis of previous model calculations with the C₆H₅F–(HF)_n cluster system,⁵¹ where we found that the contributions to the overall chemical shielding from the uniform field and the field gradient terms tend to oppose each other.

Also of interest are the various contributions to the fluctuations seen in a given shielding trajectory, and we show in Figure 6 for Trp 195 the uniform field, individual field gradient, total field gradient, and total field plus field gradient contributions to shielding. These results clearly show the dominance of \bar{A}_x and \bar{A}_{xx} in governing shielding, $\sigma(V_{yy})$ and $\sigma(V_{zz})$ being extremely small, in large part because of the small values for \bar{A}_{yy} and \bar{A}_{zz} .

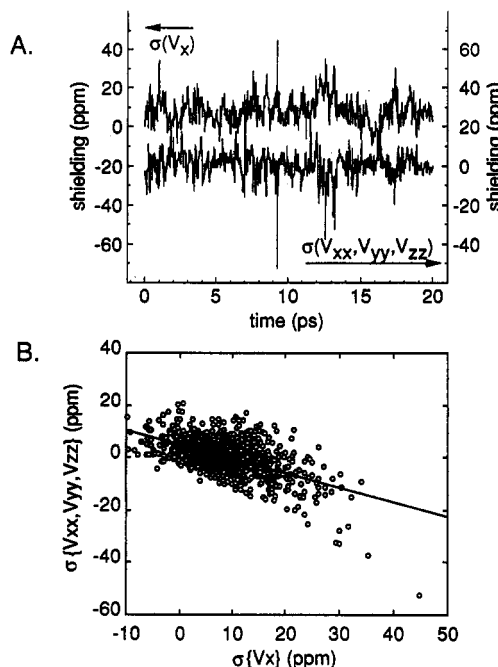


Figure 5. Field and field gradient contributions to ¹⁹F shielding of Trp 284: A, individual (V_x , V_{ii}) shielding trajectories; B, correlation between $\sigma(V_x)$ and $\sigma(V_{ii})$. Both show anticorrelation, as expected on the basis of work on model systems (refs 44 and 51).

We now consider the question of the appropriate charge to be used on F in the LRF–MD analysis, and we show in Figure 7 three separate shielding trajectories, $\sigma(E_x, V_{ii}) f(\tau)$, where τ is the time (in picoseconds) along the MD trajectory, for Trp 183 in GBP. The charges chosen were as follows: $q = -0.125$, -0.25 ,⁶⁷ and -0.375 . For the purposes of this test, both heavy atoms and hydrogens were constrained using 0.1 kcal/mol/Å². There are clearly differences in the magnitudes of the shifts observed, as might be anticipated on the basis of the different electrostatic interactions with protein and water which might be expected due to the large range in charge values chosen (the attached carbons have the same magnitude charge, but opposite sign). The choice of fluorine charge we have used for future GBP calculations is $q = -0.25$, since this (i) is very close to the value $q = -0.231$ used by Gregory and Gerig in their analysis of *p*-fluorophenylalanine, (ii) is that deduced from a 6–31G** calculation of [5-F]Trp ($q = -0.37$) when “scaled” down by a factor of ~ 0.7 to match the rest of the charges used in the ENZYMIK force field, and (iii) gives the best agreement between theory and experiment. To indicate the importance of q^F , we show in Figure 8 the results of single shielding trajectories for GBP, each 20 ps long and using the relatively large constraint discussed above, in terms of a stick diagram. As can be seen from Figure 8, there are significant differences. However, for $q^F = -0.25$, there is an overall generally good agreement between theory and experiment, and this value is that already independently deduced (for [4-F]Phe) by Gregory and Gerig.⁶⁷

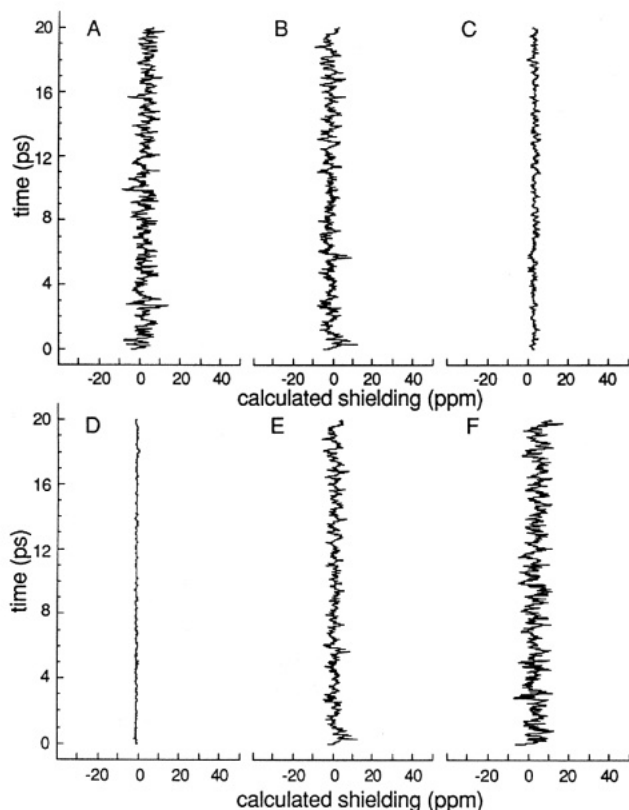


Figure 6. Shielding trajectories (20 ps; 1000 time step) for ^{19}F NMR of [5- ^{19}F]Trp 195 in GBP: A, $\sigma(E_x)$; B, $\sigma(V_{xx})$; C, $\sigma(V_{yy})$; D, $\sigma(V_{zz})$; E, $\sigma(V_{xx}, V_{yy}, V_{zz})$; F, $\sigma(E_x, V_{xx}, V_{yy}, V_{zz})$. Chemical shifts are in parts per million from an unperturbed reference, at 0 ppm.

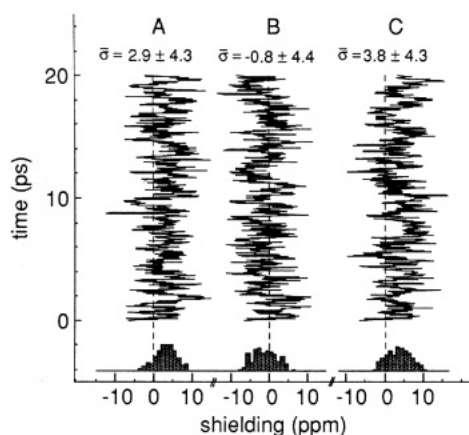


Figure 7. Shielding trajectories for [5- ^{19}F]Trp 183 using various q^F values: A, $q^F = -0.125$; B, $q^F = -0.25$; C, $q^F = -0.375$.

At this point, we should note that a number of other charge assignments need to be considered in the LRF-MD calculations. First, we have used the nonionized forms of Asp, Glu, His, Lys, and Arg residues. The basis for this is that, in separate experiments on [4- ^{19}F]Trp lysozyme, we find only rather minor spectral changes on removing surface charge, e.g. by pH titration,⁶⁸ or acetylation of lysine side chains ($\text{NH}_3^+ \rightarrow \text{NHCOCH}_3$, ref 69), and in addition, the results of much previous work^{53,70} indicate that the surface charge field will be essentially screened inside the protein by the presence of water (which has a dielectric constant of ~ 80). Second, GBP is a Ca^{2+} and glucose binding protein. We thus computed charges for the buried Ca^{2+} and its associated glutamate/aspartate counterions, together with the C and O

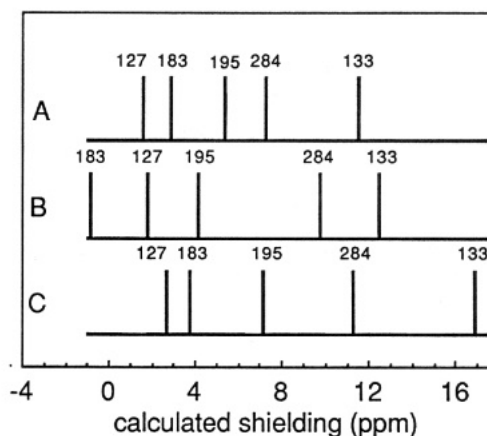


Figure 8. Theoretical shielding plots for GBP using A, $q^F = -0.125$; B, $q^F = -0.25$; and C, $q^F = -0.375$. The zero shielding reference represents no electrical influence.

Table II. Mulliken Population Charges Used for Calcium Binding Site and Glucose in GBP

| atom | charge (this work, au) | charge (ref 71, au) |
|----------------------------------|------------------------|---------------------|
| Ca^{2+} | 0.74 | |
| ^{13}C of Asp 134, 138 | 0.82 | |
| ^{13}C of Glu 205 | 0.82 | |
| δO of Asp 134, 138 | -0.70 | |
| αO of Glu 205 | -0.70 | |
| C1 | 0.194 | 0.184 |
| C2 | 0.049 | 0.066 |
| C3 | 0.053 | 0.066 |
| C4 | 0.065 | 0.055 |
| C5 | 0.062 | 0.066 |
| C6 | 0.007 | 0.008 |
| O1 | -0.315 | -0.321 |
| O2 | -0.299 | -0.310 |
| O3 | -0.305 | -0.317 |
| O4 | -0.310 | -0.312 |
| O5 | -0.270 | -0.279 |
| O6 | -0.307 | -0.306 |
| HO1 | 0.207 | ~ 0.2 |
| HO2 | 0.179 | ~ 0.2 |
| HO3 | 0.175 | ~ 0.2 |
| HO4 | 0.195 | ~ 0.2 |
| HO6 | 0.190 | ~ 0.2 |
| HC1 | 0.055 | < 0.1 |
| HC2 | 0.062 | < 0.1 |
| HC3 | 0.062 | < 0.1 |
| HC4 | 0.058 | < 0.1 |
| HC5 | 0.060 | < 0.1 |
| HC61 | 0.066 | < 0.1 |
| HC62 | 0.062 | < 0.1 |

charges for glucose, using a Gaussian 88 program (3-21G* basis set, ref 58), and the Mulliken population charges were then incorporated into ENZYME. The glucose and the Ca^{2+} cluster charges used are given in Table II. For glucose, our results are quite close to those determined independently by others,⁷¹ as also shown in Table II. As might be anticipated, variations in Ca^{2+} /Glu/Asp or glucose charge cause shielding changes in the immediately adjacent residues, Trp 127 and Trp 183, and use of a 2+ charge on Ca gave significantly worse agreement with experiment (about a 2 ppm shielding increase for Trp 127). Third, we deleted the crystallographically resolved H_2O molecules, primarily because their H atoms cannot be accurately located. Incorporation of H_2O 313, depending on its orientation, can introduce an ~ 1 –2 ppm uncertainty in the shielding of Trp 183, due to the H_2O charge field.

Figure 9 shows typical 20-ps [5- ^{19}F]Trp NMR shielding trajectories for Trp 127, 133, 183, 195, and 284 in the *E. coli* galactose binding protein. As can be seen, there is good general

(67) Gregory, D. H.; Gerig, J. T. *J. Comput. Chem.* **1989**, *10*, 711.

(68) Lian, C.; Oldfield, E. Unpublished results.

(69) Le, H.; Oldfield, E. Unpublished results.

(70) Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267.

(71) Petukhov, M. G.; Pershin, V. L. *Z. Strukt. Khim.* **1988**, *29*, 167.

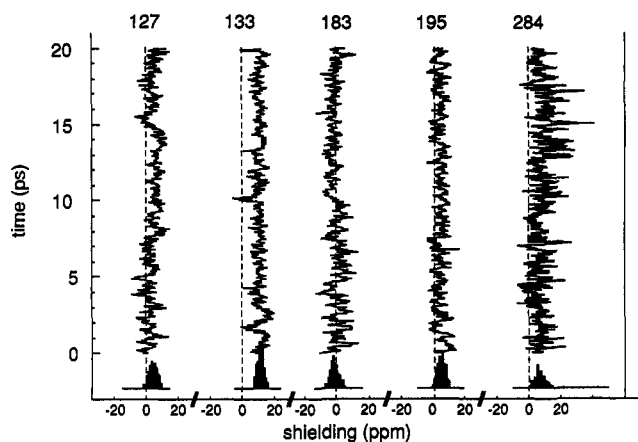


Figure 9. Individual 20-ps shielding trajectories for each of the five [5-F]-Trp residues in GBP.

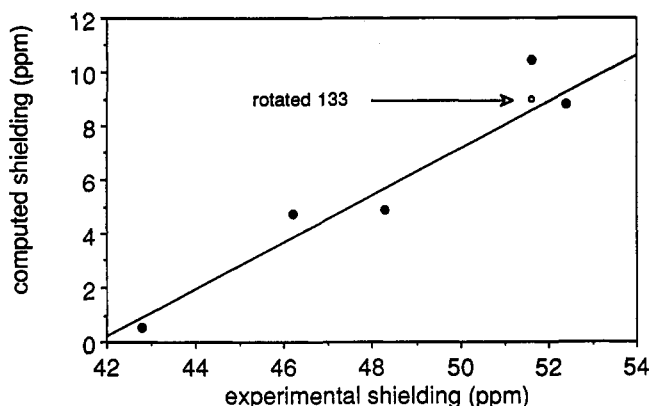


Figure 10. Plot of experimental versus theoretical shielding for each [5-F]Trp site in GBP. The experimental shieldings are those reported in ref 4. The computed values are referenced to a field-free value of 0 ppm. The open circle represents the shielding calculated for a 180° C β -C γ flip of Trp 133.

agreement in the overall spectral breadth and the ordering of chemical shifts, when compared with experiment, Figure 2, but there are naturally deviations of a few parts per million to be expected for each residue when comparing trajectories differing in initial momentum distribution. We thus generated additional trajectories in order to obtain more reliable average values by inserting differing numbers of steps at high temperature (500 K) at the beginning of the calculation. This produced dissimilar momentum distributions upon "cooling" to room temperature, allowing the calculations to sample different regions of phase space. In order to obtain our final shielding values, we averaged for each site the results of five 20-ps trajectories, as is typically done with restrained MD methods (see e.g. refs 54 and 72), including previous work aimed at determining pK_a values using the LRF model,⁵⁴ and we show in Figure 10 an experimental versus theoretical plot of the GBP shieldings, averaged over five such trajectories. We believe the agreement between theory and experiment is very promising, and when taken together with the results obtained on model systems,^{44,51} such good agreement supports the idea that weak electrical interactions dominate ^{19}F shielding nonequivalence in GBP.

While we cannot say that the approach we have outlined is an absolute proof of the correctness of the method, we can state that (i) the shielding polarizability coefficients we have used give, in model cluster calculations, chemical shifts in generally good accord with those obtained by using full *ab initio* calculations with large basis sets;⁵¹ (ii) the shielding tensor elements computed for a range of fluorobenzenes using these same programs⁴¹ are in

Table III. Shielding Trajectory and Solvent Accessibility Parameters for [5-F]Trp Residues in *E. coli* Galactose Binding Protein

| residue | $\bar{\sigma}$ (ppm) ^a | $s(\sigma)$ (ppm) ^b | A (Å ²) ^c ($r = 1.4$ Å) |
|---------|-----------------------------------|--------------------------------|---|
| 127 | 4.7 | 1.4 | 0.0 |
| 133 | 10.4 | 3.3 | 0.0 |
| 183 | 0.5 | 3.1 | 0.0 |
| 195 | 4.9 | 0.7 | 0.0 |
| 284 | 8.8 | 1.1 | 1.8 |

^a Mean shielding due to electric field effects in parts per million from an unperturbed [5-F]Trp side chain; average value from five 20-ps trajectories. ^b Standard deviation of $\bar{\sigma}$. ^c ACCESS (ref 73) solvent accessibility parameter obtained by using a 1.4-Å probe radius.

accord with those determined experimentally;⁶⁶ (iii) the fields and field gradients derived from the LRF-MD method are obtained in basically the same way as those deduced in previous studies of the pK_a values of the bovine pancreatic trypsin inhibitor,⁵⁴ although we use a higher backbone/heavy atom constraint of 0.1 kcal mol⁻¹ Å⁻²; and (iv) both the *E. coli* and *S. typhimurium* structures yield very similar shielding patterns (see below). A detailed comparison of results for the self-energies of Asp 3 and Glu 7 in the bovine pancreatic trypsin inhibitor using different force fields and harmonic constraints showed good agreement between the MOLARIS88 force field used in ref 54 and MOLARIS92 (an updated improvement on MOLARIS88, used in this work). These shielding results on GBP are thus in accord with our work on model C₆H₅F-(HF)_n clusters⁵¹ and imply the dominance of the uniform and nonuniform (gradient) electric field terms, basically as suggested previously in semiempirical work by Batchelor⁴⁷ for ^{13}C NMR.

The basic reason for the importance of the field/field gradient terms lies in the fact that the "dielectric constants" in proteins are often relatively low (see, however, ref 70), and there is only limited motional averaging of the electrical fields and field gradients. The electrostatic fields are thus often very large (up to $\sim 4 \times 10^7$ V cm⁻¹) and are highly directional. The (field)² response appears to be relatively small,^{48,51} and the van der Waals dispersion terms do not appear to dominate ^{19}F shielding nonequivalencies in proteins, as deduced from the comparison given in Figure 10, although these effects are very significant in many liquids and gases, where the quasi-static field effects we calculate are motionally averaged toward 0. In addition, of course, the van der Waals term could also make a sizable, but constant, contribution to shielding at each fluorine site.

The GBP results, Figure 9, are also of interest from the standpoint of the actual bandwidths (or standard deviations) of the individual trajectories. While these widths can be modified simply by modifying the MD constraints used, it appears that the width of the [5-F]Trp 284 trajectory is always about twice that of the other groups, and this resonance is also the most highly shielded in the experimental spectrum, Figure 2. We believe this result indicates that Trp 284 is significantly exposed to H₂O, and the large shielding is primarily due to interaction with H₂O molecules. The large width of the trajectory reflects the fact that the H₂O molecules are highly mobile, as compared to the amino acid residues neighboring the more buried sites. Although this is a tentative conclusion, we find using the Lee and Richards ACCESS program⁷³ that Trp 284 is indeed highly exposed to H₂O, and exposure values are shown in Table III, together with the mean values and standard deviations of the mean of a set of five LRF-MD trajectories. The experimental conclusion that surface groups may have large $^2\text{H}_2\text{O}$ solvent isotope shifts (SIS)^{21,74} and are relatively shielded appears to be consistent with this picture. Moreover, the observation of anomalies in the

(73) Lee, B.; Richards, F. M. *J. Mol. Biol.* 1971, 55, 379.

(74) Lauterbur, P. C.; Kaufman, B. V.; Crawford, M. K. In *Biomolecular Structure and Function*; Agris, P. F., Loepky, R. N., Sykes, B. D., Eds.; Academic Press: New York, 1978; pp 329-351.

(72) Koning, T. M. G.; Boelens, R.; van der Marel, G. A.; van Boom, J. H.; Kaptein, R. *Biochemistry*, 1991, 30, 3787.

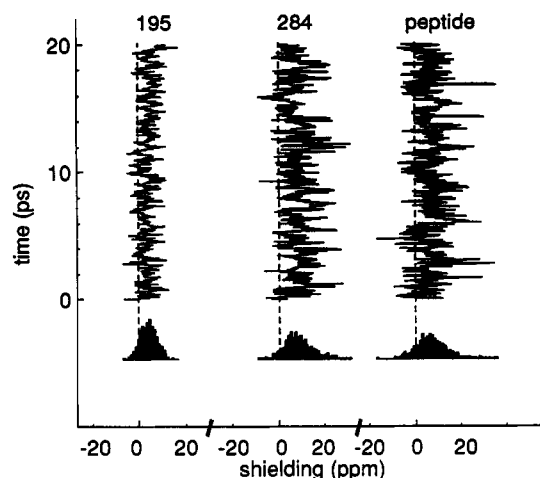


Figure 11. Shielding trajectories (20 ps) for [5- ^{19}F]Trp 195 and 284 in GBP and for Gly-Gly-[5- ^{19}F]Trp-Gly-Gly/ H_2O , at 300 K.

^2H SIS,²¹ in which very highly shielded residues may have no ^2H SIS, can be readily explained in terms of our electrical interaction model, since resonances may be highly shielded either by interaction with buried residues (e.g. Trp 133 appears buried in the X-ray structure of *E. coli* GBP and has a very small ACCESS solvent accessibility parameter) or, apparently, by interaction with H_2O molecules. We thus posit that ^{19}F resonances can be highly shielded either by interaction with buried residues (no ^2H SIS) or by interaction with water (large ^2H SIS). On the other hand, highly deshielded residues will seldom have a large ^2H SIS, since the exposure of an aromatic F atom to solvent water is unlikely to result in the residue being deshielded, although there may exist exceptions when buried but ^2H -exchangeable groups are involved, e.g. H_2O and NH_2 .

To follow up on the idea that H_2O causes shielding of F-Trp residues, we have carried out ^{19}F NMR LRF-MD trajectories on a model pentapeptide, Gly-Gly-[5-F]Trp-Gly-Gly, in H_2O . The resultant trajectory, $\sigma(\Phi)f(\tau)$, is shown in Figure 11, together with, for comparison, shielding trajectories for Trp 195 and Trp 284. Clearly, the peptide experiences a large shielding, as well as large fluctuations in σ , in large part we believe because it is highly exposed to H_2O . In proteins, large ^2H SIS are typically about 0.2–0.3 ppm, similar to the small molecule ^2H SIS values of ~ 0.25 ppm.⁷⁴ Thus, the increase in width of the Trp 284 and pentapeptide shielding trajectories can, at least qualitatively, be rationalized with the experimental data base. Trp 284 is exposed (ACCESS exposure is high), is shielded (interaction with H_2O), has a large shielding trajectory width (H_2O movement), and is in the typical chemical shift range for a surface residue (as deduced from SIS results in other systems). Similarly, the increased shielding of the pentapeptide as compared with that found in a denatured protein^{4,69} is consistent with an even larger H_2O exposure in the peptide.

Our simple physical picture for the shielding of exposed [F]-Trp residues is thus that the C–F bond symmetry in addition to the residual F charge tends to orient adjacent water molecules such that the water protons are closer to the [F] Trp fluorine than the water oxygen atoms, an arrangement which on the basis of *ab initio* calculations⁵¹ is likely to cause shielding. This result is also physically reasonable, since the alignment of water dipoles with the negatively charged fluorine site would be expected to increase electron density at the fluorine end of the C–F bond.

Table IV gives a compilation of all of the electric fields and electric field gradients of [5-F]Trp residues in GBP, averaged over five 20-ps LRF-MD trajectories, together with their associated contributions to shielding, and the standard deviations of the values given. Results for five different trajectories are shown, together with the averages of the five trajectories, for the

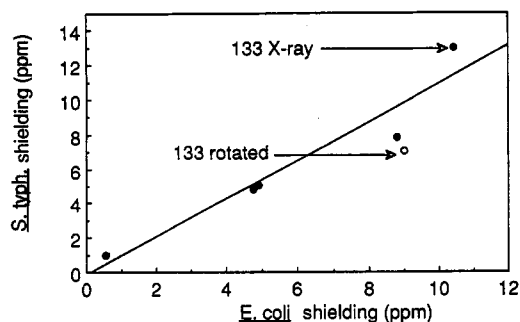


Figure 12. Graph showing the correlation between computed [5- ^{19}F]Trp shieldings in *E. coli* and *S. typhimurium* GBPs. Values shown are averages of five 20-ps shielding trajectories.

E. coli protein. Also shown are the results from equivalent calculations on the *S. typhimurium* protein and our results on the pentapeptide.

There are several important results shown in Table IV. First, we find that the overall patterns of shielding for the *E. coli* and *S. typhimurium* proteins are extremely similar, and this is illustrated graphically in Figure 12. The *Salmonella* protein has 307 residues in its amino acid sequence, compared with 309 for the *E. coli* protein, with a total of 21 substitutions/additions between the two. However, inspection of the X-ray structures^{61,62} of the two proteins shows them to be extremely similar in the regions of the five Trp residues, and essentially all of the substitutions are relatively "conservative". By this we mean that the types of changes observed are unlikely to cause major changes in shielding, as deduced from an examination of the experimental results for ^{19}F NMR of chicken, quail, duck, and turkey lysozymes we have obtained recently.⁷⁵ A similar overall similarity in shielding for ^{13}C of Trp was previously observed by Oldfield and Allerhand for adult human, chicken AII, bovine fetal, and human fetal hemoglobins,⁷⁶ and when taken together, these results suggest that both ^{13}C and ^{19}F shieldings may often be dominated by relatively local folding patterns. The results of Table IV and Figures 10 and 12 also imply that the Trp conformations in the X-ray structure and [5-F]Trp-labeled GBP in solution are quite similar, in agreement with the observation of very similar (although backbone) crystal-solution chemical shifts in *Staphylococcal* nuclease.⁷⁷

Inspection of the individual field and gradient contributions to shielding (Table IV) shows that the two most highly shielded residues, Trp 133 and Trp 284, have large field contributions to their overall shielding, but the gradient contribution for Trp 284 is small. As alluded to before, we believe the large field but small gradient terms for Trp 284 suggest that F may have an orienting effect on H_2O such that the long-range (field) term dominates. The most deshielded group, Trp 183, has a large and negative gradient contribution, due to its particular local environment, and this gradient term is partly responsible for the small overall shielding. On the other hand, the gradient term for Trp 133 is equally large but of opposite sign and contributes to the very large overall shielding for this site, as shown in Table IV and Figure 2.

We have also observed that Trp 133 and Trp 284 occasionally switch sequence in the various shielding trajectories, i.e., the experimental ordering of chemical shifts is not always obtained. This is not particularly surprising since in the experimental spectrum, Figure 2, Trp 133 is actually split into two resonances, possibly due to *cis-trans* isomerism of a neighboring proline residue,⁴ an effect not apparent in the X-ray result.^{61,62} Thus, some difference between theory and experiment might be

(75) Oldfield, E.; Montez, B.; Pearson, J.; Patterson, J.; Harrell, S.; Lian, S. Unpublished results.

(76) Oldfield, E.; Allerhand, A. *J. Biol. Chem.* **1975**, *250*, 6403.

(77) Cole, H. B. R.; Sparks, S. W.; Torchia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6362.

Table IV. Electric Fields, Field Gradients, and Chemical Shifts for *E. coli* Galactose Binding Protein and a Model Pentapeptide

| Trp residue | $E_x \times 10^3$ (au) | $V_{xx} \times 10^3$ (au) | $V_{yy} \times 10^3$ (au) | $V_{zz} \times 10^3$ (au) | $\sigma(E_x)$ (ppm) | $s(\sigma(E_x))$ (ppm) | $\sigma(V_{ii})$ (ppm) | $s(\sigma(V_{ii}))$ (ppm) | $\Delta\sigma(\Phi)$ (ppm) | $s(\sigma)$ (ppm) |
|-----------------------|---------------------------|------------------------------|------------------------------|------------------------------|------------------------|---------------------------|---------------------------|------------------------------|-------------------------------|----------------------|
| <i>E. coli</i> | | | | | | | | | | |
| 127 | | | | | | | | | | |
| a | 0.560 | -0.673 | -0.351 | -1.02 | 1.1 | 3.8 | 3.3 | 4.0 | 4.3 | 3.9 |
| b | -0.032 | -0.927 | +0.055 | +0.872 | -0.1 | 4.4 | 4.0 | 4.3 | 4.0 | 4.1 |
| c | 0.375 | -1.000 | -0.164 | +1.16 | 0.7 | 3.3 | 4.5 | 3.0 | 5.2 | 3.3 |
| d | 0.212 | -0.556 | -0.483 | +1.04 | 0.4 | 3.7 | 2.9 | 3.4 | 3.3 | 4.2 |
| e | 1.040 | -0.927 | -0.899 | +1.83 | 2.0 | 3.7 | 4.9 | 3.2 | 6.9 | 3.5 |
| av | 0.431 | -0.817 | -0.368 | +0.776 | 0.8 | 3.8 | 3.9 | 3.6 | 4.7 | 3.8 |
| 133 | | | | | | | | | | |
| a | 4.61 | -0.768 | +0.790 | -0.023 | 8.7 | 3.3 | 2.6 | 3.2 | 11.3 | 3.4 |
| b | 4.99 | -0.541 | +0.811 | -0.271 | 9.4 | 2.7 | 1.6 | 3.2 | 11.0 | 3.2 |
| c | 3.42 | +0.223 | +0.928 | -1.15 | 6.5 | 3.0 | -1.8 | 2.9 | 4.6 | 4.4 |
| d | 5.09 | -0.918 | +0.975 | -0.057 | 9.6 | 2.7 | 3.1 | 2.3 | 12.7 | 2.8 |
| e | 4.75 | -1.000 | +0.953 | +0.048 | 9.0 | 2.9 | 3.5 | 2.7 | 12.5 | 2.6 |
| av | 4.57 | -0.601 | +0.891 | -0.291 | 8.6 | 2.9 | 1.8 | 2.9 | 10.4 | 3.3 |
| 183 | | | | | | | | | | |
| a | 0.02 | +0.645 | -0.159 | -0.486 | 0.0 | 7.3 | -2.7 | 7.9 | -2.6 | 4.4 |
| b | 1.04 | +0.565 | +0.387 | -0.952 | 2.0 | 6.9 | -2.8 | 8.0 | -0.9 | 4.3 |
| c | 1.75 | +0.023 | +0.351 | -0.374 | 3.3 | 5.9 | -0.4 | 5.7 | 2.9 | 4.3 |
| d | 2.95 | +0.172 | +0.055 | -0.227 | 5.6 | 4.7 | -0.8 | 5.5 | 4.7 | 4.2 |
| e | 1.61 | +1.050 | -0.130 | -0.918 | 3.0 | 7.9 | -4.5 | 6.9 | -1.4 | 4.3 |
| av | 1.47 | +0.491 | +0.101 | -0.591 | 2.8 | 6.5 | -2.2 | 6.8 | 0.5 | 4.3 |
| 195 | | | | | | | | | | |
| a | 1.46 | +0.226 | -2.29 | +2.07 | 2.7 | 3.2 | 1.2 | 2.1 | 3.9 | 3.5 |
| b | 1.20 | -0.070 | -2.27 | +2.34 | 2.3 | 3.0 | 2.4 | 2.1 | 4.7 | 3.3 |
| c | 1.73 | +0.072 | -2.23 | +2.16 | 3.3 | 3.0 | 1.8 | 1.7 | 5.0 | 3.4 |
| d | 1.53 | +0.029 | -2.44 | +2.41 | 2.9 | 2.7 | 2.2 | 1.6 | 5.0 | 3.0 |
| e | 2.18 | +0.136 | -2.58 | +2.44 | 4.1 | 3.1 | 1.8 | 2.0 | 5.9 | 3.2 |
| av | 1.62 | +0.079 | -2.36 | +2.28 | 3.1 | 3.0 | 1.9 | 1.9 | 4.9 | 3.3 |
| 284 | | | | | | | | | | |
| a | 4.56 | -0.170 | +0.513 | -0.343 | 8.6 | 7.0 | 0.3 | 6.8 | 8.9 | 6.4 |
| b | 3.23 | -0.383 | -0.721 | +1.10 | 6.1 | 6.6 | 2.4 | 7.4 | 8.4 | 6.6 |
| c | 5.28 | +0.500 | -0.133 | -0.367 | 10.0 | 8.8 | -2.1 | 8.1 | 7.9 | 5.6 |
| d | 4.24 | +0.025 | -0.312 | +0.287 | 8.0 | 5.8 | 0.2 | 5.9 | 8.2 | 5.5 |
| e | 4.42 | -0.422 | -0.457 | +0.879 | 8.3 | 7.5 | 2.3 | 9.0 | 10.6 | 8.4 |
| av | 4.35 | -0.090 | -0.222 | +0.311 | 8.2 | 7.1 | 0.6 | 7.4 | 8.8 | 6.5 |
| 133 (flip) | | | | | | | | | | |
| a | 7.97 | 0.682 | 0.608 | -1.29 | 15.0 | 8.8 | -3.6 | 11.5 | 11.5 | 7.8 |
| b | 7.02 | 1.33 | 0.076 | -1.40 | 13.2 | 9.5 | -5.9 | 11.4 | 7.3 | 7.9 |
| c | 6.59 | 0.795 | 0.216 | -1.01 | 12.4 | 8.1 | -3.7 | 9.1 | 8.7 | 7.2 |
| d | 6.80 | 0.729 | 0.538 | -1.27 | 12.8 | 8.8 | -3.7 | 10.7 | 9.1 | 8.9 |
| e | 6.76 | 0.957 | 0.174 | -1.13 | 12.7 | 9.1 | -4.4 | 10.2 | 8.4 | 7.3 |
| av | 7.03 | 0.899 | 0.322 | -1.22 | 13.2 | 8.9 | -4.3 | 10.6 | 9.0 | 7.8 |
| <i>S. typhimurium</i> | | | | | | | | | | |
| 127 | | | | | | | | | | |
| a | 0.428 | -0.485 | -0.399 | +0.884 | 0.9 | 3.3 | 4.0 | 3.2 | 4.8 | 3.4 |
| b | -0.289 | -0.566 | -0.504 | +1.071 | 1.3 | 3.4 | 2.5 | 3.2 | 3.8 | 4.1 |
| c | -0.009 | -0.565 | -0.582 | +1.147 | 21.9 | 10.8 | -17.0 | 14.1 | 4.9 | 5.8 |
| d | -0.511 | -0.527 | -0.437 | +0.964 | 1.5 | 3.4 | 4.3 | 3.4 | 5.8 | 3.4 |
| av | -0.095 | -0.536 | -0.480 | +1.016 | 6.4 | 5.2 | -1.6 | 6.0 | 4.8 | 4.2 |
| 133 | | | | | | | | | | |
| a | 5.93 | -0.881 | +0.823 | +0.058 | 10.6 | 2.7 | 2.6 | 2.7 | 13.2 | 2.9 |
| b | 5.99 | -0.769 | +0.492 | +0.277 | 12.1 | 2.8 | 1.3 | 3.0 | 13.4 | 3.7 |
| c | 6.10 | -0.814 | +0.509 | +0.305 | 11.4 | 2.3 | 1.0 | 2.3 | 12.3 | 2.6 |
| d | 5.95 | -0.819 | +0.454 | +0.365 | 12.2 | 2.7 | 1.4 | 2.6 | 13.1 | 3.3 |
| av | 5.99 | -0.821 | +0.570 | +0.251 | 11.6 | 2.6 | 1.4 | 2.6 | 13.0 | 3.1 |
| 183 | | | | | | | | | | |
| a | -0.988 | -0.636 | +0.318 | +0.318 | 3.5 | 6.7 | -1.8 | 6.0 | 1.7 | 3.9 |
| b | -2.50 | -1.71 | +0.398 | +1.32 | 6.6 | 6.5 | -2.0 | 6.6 | 4.6 | 4.6 |
| c | -1.23 | -1.23 | +0.948 | +0.281 | 3.7 | 6.6 | -5.0 | 7.9 | -1.2 | 5.2 |
| d | -1.60 | -1.43 | +0.462 | +0.969 | 2.9 | 7.2 | -4.2 | 8.2 | -1.3 | 4.1 |
| av | -1.58 | -1.25 | 0.532 | +0.721 | 4.2 | 6.8 | -3.2 | 7.2 | 1.0 | 4.4 |
| 195 | | | | | | | | | | |
| a | 3.11 | +0.689 | -2.55 | +1.86 | 9.1 | 7.3 | -3.2 | 8.8 | 6.0 | 4.8 |
| b | 1.78 | +0.633 | -2.57 | +1.94 | 5.5 | 3.9 | -1.2 | 3.8 | 4.3 | 2.9 |
| c | 2.66 | +0.751 | -2.65 | +1.90 | 8.2 | 7.2 | -3.2 | 8.0 | 5.1 | 5.3 |
| d | 2.73 | +0.249 | -1.83 | +1.58 | 5.4 | 3.8 | -0.7 | 3.1 | 4.8 | 3.1 |
| av | 2.57 | +0.580 | -2.40 | +1.82 | 7.0 | 5.6 | -2.1 | 5.9 | 5.0 | 4.0 |
| 284 | | | | | | | | | | |
| a | 3.35 | -0.103 | -0.017 | +0.120 | 6.3 | 5.7 | 0.5 | 5.4 | 6.8 | 5.9 |
| b | 5.24 | +0.328 | -0.202 | -0.126 | 9.9 | 7.2 | -1.3 | 8.0 | 8.6 | 6.0 |
| c | 5.32 | +0.563 | -0.899 | +0.336 | 10.0 | 6.6 | -1.6 | 8.3 | 8.4 | 6.2 |
| d | 4.68 | +0.271 | +0.005 | -0.277 | 8.8 | 6.6 | -1.2 | 7.5 | 7.6 | 5.0 |
| av | 4.65 | +0.265 | -0.278 | +0.013 | 8.8 | 6.5 | -0.9 | 7.3 | 7.8 | 5.8 |

Table IV. (Continued)

| Trp residue | $E_x \times 10^3$ (au) | $V_{xx} \times 10^3$ (au) | $V_{yy} \times 10^3$ (au) | $V_{zz} \times 10^3$ (au) | $\sigma(E_x)$ (ppm) | $s(\sigma(E_x))$ (ppm) | $\sigma(V_{ii})$ (ppm) | $s(\sigma(V_{ii}))$ (ppm) | $\Delta\sigma(\Phi)$ (ppm) | $s(\sigma)$ (ppm) |
|---|---------------------------|------------------------------|------------------------------|------------------------------|------------------------|---------------------------|---------------------------|------------------------------|-------------------------------|----------------------|
| <i>S. typhimurium</i> | | | | | | | | | | |
| 133 (flip) | | | | | | | | | | |
| a | 5.49 | 0.688 | -1.05 | 0.358 | 10.3 | 8.7 | -2.0 | 10.8 | 8.3 | 8.2 |
| b | 5.73 | 1.01 | 0.350 | -1.36 | 10.8 | 9.6 | -4.7 | 11.1 | 6.1 | 7.8 |
| c | 5.44 | 0.987 | 0.466 | -1.45 | 8.5 | 8.1 | -1.0 | 9.0 | 7.5 | 7.2 |
| d | 3.87 | -0.126 | -0.127 | 0.252 | 7.3 | 7.1 | 0.7 | 8.0 | 8.0 | 8.3 |
| av | 5.13 | 0.640 | -0.090 | -0.550 | 9.2 | 8.4 | -1.8 | 9.7 | 7.5 | 7.9 |
| Gly-Gly-[5-F]Trp-Gly-Gly/H ₂ O | | | | | | | | | | |
| [5-F]Trp | 4.42 | 0.266 | -0.035 | -0.232 | 8.3 | 6.0 | -1.2 | 7.8 | 7.1 | 6.0 |

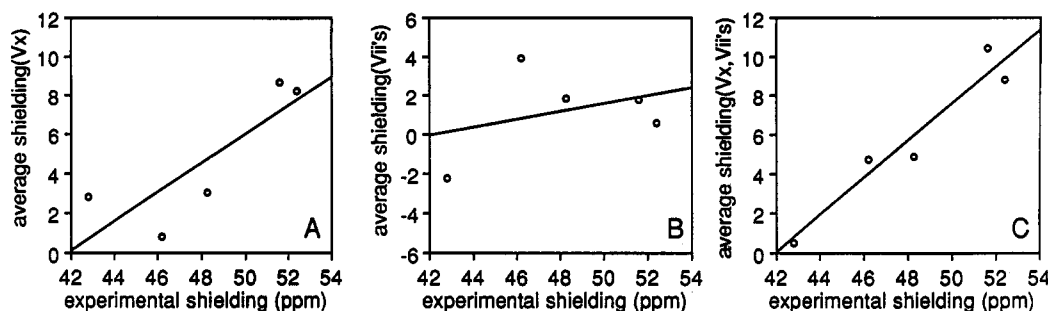


Figure 13. Field, field gradient, and total electrical contributions to shielding for all five [5-¹⁹F]Trp sites in *E. coli* GBP: A, uniform field contribution; B, field gradient contribution; and C, total field plus field gradient contribution, all plotted as a function of the experimental shieldings.

anticipated for Trp 133. Also, it must be noted that we are using a nonfluorinated crystal structure to represent a fluorinated solution structure, so other specific perturbations cannot be ruled out. For example, examination of the X-ray structure^{61,62} suggests that a 180° rotation about C^β-C^γ is quite possible for Trp 133, and in this case we find a ~2 ppm deshielding, as shown in Table IV and Figures 10 and 12. Results on other ¹⁹F-labeled GBPs⁴ or e.g. on C^γ of Trp in GBP may help resolve some of these uncertainties, as might the results of a ²H SIS experiment. Given that three elements of the dipole shielding polarizability tensor, nine elements of the quadrupole shielding polarizability tensor, one field and nine field gradient tensor elements, and the fluorine (and other e.g. Ca²⁺, C, O sugar) charges all have to be incorporated into the shielding calculations, the good agreement between the *E. coli* and *S. typhimurium* calculations and the experimental *E. coli* GBP spectrum is quite encouraging.

Further inspection of the results in Table IV shows that, of all the various contributions to shielding, it appears to be the uniform field component which dominates, a conclusion we drew previously on the basis of the results of both NMR and IR experiments.^{12,16,41} We show in Figure 13 experimental *versus* theoretical plots of chemical shielding for the field, field gradient, and total shifts computed for *E. coli* GBP, using the average values obtained from the five shielding trajectory results given in Table IV. The fact that the field/dipole shielding polarizability term dominates the overall shielding patterns is to be expected since convergent results are only obtained when large numbers of atoms are incorporated. In model calculations, we find that the uniform field term dominates at $r \gtrsim 7$ Å,^{44,51} while at shorter distances ($r \lesssim 5$ Å), the gradient terms become increasingly important.⁵¹ Thus, the importance of the uniform field in influencing many NMR and IR frequency shifts can be appreciated, although the gradient terms may also be very important in particular cases (e.g. Trp 183) and clearly contribute significantly to the overall goodness-of-fit to the experimental results, as seen in Figure 13.

Conclusions

The results we have presented above are of interest for a number of reasons. First, they represent the first reasonably successful attempt at calculating the ¹⁹F NMR spectrum of a protein. The overall shielding appears to be dominated by weak electrical

interactions, describable by use of a series of multipole shielding polarizability tensors and the electrostatic potential of the protein. Second, the observation that there is quite good agreement between the shielding patterns and overall spectral breadths between theory and experiment indicates that the LRF-MD method of computing the average fields and field gradients is probably accurate, since elsewhere we have shown that the shielding polarizability tensors yield, in model systems, results which are quite close to those obtained by using full *ab initio* calculations with large basis sets.⁵¹ Third, our results imply that surface charge does not contribute much toward the ¹⁹F shielding observed experimentally, although H₂O molecules do play a role. Fourth, the use of the LRF-MD method appears to give a good means of obtaining average fields and field gradients in proteins. Fifth, we find that the computed shielding patterns of the *E. coli* and *S. typhimurium* galactose binding proteins are very similar to each other (and to the *E. coli* experiment), due to equivalent folding patterns (i.e., side-chain orientations, CONH locations, and proximity to H₂O molecules). Sixth, our results tend not to support the dominance of van der Waals interactions in governing ¹⁹F shielding nonequivalencies in proteins, on the basis of the observation that there is good agreement between theory and experiment using only the fields and field gradient interactions. The Taylor series expansion of the potential used in the LRF model to describe the protein energetics also seems the most appropriate expansion to describe shielding, a point suggested in previous work by Batchelor.⁴⁸ Seventh, our results suggest a possible explanation for previous observations on ²H solvent isotope shifts, and ¹⁹F relaxation, in proteins. For example, several workers have noticed^{21,74} that the more shielded resonances in the ¹⁹F NMR spectrum of a protein may often exhibit a ²H SIS, consistent with the observation that the exposed Trp 284 in GBP is highly shielded, as is the [5-F]Trp in a model Gly-Gly-[5-F]Trp-Gly-Gly pentapeptide calculation, due to interaction with solvent water. However, it has also been noted that some highly shielded resonances may exhibit no ²H SIS, which would be consistent with a large field/field gradient shift of a buried residue. In our picture, deshielded residues would seldom have large ²H solvent isotope shifts, as has been noted experimentally.²¹ The relaxation rate-chemical shift profile seen in [3-F]Tyr *E. coli* alkaline phosphatase²¹ is also consistent with the ideas presented above

in that exposed or partially exposed residues will be relatively shielded (near the denatured chemical shift position) and have relatively low relaxation rates,²¹ while resonances which are buried are expected to have faster relaxation rates, but can be either shielded or deshielded, depending on the fields and field gradients experienced. The highly directional model we have described thus appears to give a consistent picture of previous shift, relaxation, and isotope-shift results. Eighth, our results using the LRF-MD shielding trajectory approach are similar to those we have obtained by using an *ab initio* approach where we have introduced the protein's charge field at the SCF level, using AMBER charges.⁴⁴ Thus, in future work, there is hope for determining absolute shieldings, rather than the relative shieldings

we have used here. Ninth, and finally, it is possible that the present strategy may prove to be a useful experimental method with which to probe local electric fields in proteins. This could have implications in establishing the importance of electrostatic interactions in enzyme catalysis^{57,78} and may be applicable to nucleic acids as well.⁷⁹

Acknowledgment. One of us (E.O.) thanks Michael Schaefer, Klaus Schulten, Clifford Dykstra, and J. Augspurger for helpful advice.

(78) King, G.; Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366.

(79) Chu, W.-C.; Horowitz, J. *Biochemistry* **1991**, *30*, 1655. Chu, W.-C.; Horowitz, J. *FEBS Lett.* **1991**, *294*, 159.