Proteins and Model Systems: Spectral Analyses

Eric Oldfield

University of Illinois at Urbana-Champaign, Urbana, IL, USA

1 INTRODUCTION

The first NMR spectra of proteins, obtained in the 1950s and 1960s, focused on the ^{1}H nucleus because of its spin $I = 1/2$ nature, high natural abundance, and sensitivity.[1–6] However, because of the small range of chemical shifts (due to folding) and the low spectral resolution available on 60-, 100-, and even 220-MHz instruments, there was relatively little spectral resolution available, and only a handful of individual proton resonances could be assigned, precluding use of computational methods to predict spectra from the structure. This problem of a small ^{1}H chemical shift range was well known to organic chemists, and several groups, in particular those of Roberts,[7] Sternlicht[8], and Grant[9], had embarked on the observation of ^{13}C nuclei that, while having only a 1% natural abundance and low sensitivity, were known to have a $\sim$10$\times$ larger chemical shift range than ^{1}H, opening up the intriguing prospect of ^{13}C NMR of macromolecules, such as proteins. To get improved sensitivity (or at least, signal-to-noise ratios), these workers developed much larger (10–12 mm) sample probes (5 mm was, and still is, the standard for most solution samples); in addition, the advent of ^{1}H-decoupling[10] improved spectral resolution considerably. However, it took Ernst and Anderson's development of the Fourier transform method[11] to provide genuine sensitivity enhancements, opening the way to protein structure studies.

2 SOLUTION NMR

2.1 Chemical Shift Nonequivalence in Proteins

We attempted such studies in 1971 on cytochrome *c*, but our studies, as well as those of others, did not succeed in resolving single carbon atom sites in proteins, and the question arose as to whether ^{13}C chemical nonequivalencies in proteins, due to folding into their native conformations, were actually going to be large enough (compared with their linewidths) to enable observation of individual sites. Here, early theoretical work by Kuhlmann *et al*.[12] and Doddrell *et al*.[13] using the Solomon equations[14] provided important insights since their work enabled the prediction of linewidths (W), spin-lattice relaxation times (T_1) as well as nuclear Overhauser effects (NOEs) as functions of the rotational correlation time, τ_R. These theoretical predictions were supported by the results of W, T_1, and NOE measurements[15] on homopolymeric polypeptides, in particular, poly(γ-benzyl-L-glutamate), which adopts either an α-helical or a random coil configuration, depending on the composition of the solvent. The $C^\alpha T_1$ values changed by less than a factor of 2 when going from an α-helical to a random coil, but the NOE changed from 1.1 (a 10% enhancement) to >2 (a 100% enhancement), since the similarity in T_1 values was simply due to correlation times that were on either side of the T_1 minimum and corresponded to τ_R values of 24–32 ns (helix) and $\sim$0.8 ns (for the random coil). The corresponding linewidths for the helices were predicted to be approximately 40–50 Hz (approximately 30–50 Hz experimental), to be compared with the approximately 8–10 Hz predicted for the random coil structures. In the 1-D world of the early 1970s, these observations made it unlikely that irrotationally bound nuclei, such as C^α, would be accessible, since while 50 Hz these days does not sound so large, in 1972 and on a 60-MHz ^{1}H (15-MHz ^{13}C) NMR instrument, 50 Hz corresponds to a $\sim$3 ppm linewidth, which might reasonably be expected to be a large fraction of any chemical shift range due to conformational or nonbonded interactions, as exemplified, for example, by the $\sim$3–4 ppm downfield shift of C^α in the α-helical versus random coil samples.[15] Thus, observing rigid CH groups would be difficult, and rigid CH_2 groups would be twice as bad, since relaxation is purely dipolar.

Fortunately, however, the dipolar nature of ^{13}C–^{1}H relaxation in these systems (at 15 MHz) applies to both aliphatic (C^α) and aromatic systems, and in aromatics (His, Tyr, Phe, and Trp) there are an interesting, second class of carbons: nonprotonated aromatic carbons (Figure 1). For example, in tryptophan, there are three nonprotonated aromatics, C^γ, $C^{\delta 2}$, and $C^{\varepsilon 2}$. For these carbons, at low fields, relaxation is due to dipolar interactions with distant protons, and since the linewidth for dipolar relaxation is given by the following equation:

$$W = (20\pi)^{-1}\hbar^2\gamma_c{}^2\gamma_H{}^2 N r_{CH}{}^{-6} f(\tau_R) \tag{1}$$

where W is the linewidth in Hertz, γ_C and γ_H are the gyromagnetic ratios of ^{13}C and ^{1}H, respectively, N is the number of directly attached hydrogens, r_{CH} is the C–H bond length, and $f(\tau_R)$ is defined as

$$f(\tau_R) = 4\tau_R + \frac{\tau_R}{1+(\omega_H-\omega_C)^2\tau_{R^2}} + \frac{3\tau_R}{1+\omega_c^2\tau_R^2} + \frac{6\tau_R}{1+(\omega_H+\omega_C)^2\tau_R^2} + \frac{6\tau_R}{1+\omega_H^2\tau_R^2} \tag{2}$$

where ω_c and ω_H are the Larmor frequencies. The linewidths for these nonprotonated carbons is, therefore, predicted to be only a few Hertz, due to the r_{CH}^{-6} distance term. But still, the question is, what about sensitivity or signal-to-noise ratios?

There were three key developments that enabled the detection of single carbon atom sites in proteins. The first was the use of very large (20-mm) sample tube probes that used $\sim$13-mL protein solution, typically $\sim$1 g of any protein that

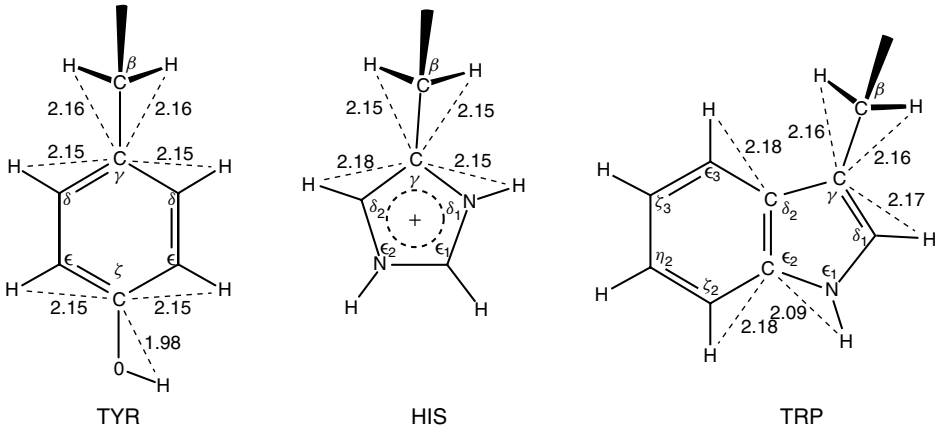

Figure 1 Estimated distances of nonprotonated aromatic carbons (of tyrosine, histidine, and tryptophan residues) to hydrogens that are two bonds removed. These distances were computed with the use of known bond lengths and angles in some crystalline amino acids and small peptides and on known CH, OH, and NH bond lengths in smaller molecules. The calculated two-bond CH distances for C^{γ} of Phe are about the same as those shown for Tyr. (Reproduced from Ref. 16. © American Chemical Society, 1975)

was not too expensive (lysozyme, cytochrome *c*, myoglobin, and hemoglobin). The second was the use of noise off-resonance proton decoupling,[17] a weak effective decoupling field that removed the long-range $^{13}C-^{1}H$ J-couplings while broadening directly bonded ($^{1}J_{CH}$) interactions. The third was the use of an RF crystal filter[18] that removed image noise—essentially the same effect as obtained with quadrature detection. All of these features were implemented in Adam Allerhand's laboratory and resulted in the first ^{13}C NMR spectrum of a protein in which numerous resolved, single carbon atom sites could be observed.[19,20] The results were exciting in that with lysozyme, which contains six Trp residues, five peaks were observed in the region of C^{γ}, one a two-carbon resonance; the chemical shift range was ~6.5 ppm; and linewidths, as expected, were very narrow. At higher field and with the use of a 20-mm sideways spinning solenoidal coil geometry in a superconducting magnet,[21] sample volumes decreased and sensitivity increased, and early results on lysozyme showing the large C^{γ} shift range, as well as the large Tyr C^{ζ} shift range in ribonuclease, are shown in Figure 2. The shift ranges were due to folding, since they disappeared on denaturation, but the question was, what actually caused these shifts? They did not correlate with backbone ϕ, ψ torsion angles, and so the question arose as to whether they were due to χ_1, χ_2, hydrogen bonding, electrostatics, solvation, or dynamics. Answering this remained a problem for the next 30–35 years. In addition, in these early studies, major differences between the ^{13}C spectra of diamagnetic and paramagnetic proteins were demonstrated (Figure 3),[22] something that had been observed previously with ^{1}H NMR by Wüthrich[5] and Gupta and Redfield,[23] and these and other related hyperfine effects present yet another challenge for prediction. However, first, we need to consider the "easy," diamagnetic proteins.

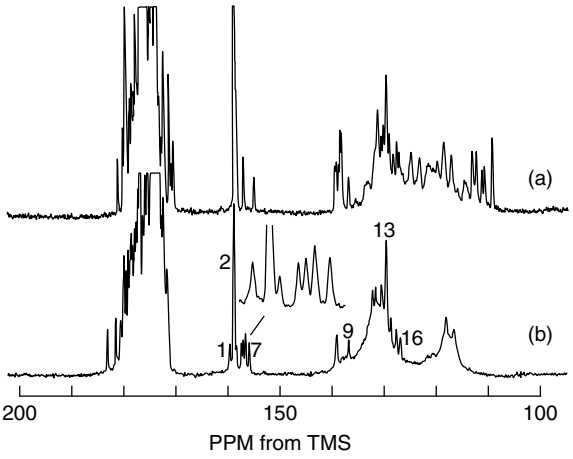

Figure 2 Aromatic, carbonyl and C^{ζ} of arginine region of the proton-decoupled natural-abundance carbon-13 Fourier transform NMR spectra of 6.5-mL aqueous solutions of (A) hen-egg-white lysozyme and (B) bovine pancreatic ribonuclease. (Reproduced from Ref. 21. © Elsevier, 1978)

2.2 Electric Field Effects on Chemical Shifts: The Basics

There was almost a 20-year gap between the first observation of the effects of protein structure on ^{13}C chemical shifts in proteins and the first successful prediction of these shifts, which necessitated the use of quantum mechanical (QM) or quantum chemical methods. In our group, we had been investigating ^{17}O NMR shifts and electric field gradients (EFGs) in inorganic solids, such as minerals,[24] as well as ^{13}C NMR spectra of metal carbonyls on catalyst surfaces. This led to the idea that it might be interesting to look at ^{13}CO and $C^{17}O$ bound to heme proteins, since there might be a large range of ^{13}C, ^{17}O chemical shifts that might be correlated with each other and with other properties, such as the EFG, and even the CO infrared vibrational frequency (ν_{C-O})—which was, in fact, known to vary widely between different heme proteins. As shown in Figure 4, there were indeed good correlations between the ^{17}O shift and ν_{C-O}, between the ^{17}O EFG (e^2qQ/h) and ν_{C-O}, as well as between the ^{17}O EFG (e^2qQ/h) and the isotropic ^{17}O shift, $\delta_i(^{17}O)$.[25] This suggested that there must be a common origin for these diverse correlations, and we proposed[25] an electrostatic charge field or a polarization effect, something that was later calculated theoretically[26] by

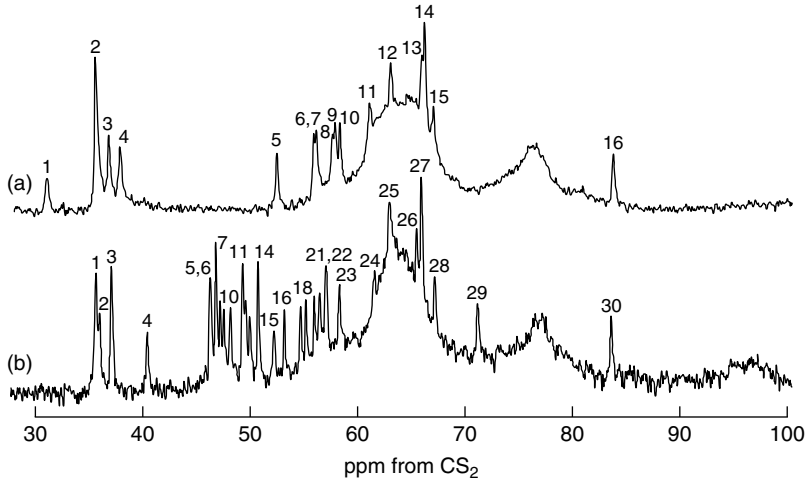

Figure 3 Region of aromatic carbons (and C^{ζ} of the arginine residues) in the noise-modulated, off-resonance, proton-decoupled, natural-abundance ^{13}C NMR spectra of hoarse-heart cytochrome *c* [in 50 mM phosphate buffer (pH 6.7) 41 °C], recorded at 15.18 MHz. (a) 14.4 mM ferricytochrome *c*, after 46 000 accumulations (14 h). (b) 11.5 mM ferrocytochorme *c*, after 16 384 accumulations (5 h). (Reproduced from E. Oldfield, A. Allerhand, Proc. Natl. Acad. Sci. USA, 1973, 70, 3531)

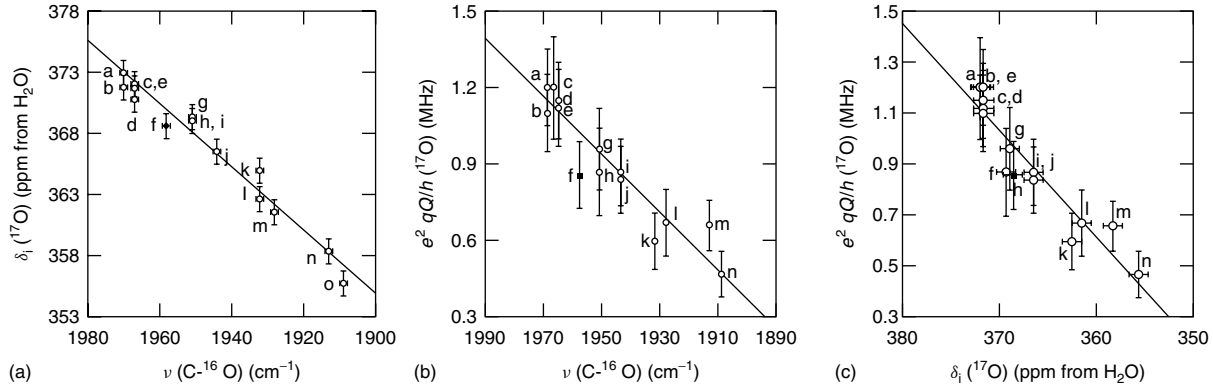

Figure 4 IR/NMR correlations. (a) Graph showing relation between infrared CO vibrational frequency [ν(C–O), cm^{-1}] and ^{17}O NMR isotropic chemical (^{17}O), for heme proteins. (b) Graph showing relation between ν(C–O) and ^{17}O e^2qQ for CO-liganded heme proteins. (c) Graph showing relation between $\delta_i(^{17}O)$ and $C^{17}O$-labeled heme proteins. (Reproduced from Ref. 25. © American Chemical Society, 1991)

using derivative Hartree–Fock (DHF) theory. By analogy to the electrical polarizability, $P_{x,x}$,

$$P_{x,x} = \frac{\partial^2 E}{\partial V_x^2} \tag{3}$$

a second-rank tensor property, the chemical shift is also a second-rank tensor:

$$\sigma_{xy} = \frac{\partial^2 E}{\partial \mu_x \partial H_y} \tag{4}$$

and its derivatives with respect to external electric fields are given by the "dipole shielding polarizabilities":

$$A_{yz,x} = \frac{\partial \sigma_{yz}}{\partial V_x} \tag{5}$$

while the response to a field gradient (e.g., V_{xx}) is given by a fourth rank tensor, the quadrupole shielding polarizability:

$$A_{\alpha\beta,\gamma\delta} = \frac{\partial \sigma_{\alpha\beta}}{\partial V_{\gamma\delta}} \tag{6}$$

Being derivative properties, the *A*-tensors can be obtained analytically. Evaluation of the dipole and quadrupole shielding polarizability (and hyperpolarizability) tensors for CO yielded both large dipole and quadrupole shielding polarizabilities and, by using a variety of electrical perturbations,[26] it was possible to predict each of the spectroscopic perturbations (^{17}O shift, ^{13}C shift, ν_{C-O}, and the ^{17}O-EFG) and hence the correlations seen in the proteins, within a factor of ~2—which is a good result, given that the heme was not part of the calculations. These results meant that the correlations seen experimentally[25] could be attributed primarily to an electrostatic field perturbation of the electronic structure of CO, with the uniform field contribution dominating. In addition, there was no evidence for any hyperpolarizability (V_i^2, V_{ii}^2) contributions to shielding, a significant difference

from the conventional Buckingham power series expansion in the uniform field.

But how much do these electrostatic fields contribute to amino-acid chemical shifts? An analysis of the dipole and quadrupole shielding polarizabilities in a broad range of molecules[27,28] revealed extremely large values for ^{19}F in aromatic systems (such as fluorobenzene), and these results were confirmed by using a charge field perturbation (CFP) approach on fluorobenzene–HF supermolecules and fluorobenzene (+-) clusters,[29] validating the DHF results. Since the range of shifts in ^{19}F-labeled proteins (containing e.g. ^{19}F-Trp, ^{19}F-Tyr residues) was known to be large (10 ppm or more),[30] it seemed it would be possible to calculate ^{19}F shifts in proteins—once the fields and field gradients were known. That is, it should be possible to calculate the electrostatic field contributions to shielding from the following relation:

$$P_{\alpha\beta} - P_{\alpha\beta}{}^{(0)} = P_{\alpha\beta,\gamma}{}^{(1)}F_{\gamma} + P_{\alpha\beta,\gamma\delta}{}^{(1)}F_{\gamma\delta} \qquad (7)$$

where $P_{\alpha\beta} = \sigma_{\alpha\beta}$, the $\alpha\beta$ element of the ^{19}F chemical shielding tensor, $P_{\alpha\beta,\gamma}{}^{(1)} = A_{\alpha\beta,\gamma}$ and $P_{\alpha\beta,\gamma\delta}{}^{(1)} = A_{\alpha\beta,\gamma\delta}$, the shielding polarizabilities. We first investigated[31] the galactose binding protein from *Escherichia coli*, where the ^{19}F shifts and assignments of the five [5-^{19}F] tryptophan residues had been reported by Luck and Falke.[32] To calculate the fields and field gradients, we used the local reaction field model. The effects of motion as well as possible hydration effects were taken into account by using a molecular dynamics (MD) method, which resulted in the production of individual *shielding trajectories*: how the ^{19}F chemical shift varied as a function of time, along a given MD trajectory. As can be seen in Figure 5, there are large fluctuations in the ^{19}F shifts at each of the five [5-^{19}F]Trp sites as a function of time, but when averaged out over several trajectories, a good correlation with the experimental ^{19}F NMR shifts is found.[31] Thus, ^{19}F shifts of ^{19}F-labeled amino-acids, as well as CO shifts, have large contributions from electric field effects. However, ^{13}C nuclei in a variety of systems were found to have much smaller A_i, A_{ii} values, even though it became clear, based on the work of Spera and Bax,[33] that C^{α} (and C^{β}) shifts in proteins had similarly large (up to ~10 ppm) ranges of chemical shifts due to folding. Moreover, these shifts—downfield from the random coil for C^{α} in helices and upfield for C^{α} in sheets (and the opposite trends for C^{β})—clearly correlated with protein secondary structure,[33] but the question is due to backbone ϕ,ψ angles, electrostatics, or to both effects. To answer these questions, a new approach would be needed to predict C^{α} shifts.

2.3 ^{13}C Shift Predictions in Proteins

The electrostatic field models described above were capable of predicting ^{19}F and $^{13}C^{17}O$ ligand shifts in proteins, but on the basis of the computed shielding polarizabilities, electric field effects were unlikely to dominate for C^{α} (or C^{β}). A new approach was thus required to compute these shift properties and fortunately, shortly before Spera and Bax's experimental report,[33] Wolinski *et al*. had reported good chemical shift predictions for ^{17}O in water clusters using Hartree–Fock theory implemented in their TEXAS 90 program.[34] It thus seemed possible that we might solve the chemical shift problem for C^{α}, C^{β} in proteins using a direct method in which we used protein fragments—basically just *N*-formyl-amino acid amides (with or without hydrogen bond partners)—since the observation that the experimental C^{α}, C^{β} shifts correlated in a general way with helix/sheet structure probably meant that ϕ,ψ were dominant, although it might still be that electrostatic field effects (such as helix dipole fields) would contribute to the shifts. In any case, we set about trying to calculate C^{α}, C^{β} shifts in proteins using Hartree–Fock theory.

The first results, obtained by Angel de Dios, were not encouraging since there was a lot of scatter in the theory versus experiment correlations. However, Cynthia Jameson raised the possibility that the X-ray structure we were using might not be accurate enough for QM calculations. This turned out to be the case, since there were considerable variations in bond lengths and bond angles between residues that were not expected, but on protein geometry optimization using a molecular mechanics force-field, the agreement with theory improved considerably and using either this approach or simply by using *N*-formyl-amino acid amides having their torsions angles set to those seen in the X-ray structures, there was good agreement between theory and experiment[35] (Figure 6). The incorporation of bond length/bond angle errors, as well as rovibrational corrections, then enabled good predictions of glycine, alanine, and valine C^{α} shifts in proteins, with the overall ~24 ppm shielding range being predicted with a ~1.4 ppm error[36,37] and, when extended to other proteins,

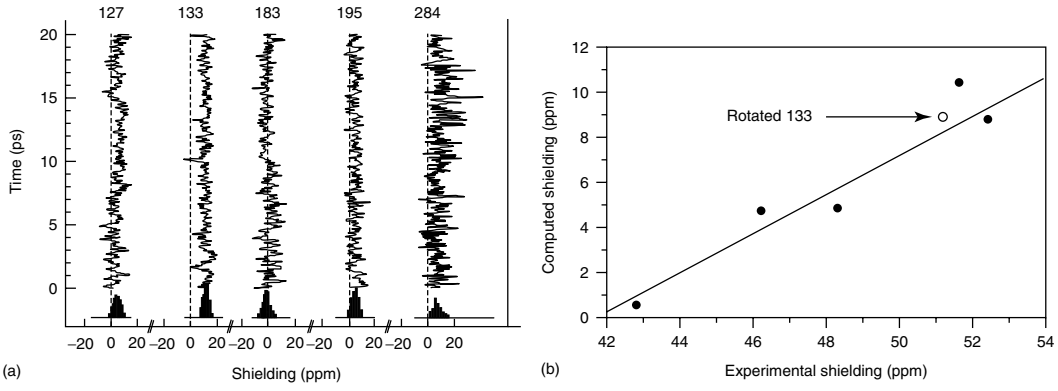

Figure 5 Fluorine shift results. Left, Individual 20-ps shielding trajectories for each of the five [5-F]-Trp residues in GBP. Right, Plot of experimental *versus* theoretical shielding for each [5-F]Trp site in GBP. The computed values are referenced to a field-free value of 0 ppm. The open circle represents the shielding calculated for a 180°C^{β}-C^{γ} flip of Trp 133. (Reproduced from Ref. 31. © American Chemical Society, 1993)

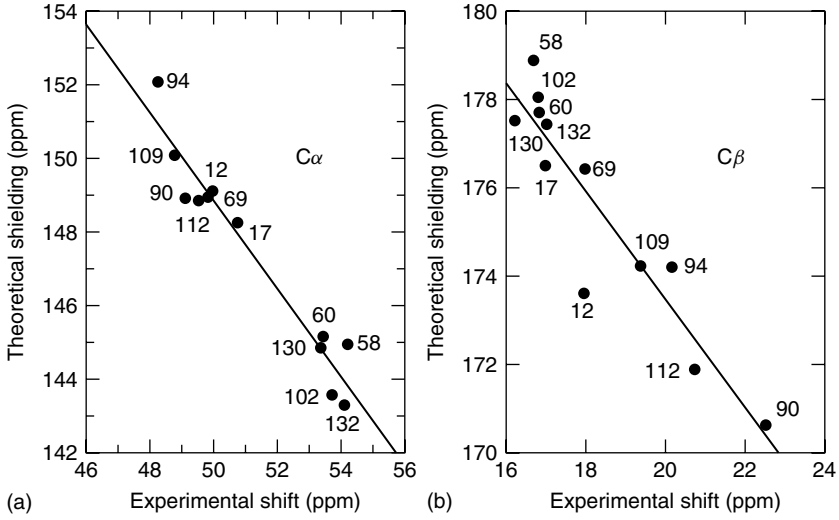

Figure 6 Chemical shift predictions. (a) Experimental C^{α} chemical shifts of the 12 Ala sites in Staphylococcal nuclease (SNase) versus computed shieldings, $m = 1.2$ and $r = 0.97$. (b) Experimental C^{β} chemical shifts for the 12 Ala sites in SNase versus computed shieldings, $m = 1.2$ and $r = 0.93$. (Reproduced from Ref. 37. © American Chemical Society, 1993)

showed that using geometry optimization to select the lowest energy χ_1 conformer for 19 valine residues in 3 proteins gave very good agreement with experimental shifts for both C^{α} and C^{β}.[38] Similar results were obtained by using a shielding hypersurface approach in which we computed $\delta(\phi, \psi, \chi_1)$ for 1728 different structures.[39]

This dominance of ϕ,ψ effects on shielding[40] then led to the use of $^{13}C^{\alpha}$, $^{13}C^{\beta}$ isotropic chemical shifts to try and optimize protein structures.[41,42] The method is called the *Z-surface approach*, the idea being simply to predict the most likely value of an experimental parameter (e.g., ϕ, ψ), that is, the value(s) that give the best accord with experiment. Specifically, for a parameter P that is a function of a single angle, α,

$$P = f(\alpha) \tag{8}$$

we defined the probability that the experimental value of P, P_{expt}, corresponds to a given angle α as an unnormalized probability or Z-surface:

$$Z = e\left(-\frac{(P_{\text{expt}} - f(\alpha))^2}{W}\right) \tag{9}$$

where W represents a search width (to take into account any computational inadequacies or experimental uncertainties). The Z-surface is not normalized in order to retain an absolute measure of how closely the given ϕ,ψ pairs match the experimental data set, $\{P_i\}$. Typical results for ϕ,ψ predictions based on C^{α}, C^{β} shift surfaces (and an empirical H^{α} surface) are shown in Figure 7[43] and provide a new route to structure refinement.

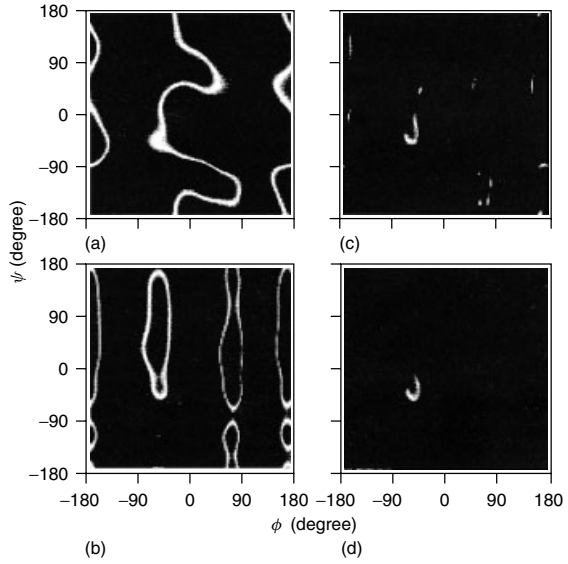

Figure 7 Chemical shift surface prediction of peptide backbone ϕ,ψ torsion angles for Ala^{60} in SNase. (a) $^{1}Z\delta(C^{\alpha})$; (b) $^{1}Z\delta(C^{\beta})$; (c) $^{2}Z\delta(C^{\alpha})$ $\delta(C^{\beta})$; and (d) $^{3}Z\delta(C^{\alpha})$ $\delta(C^{\beta})$ $\delta(H^{\alpha})$. (Reproduced from Ref. 43. © Springer, 1995)

3 SOLID-STATE NMR

3.1 C^{α} Shielding Tensors in Amino-Acids and Peptides

The results described above pertain primarily to solution NMR and open up the possibilities of refining or predicting elements of protein structure, and elsewhere,[44] Vendruscolo and Dobson, and Baker and Bax,[45] have shown how isotropic chemical shifts can be used in the determination of full protein structures, work that has very recently been extended to solids.[46] That is, in both cases, the accurate determination of protein structures was obtained by solely using chemical shifts as experimental input. But there are not many chemical shifts in a protein: at the most N, where N is the number of nuclei whose shifts can be resolved and assigned. However, the chemical shift is actually the trace of the chemical shift (or shielding) tensor:

$$\delta = \frac{1}{3}(\delta_{11} + \delta_{22} + \delta_{33}) \tag{10}$$

where δ_{ii} are the principal components of the tensor, so there are actually $3N$ shift or shielding tensor observables. In addition, the shielding tensor has a specific orientation. There are three Euler angles needed to describe this, so for a given site, there could actually be up to $6N$ observables. If all these can be measured, they could be a great aid in structure determination—especially if they can be accurately calculated.

In early work,[47] we used single-crystal NMR to deduce the shift tensor magnitudes and orientations for each of the ^{13}C nuclei in L-threonine, and in later work, we computed these shielding tensor magnitudes as well as their orientations, together with the tensor magnitudes in L-tyrosine.[48] We found very good agreement for the tensor magnitudes and for threonine (in the icosahedral representation), for both the tensor magnitudes and their orientations.[48] These results led to the idea that it should be possible to predict ϕ,ψ values in peptides with reasonable accuracy, based on C^α shielding tensor magnitudes alone. This proved to be the case with the tripeptide glycylalanyl valine. For G[*A]V, there were three high probability ϕ/ψ pairs predicted using the Z-surface approach, with the highest probability ϕ,ψ result being within $\pm 12^\circ$ of the values determined crystallographically.[49] When the results of shielding tensor magnitudes as well as numerous (C^α–N, C^α–H) orientational restraints were applied to two other tripeptides, NA[*V] and G[*A]L, together with a computed energy function for NAV, the results were even better: ϕ, ψ, χ_1 angles $<6^\circ$ different to those seen crystallographically.[50] But can shielding tensors be measured (and calculated) in proteins?

3.2 C^α Shielding Tensors in Proteins

The $^{13}C^\alpha$ shielding tensors for all 20 naturally occurring amino acids (as N-formyl amino-acid amides) have now been computed for a wide range of ϕ,ψ, and χ_1 conformations.[51–53] An interesting result that emerges is that the overall breadth (or span) of the shielding tensor, $\Omega = \sigma_{33} - \sigma_{11}$ is, for most amino acids, rather constant ($\approx$ 34 ppm) between helical and sheet regions, but for the three β-branched amino acids, Val, Ile, and Thr, $\sigma_{33} - \sigma_{11}$ is only $\sim$22 ppm.[53] There is also a large difference in shielding tensor orientation between helical and sheet residues, which permits good predictions of the "CSA" values $\Delta\sigma^*$:

$$\Delta\sigma^* = \sigma_{\text{orth}} - \sigma_{\text{par}} \tag{11}$$

deduced from solution NMR,[54] where σ_{par} is the shielding in the direction parallel to the C–H bond vector and σ_{orth} is the average shielding orthogonal to this bond. On average in α-helices, $\Delta\sigma^* = 6.1 \pm 4.9$ ppm, whereas in β-sheets, $\Delta\sigma^* = 27.1 \pm 4.3$ ppm. This effect was predicted theoretically by Walling *et al.*,[55] who investigated "idealized" helical ($\phi = -60^\circ$, $\psi = -60^\circ$) and sheet ($\phi = -120^\circ$, $\psi = 120^\circ$) geometries for 18 amino acids. Then Sitkoff and Case made a direct comparison of the $^{13}C^\alpha$ CSA values for ubiquitin and calmodulin/M13 using an alanine fragment, and we made similar correlations for several other amino acids.[53]

But can CSA values (δ_{11}, δ_{22}, and δ_{33}) be accurately determined for individual sites in proteins? In early work, Cole *et al.*[56] reported the first magic-angle sample spinning NMR spectra of ^{13}C-labeled protein microcrystals and showed

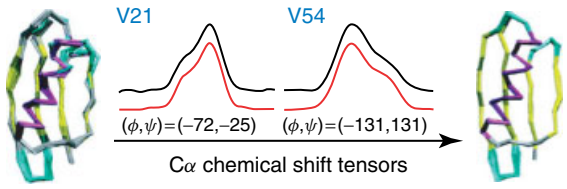

Figure 8 CSA lineshapes (black) and simulations (red) for C^α of V21 and V54 in GB1 together with (left) distance-based backbone structure and (right) CSA, TALOS, vector angle, and distance-restraint-based structure. (Reproduced from Ref. 60. © American Chemical Society, 2009)

that lines were narrow (as expected) and exhibited spinning sideband structure, opening up the possibility of determining of CSA in solid proteins using, e.g., the Herzfield–Berger method.[57] Indeed, this general approach has recently been applied to uniformly labeled proteins.[58] A second method to measure shielding tensors in proteins is to use the "recovery of CSA" or ROCSA method of Chan and Tycko,[59] which yields pseudostatic powder pattern lineshapes, from which δ_{ii} can be extracted.[60]

The agreement between the experimental shift tensor results and simulations is good, leading to the idea that it might be possible to use these CSA values in refinement. The shielding tensor elements σ_{ii} (ϕ,ψ) for each amino acid have now been incorporated into the Xplor-NIH program, and using a shielding tensor restraint, E_{CST}

$$E_{\text{CST}} = K_{\text{CST}} \Sigma (\delta_{ii} - \delta_{ii}{}^{\text{obs}})^2 \tag{12}$$

during simulated annealing, combined with isotropic chemical shift,[61] distance and vector angle restraints, the resulting refined structures exhibit root-mean-square errors with respect to existing X-ray structures of $\lesssim$1 Å (some of which must come from X-ray structural errors), and these structures show improved back-calculated shift and shift tensor properties, as well as no significant distance or other violations.[60] The structures show very tight "clustering", as can be seen in Figure 8, and can reasonably thought to be of improved accuracy and precision, due to the addition of the shielding tensor results.[60] Given the recent structure determination[46] of GB1 with just solid-state shifts (as experimental input), it will clearly be of interest to see what improvements are obtained by adding shift tensor information.

3.3 Aromatic Amino-Acid Shielding Tensors

Next, we return to the question, what causes the large range of chemical shift nonequivalence for the nonprotonated aromatic carbons in proteins? To investigate this question, we first obtained the solid-state NMR shifts of all nonprotonated aromatics in a series of Trp-containing peptides; C^γ, $C^{\delta 2}$, and $C^{\varepsilon 1}$ in His peptides; and C^γ in Phe and Tyr.[16,62,63] In none of these systems were ϕ,ψ effects important in shielding, unlike their role in C^α and C^β shielding. In the case of Trp residues, the dominant contributors to shielding were the side-chain torsion angles χ_1 and χ_2, and shifts could be predicted in peptides as well as in proteins over a $\sim$11 ppm shift range with a $\sim$1.4 ppm rms error.[62] When applied to 15 Trp C^γ in the carbon monoxyheme proteins reported earlier,[16] the use of homology

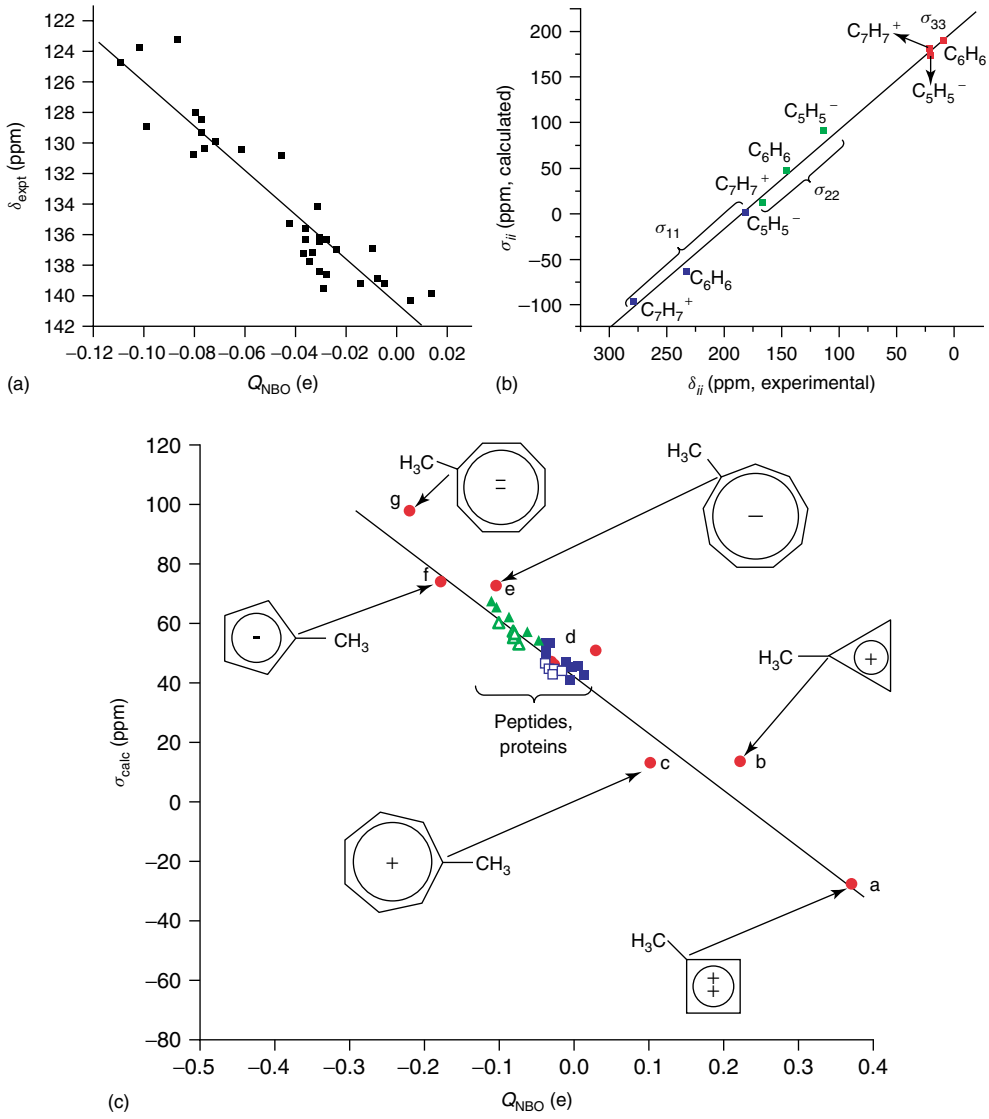

Figure 9 Correlations between shifts/shielding and NBO charge: (a) plot of experimental ^{13}C NMR shifts versus carbon NBO charges for C^{γ} of Phe and Tyr in peptides and proteins; (b) plot of the computed shielding tensor element magnitudes of aromatic systems ($C_7H_7^+$, C_6H_6, $C_5H_5^-$) versus the respective experimental values; and (c) plot of the calculated C^{γ} chemical shieldings versus C^{γ} NBO charges for Phe and Tyr residues in peptides and proteins and in various aromatic ring systems (green filled triangles, tyrosine dipeptides; green open triangles, tyrosine residues in proteins; blue filled squares, phenylalanine dipeptides; blue open squares, phenylalanine residues in protein). The slope of the correlation through all the data points is 191 ppm e^{-1}. (Reproduced from Ref. 64. © American Chemical Society, 2007)

models for these proteins yielded a good theory–experiment correlation for 13/15 residues, with two terminal Trps (Trp^3 in chicken AII hemoglobin, Trp^7 in myoglobin) being at about the random coil shift value.[62] In sharp contrast, results[63] for a series of His-containing peptides showed that C^{γ} and $C^{\delta 2}$ shifts were not only very large (12.7–13.8 ppm) but were also highly correlated and were shown to be dominated by ring tautomer (π, $N^{\delta 1}$–H and τ, $N^{\varepsilon 2}$–H) effects and intermolecular interactions. Similar effects were found in proteins, but only for buried residues. All shifts were predicted with an overall 1.6–1.9 ppm rms error over a 26 ppm range, but unlike, e.g., C^{α} shifts, it was essential to incorporate the effects of hydrogen bond partner molecules. For Phe and Tyr C^{γ}, there were no obvious correlations with ϕ, ψ, χ_1, or χ_2 torsion angles. However, good results ($R^2 = 0.94$, rmsd = 1.6 ppm, range = 17.1 ppm) were obtained by using a self-consistent reaction field continuum model.[64] This implies the importance of polarization effects, and indeed we found a linear relation between computed C^{γ} atomic charges and shifts for the 14 peptides, as well as in the 18 protein residues investigated (Figure 9). Remarkably, this result[64] is very similar to the correlation reported in the 1960s between ^{13}C shifts and π-electron densities in a variety of other classical, $4n + 2\,\pi$ electron systems.[65] The use of hydrogen bond partners, as well as the self-consistent reaction field approach, was essential to describe these effects, in which in essence electrostatic interactions between the aromatic ring and neighboring groups affect π-electron densities and heme C^{γ} atomic charges, with enhanced charges correlating with increased shielding—as seen in the other $4n + 2\,\pi$ systems.[65] Thus, ^{13}C shifts in proteins can be dominated by a large range of properties or interactions: ϕ, ψ, χ_1, χ_2, hydrogen bonding, electrostatics, tautomer state,

and atomic charges, but in all cases, one can now obtain good accord between theory and experiment with typically just one or two interactions dominating.

4 THROUGH-SPACE INTERACTIONS AND HYDROGEN BONDING

The ^{19}F NMR spectrum of [5-^{19}F]Trp-labeled galactose binding protein shows five peaks for the five labeled sites and no interactions between the sites. Remarkably, however, the spectra of [6-^{19}F]Trp-labeled dihydrofolate reductase (DHFR) reported by Kimber *et al*.[66] showed a field-independent splitting (in Hertz) between two [6-^{19}F] Trp residues in a *Lactobacillus casei* DHFR-NADPH-methotrexate complex. This coupling could either be $^{398}J_{FF}$ or a direct through-space J-coupling. As proposed[66] and as later revealed crystallographically, two Trp (Trp5 and Trp133) are indeed close (Figure 10), and on the basis of previous observations of large but "long-range" J-couplings in compounds such as a difluorophenanthrene ($^5J \sim 170$ Hz), it is clear that these types of interactions have to be through space and not through bonds. If correct, it should, therefore, be possible to compute such long-range couplings (which actually are short range) by using quantum chemistry since, as with the chemical shift, the J-coupling is just another second-rank tensor:

$$J_{\alpha\beta} = \frac{\partial^2 E}{\partial \mu_1 \partial \mu_2} \tag{13}$$

where μ_i are the magnetic moments. To test whether such long-range, non-bonded interactions might be computed, we used density functional theory to compute J_{FF} for a variety of small organic molecules containing spatially close ^{19}F nuclei, separated, however, by numerous bonds. We used several computational models, a particularly simple one being $(CH_3F)_2$, a fluoromethane dimer. The results (Figure 10b) were surprisingly good ($R^2 = 0.90$, slope $= 1.1$, $N = 15$) and, as shown in Figure 10a, reveal that J_{FF} decreases exponentially with the F–F internuclear separation. Moreover, the Fermi contact interaction makes the overwhelming contribution to the computed J_{FF}.[67]

What is of particular interest about these results is that there is no real bond between the ^{19}F nuclei, just a penetration of the electron clouds that enables Fermi contact between the two fluorine nuclei. Indeed, such an F–F interaction (in, say, the CH_3F dimer) would likely be antibonding. This leads to the question—are long-range, nonbonded J-couplings seen in other systems indicative of bonding? To avoid, as much as possible, purely linguistic or semantic debates as to what is a bond, what is a covalent bond, and so forth, we used[68] the language of Bader's AIM (atoms in molecules) theory.[69] When investigating long-range (or trans) hydrogen bonds, we found[68] the following: local energy densities (computed by using AIM theory) in hydrogen bonds were correlated with several NMR observables. In particular, the chemical nature of these bonds correlated with the magnitudes of the trans hydrogen bond scalar couplings and the proton chemical shifts. The AIM results also indicated that the $^{3h}J_{NC'}$ couplings observed in proteins are mediated by closed-shell, noncovalent NH$\cdots$OC interactions. The inductive mechanism that allows the nitrogen and carbon nuclei to couple is provided by a mutual penetration of isolated donor and acceptor nonbonding van der Waals charge densities. The AIM results also showed that the same mechanism explains the through-space scalar couplings observed between fluorine nuclei in organic molecules, and in proteins, with the magnitudes of both $^{3h}J_{NC'}$ and J_{FF} depending on a very similar exponential function of the van der Waals penetration.

Hydrogen bond ^{1}H chemical shifts describe a wide range of chemical interactions, from the closed-shell limit in protein backbone hydrogen bonds to the partially covalent in enzymatic short, strong, hydrogen bonds (SSHBs), to the genuine shared-electron in low-barrier hydrogen bonds. As the proton resonates to lower and lower field, there is a smooth, exponential increase in the degree of covalence in the hydrogen bond, until at about 20–21 ppm when the bond becomes a genuine shared-electron (covalent) interaction. The AIM results also show that the further downfield the ^{1}H resonates, the greater is the covalent character. Since it has already been demonstrated that $\delta(^1H)$ increases exponentially with decrease in the hydrogen bond distance (OH$\cdots$O), and since enzymatic SSHBs are found at the steep ends of the energy density exponentials, small perturbations in donor–acceptor distances have very large energetic consequences. In this sense,

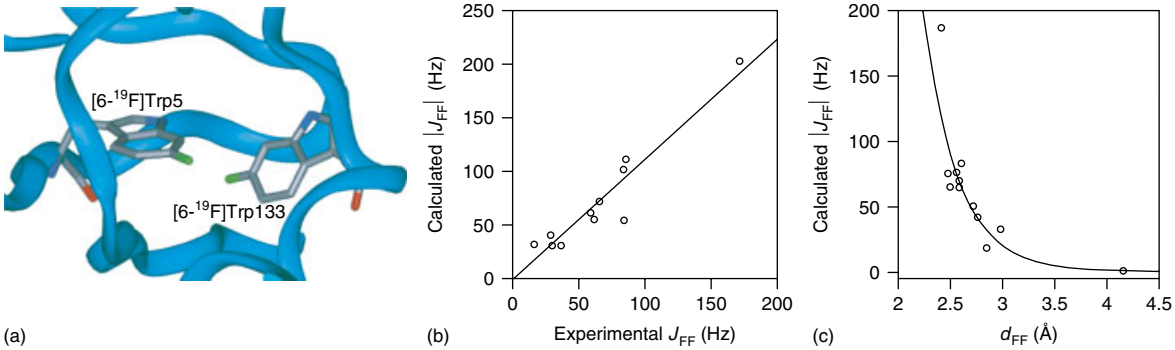

Figure 10 Long-range ^{19}F–^{19}F J-coupling. (a) The proximity of the [6-F]Trp5 and [6-F]Trp133 side chains in the protein [6-F]Trp DHFR complexed with MTX and NADPH. The fluorine nuclei (green) are approximately 3 Å apart. (b) Calculated versus experimental J_{FF} in several organic molecules and in the protein DHFR calculated using only fluoromethane dimer models (slope $= 1.13$ and $R^2 = 0.90$). (c) The exponential distance dependence of the calculated J-couplings, $|J_{FF}|$, versus the fluorine–fluorine internuclear separation, d_{FF} ($R^2 = 0.96$). (Reproduced from Ref. 67. © American Chemical Society, 2000)

it is a relatively short trip from the partially covalent SSHBs at 15–20 ppm to the shared-electron, single-well LBHB at 21 ppm.

5 METALS AND METALLOPROTEINS

5.1 NMR of Metals

Many proteins contain metal centers and in some cases these can have very large effects on protein shifts. Metal ion NMR itself is also, in principle, possible and, indeed, in the history of the nature or theory of chemical shifts, metal ion NMR played an important role. For example, Proctor and Yu in 1951 found that ^{59}Co NMR chemical shifts varied greatly (>10 000 ppm) from compound to compound,[70] an observation noted by Ramsey,[71] and a very important correlation between chemical shifts and crystal field splittings was made over 50 year ago by Freeman *et al*.[72] However, early attempts to accurately predict ^{59}Co shifts in diamagnetic CoIII (d^6) complexes using Hartree–Fock or density functional theory (DFT) failed, as did attempts at computing ^{57}Fe NMR shifts. However, when hybrid DFT functionals were used, Bühl found that both ^{57}Fe and ^{103}Rh shifts in small organometallics could, in fact, be well predicted.[73] This led to the idea that it should be possible to calculate ^{59}Co shifts, which turned out to be the case.[74] However, the large quadrupole moment of ^{59}Co makes protein studies difficult—but not so with ^{57}Fe, where in early work we and others had observed ^{57}Fe NMR chemical shifts in several different diamagnetic heme proteins.[75]

To compute ^{57}Fe shieldings in metalloproteins, large metalloporphyrin models such as a CO, methyl imidazole (1-MeIm) ligated iron tetraphenylporphyrin (TPP) are needed in order to encompass most of the key structural features found in a CO heme protein, such as carbonmonoxymyoglobin, with other models (such as a dimethyl sulfide porphyrin adduct for cytochrome *c*) being constructed computationally, based (in general) on protein X-ray structures. The ^{57}Fe shift predictions were good, as were the tensor spans ($\Omega = \sigma_{33} - \sigma_{11}$), and when the isotropic shifts were combined with Bühl's results, it yielded a theory versus experiment $R^2 = 0.992$.[76] But there was a big problem with carbonomonomyoglobin. Specifically, when a Fe(TPP)(1-MeIm)(CO) model for carbonmonoxy myoglobin (MbCO) was used for shift calculations, good results were obtained, but when using reported protein X-ray structures, there were errors of approximately 5000–7000 ppm. There were several possible explanations, including that the X-ray was wrong, the NMR shift was wrong, or the calculation was wrong (and in principle, any combination of the above). Since the same ^{57}Fe shift had been seen by several groups and the small molecule structure gave a reasonable answer, the X-ray seemed wrong. However, the question was, how can one test this hypothesis? One approach was to ask the following questions: how well can one predict other properties using this small molecule model structure? And can we do better with the protein X-ray structure? But what other properties should we calculate for ^{57}Fe? Obviously, ^{57}Fe Mössbauer isomer shifts and quadrupole splittings, both of which are (in principle) easy to compute ground, state properties.

5.2 Combining NMR and Mössbauer Spectroscopy: Chemical/Isomer Shifts and EFGS

Surprisingly, even after many decades of experimental work in the Mössbauer community, there had been no computational studies of the Mössbauer properties of large molecular systems such as MbCO, with only a handful of inorganics ($K_4[Fe(CN)_6]$, $K_3[Fe(CN)_6]$, $KFeF_3$, and $BaFeO_4$) being investigated. Although Mossbauer calculations might seem off-topic in this article, they are really not since the actual properties are similar or identical to those of interest in NMR: the EFG and charge densities. Moreover, of more practical interest, there is essentially no additional cost in such calculations if the NMR chemical (or hyperfine) shifts have already been computed, since the wavefunctions are already available.

^{57}Fe Mössbauer spectra are typically dominated by two interactions: the quadrupole splitting, which arises from the nonspherical nuclear charge distribution in the $I^* = 3/2$ excited state in the presence of an EFG at the ^{57}Fe nucleus, and an isomer (or chemical) shift, which arises from differences in the electron density at the nucleus between the absorber (the molecule or the system of interest) and a reference compound (usually α-Fe at 300 K). This interaction is given as:

$$\delta_{\mathrm{Fe}} = E_A - E_{\mathrm{Fe}} = \frac{2\pi}{3} Z e^2 \left(\left\langle R^2 \right\rangle^* - \left\langle R^2 \right\rangle \right) \left(|\psi(0)|^2_{\mathrm{Fe}} \right) \quad (14)$$

where Z represents the atomic number of the nucleus of interest (iron) and R, R^* are average nuclear radii of the ground and excited states of ^{57}Fe. Since $|\psi(0)|^2_{\mathrm{Fe}}$ is a constant, the isomer shift (from Fe) can be written as

$$\delta_{\mathrm{Fe}} = \alpha[\rho^{\mathrm{tot}}(0) - c] \quad (15)$$

where α is the so-called calibration constant and $\rho^{\mathrm{tot}}(0)$ is the computed charge density at the nucleus. The second property is the Mössbauer quadrupole splitting (and its asymmetry parameter), which is related to the components of the EFG tensor at the nucleus as follows:

$$\Delta E_Q = \frac{1}{2} e Q V_{zz} \left(1 + \frac{\eta^2}{3} \right)^{1/2} \quad (16)$$

where e is the electron charge, Q is the quadrupole moment of the $E^* = 14.4$ keV excited state, and the principal components of the EFG tensor are labeled according to the following convention:

$$|V_{zz}| > |V_{yy}| > |V_{xx}| \quad (17)$$

with the asymmetry parameter, η, being defined as:

$$\eta = \frac{V_{xx} - V_{yy}}{V_{zz}} \quad (18)$$

Thus, the Mössbauer quadrupole splitting is just the EFG ($e^2 q Q/h$) we are familiar with in, e.g., ^{17}O NMR, the asymmetry parameter is also the same, and the isomer shift (sometimes also called the *chemical shift*) just scales as the charge density at the nucleus. Since there is no mixing-in of excited states, it should be possible to compute both δ_{Fe} as

well as ΔE_Q and η. Indeed, the V_{ii} components are readily available from a chemical (or hyperfine) shift calculation using the Gaussian program, whereas for the charge density, we used the Aim2000 program.[77–79] This functionality is now available in the latest Gaussian releases.

From the results of the Mössbauer quadrupole splitting (EFG) calculations, it was immediately apparent that the small molecule Fe(TPP)(1-MeIm)(CO) structure yielded an accurate prediction for the MbCO ΔE_Q (0.44 mm s^{-1} predicted, 0.36–0.37 mm s^{-1} experimental), to be compared with calculations based on the protein X-ray structures ($\Delta E_Q = 1.89$–2.38 mm s^{-1}), the main difference between the protein and the model compound structures being the fact that the former had a bent and tilted FeCO subunit, which was essentially linear and not tilted in the model metalloporphyrin.[79] We then computed δ_{Fe} and ΔE_Q for many different diamagnetic and paramagnetic systems[77,78]—but that is another story.

The question is, what are the actual geometries in MbCO? And, can we combine NMR results with IR, Mössbauer isomer shifts, quadrupole splittings, and ^{17}O EFGs—all of the properties we have discussed above—using, for example, the Z-surface approach, to get at structure? To do this, it is necessary to calculate a series of property surfaces, not $P(\phi, \psi)$ as in the study of amino-acids, but $P(\tau, \beta)$, where τ and β are the Fe–C–O tilt and bend angles, respectively. We performed such calculations for $\delta_i(^{13}C)$, $\Delta\sigma(^{13}C)$, $\delta_i(^{17}O)$, e^2qQ/h (^{17}O), $\delta_i(^{57}Fe)$, and ΔE_Q (^{57}Fe) and used the Z-surface approach to predict the most probable values of τ,β.[80] The results of the ^6Z surface prediction yielded $\tau = 4^\circ$, $\beta = 7^\circ$[80] and about a year after this work was published, a group of X-ray crystallographers[81] redetermined the MbCO structure, concluding that the $\sim 140^\circ$ combined tilt/bend reported earlier was most likely due to sample decomposition and detection of an off-axis water (not a CO). Moreover, the actual τ,β values they reported[81] were $<1^\circ$ different to those predicted by using the NMR/Mössbauer QM analysis.[80] This X-ray reanalysis was challenged by another group, with a later tie-breaking X-ray reanalysis by yet another group,[82] who agreed with the Z-surface result. Likewise, we and others found poor accord between structures determined by using crystallography with those deduced by using spectroscopy and quantum chemistry in several other systems, for example, in the case of NO binding to heme proteins,[83] so it seems that there has been a tendency to overestimate the accuracy of some metalloprotein crystallographic structures.

5.3 Paramagnetic Proteins and Model Systems

Finally, we need to consider paramagnetic systems. The observed chemical shift in paramagnetic species can be described as the sum of two terms:

$$\delta_{obs} = \delta_{dia} + \delta_{hf} \tag{19}$$

where δ_{obs}, δ_{dia}, and δ_{hf} are the observed, diamagnetic, and paramagnetic (or hyperfine) shifts, respectively. δ_{dia} is the shift that would be observed if the molecule contained no unpaired electrons, the conventional orbital or chemical shift, and is generally taken to be the shift value found for a suitable diamagnetic reference compound. The hyperfine shift, δ_{hf}, can itself be divided into two terms:

$$\delta_{hf} = \delta_{con} + \delta_{dip} \tag{20}$$

where δ_{con} is the contact shift and δ_{dip} is the dipolar or pseudocontact shift. The contact shift arises from spin delocalization from the unpaired electrons, usually located on a metal atom, to the nuclei at the periphery of the molecule through chemical bonds, and is directly proportional to the Fermi contact spin density, $\rho_{\alpha\beta}$, at the nucleus of interest:

$$\delta_{con} = \frac{\mu_o \mu_B^2 g_e^2 (S+1)}{9kT} \rho_{\alpha\beta} \tag{21}$$

Here, μ_0 is the vacuum permeability ($4\pi \times 10^{-7}$ NA^{-2}), μ_B is the Bohr magneton (9.274×10^{-24} JT^{-1}), g_e is the free electron g-factor, S is the total spin, k is Boltzmann's constant, and T is the absolute temperature. The Fermi contact spin density, $\rho_{\alpha\beta}$, represents the net imbalance between α and β spins at the position of the nucleus of interest. Positive spin density corresponds to a downfield contact shift, whereas negative spin density corresponds to an upfield contact shift. For heavy atoms (such as ^{13}C and ^{15}N) in metalloporphyrin complexes, the contact shift can be very large, hundreds or thousands of parts per million, and it typically overwhelmingly dominates the dipolar or pseudocontact shift. As a result, the relationship between the hyperfine shift and the Fermi contact spin density can be expressed in a simplified form:

$$\delta_{hf} \approx \delta_{con} = m \frac{S+1}{T} \rho_{\alpha\beta} \tag{22}$$

where m is a collection of physical constants:

$$m = \frac{\mu_o \mu_B^2 g_e^2}{9k} = 23.5 \times 10^6 \text{ppm K au}^{-1} \tag{23}$$

The first predictions of hyperfine shifts in proteins were carried out by the groups of Weinhold and Markley [84] on rubredoxin, an Fe/S cluster protein, and there was a good prediction of the overall chemical shift range expected. We later investigated a broader range of systems, having $S = 1/2$, $S = 2$, and $S = 5/2$ spin states, the states that are most frequently found in heme proteins. Using density functional theory (DFT) methods, the >3500 ppm range in experimentally observed hyperfine shifts was well predicted. Using spin-unrestricted methods together with large, locally dense basis sets, we obtained[85] very good correlations between experimental and theoretical results: $R^2 = 0.941$ ($N = 37$, $p < 0.0001$) when using the pure BPW91 functional and $R^2 = 0.981$ ($N = 37$, $p < 0.0001$) when using the hybrid functional, B3LYP. The correlations were even better for C^α and C^β shifts alone: C^α, $R^2 = 0.996$ ($N = 8$, $p < 0.0001$, B3LYP); C^β, $R^2 = 0.995$ ($N = 8$, $p < 0.0001$, B3LYP), but were worse for C^{meso}, in part due to the small range in C^{meso} shifts. With paramagnetic metalloproteins, linewidths in solution can be rather problematic, due to the dominance, in many cases, of Curie relaxation. That is, the linewidth increases significantly with increase in the correlation time, and linewidths can be on the order of 0.5 MHz—but if shifts are measured, they can still be calculated with moderate accuracy.[86] It would be desirable

to remove this Curie relaxation, and the way to do that is just to obtain solid-state NMR spectra, since then there are essentially no linewidth contributions arising from relaxation mechanism.

Early studies of paramagnetic solids, such as Cu^{II} (DL-alanine)$_2$·H_2O, using magic-angle sample spinning showed that all three carbons–C′, C^{α}, and C^{β}–were well resolved and could be assigned,[87,88] but initial attempts in computing these shifts were not successful. The reason for this is that Cu^{II} (DL-alanine)$_2$·H_2O is not a simple, discrete molecular entity. Rather, the crystallographic structure[89] reveals that Cu(DL-alanine)$_2$·H_2O is not the five-coordinate complex anticipated. Instead, it is a 1-D polymer with water–Cu_{∞} chains (d_{Cu-O} = 2.653 Å) zigzagging along the crystallographic *c*-axis, and these chains are further connected by a large 3-D hydrogen bond network. The central unit (H_2O–Cu–H_2O) has 14 hydrogen bonds, with eight neighboring Cu(DL-alanine)$_2$ and two water molecules. These structural features have significant effects on the hyperfine interactions, and necessitated the use of a 9-copper containing "supermolecule" in order to adequately describe the system's electronic structure, and hyperfine shifts. Once this was done, however, the computational results[89] provided excellent predictions of the ^{13}C and ^{1}H shifts for this and a second paramagnetic system, $V(acac)_3$, together with accurate single-crystal ENDOR hyperfine tensor results:[89]

$$A_i = A_{iso} + T_{ii} \tag{24}$$

as well as the ^{1}H hyperfine shifts seen in single-crystal NMR studies of Cu(DL-alanine)$_2$·H_2O. These NMR shifts (ΔH_i) are related to the hyperfine couplings $A_i (i = 1,2,3)$ as

$$\Delta H_i = -A_i S_c / \hbar \gamma_H \tag{25}$$

where S_c is the Curie spin, $\hbar$ is Plank's constant divided by 2π, and γ_H is the ^{1}H gyromagnetic ratio. A plot of A_i^{calcd} versus H_i^{expt} showed an excellent correlation between theory and experiment ($R^2 = 0.961$) with an intercept of −1.05 MHz (4.3% of the range of 24.20 MHz) and a slope of 0.00537 MHz G^{-1}. The expected value is 0.00426 MHz G^{-1}. The origins of the discrepancy are not certain, but since the ENDOR and MAS NMR results have <1% errors, it could simply be due to difficulties in accurately measuring sample temperatures (e.g., a 1 K error in T would result in a 20–30% error in S_c since $T - T_c = 4.7$ K).[89]

6 CONCLUSIONS AND PROSPECTS

It is now possible to calculate, with good accuracy, most NMR observables in a protein–isotropic chemical shifts, chemical shift tensor elements and their orientations, and EFG tensors–and to use these in structure refinement. These studies then led to investigations of paramagnetic species where hyperfine shifts, ESR, and ENDOR spectra can be predicted. In addition, the output of such calculations readily led to the interpretation of Mössbauer spectra, which are functions of the EFG at the nucleus and charge densities. Long-range, non-bounded interactions, such as ^{19}F–^{19}F J-couplings, are also accessible, again with little added computational expense. In brief, essentially all spectroscopic observables can now be predicted and, consequently, can be used in structure prediction and refinement. Given the success in de novo structure determination using isotropic chemical shifts,[44,46] it seems likely that using solely shift tensors (and tensor orientations) will be a particularly fruitful area for future development, applicable in particular to noncrystalline proteins. However, even when crystallographic structures are available, the use of QM methods, such as those described here and elsewhere[90] to improve structure quality, is also clearly an important area for development, since the spectroscopic observables are extremely sensitive to local (bond length, bond angle, and torsion angle) structure. Thus, if a really accurate structure is of interest, combining crystallography with spectroscopy is highly desirable, and can be expected to lead to, e.g., a better understanding of the mechanism of action of some enzymes, and how they are inhibited, of interest in drug discovery. While it is not essential to have an ultrahigh-resolution structure (or indeed any structure) to discover a new drug, it seems logical that any structure-based design approach would benefit from the best possible structures. Here, NMR can offer unique information complementary to X-ray crystallography on, e.g., the protonation states of different ligand groups when bound to a target protein. In addition, interpreting NMR shift changes on ligand binding in terms of protein structural charge clearly warrants further development (e.g., using shielding surfaces) as a rapid means of determining target structural changes.

7 REFERENCES

1. M. Saunders and A. Wishnia, *Ann. N. Y. Acad. Sci.*, 1958, **70**, 870.
2. O. Jardetzky and C. D. Jardetzky, *J. Biol. Chem.*, 1958, **233**, 383.
3. A. Kowalsky, *J. Biol. Chem.*, 1962, **237**, 1807.
4. C. C. McDonald and W. D. Phillips, *J. Am. Chem. Soc.*, 1967, **89**, 6332.
5. K. Wuthrich, *Proc. Natl. Acad. Sci. U S A*, 1969, **63**, 1071.
6. R. G. Shulman, et al. *Science*, 1969, **165**, 251.
7. J. D. Roberts, F. J. Weigert, J. I. Kroschwitz, and H. J. Reich, *J. Am. Chem. Soc.*, 1970, **92**, 1338.
8. W. J. Horsley and H. Sternicht, *J. Am. Chem. Soc.*, 1968, **90**, 3738.
9. A. J. Jones, M. W. Winkley, D. M. Grant, and R. K. Robins, *Proc. Natl. Acad. Sci. U S A*, 1970, **65**, 27.
10. E. G. Paul and D. M. Grant, *J. Am. Chem. Soc.*, 1964, **86**, 2977.
11. R. R. Ernst and W. A. Anderson, *Rev. Sci. Instrum.*, 1966, **37**, 93.
12. K. F. Kuhlmann, D. M. Grant, and R. K. Harris, *J. Chem. Phys.*, 1970, **52**, 3439.
13. D. Doddrell, V. Glushko, and A. Allerhand, *J. Chem. Phys.*, 1972, **56**, 3683.
14. I. Solomon, *Phys. Rev.*, 1955, **99**, 559.
15. A. Allerhand and E. Oldfield, *Biochemistry*, 1973, **12**, 3428.
16. E. Oldfield and A. Allerhand, *J. Biol. Chem.*, 1975, **250**, 6403.
17. E. Wenkert, A. O. Clouse, D. W. Cochran, and D. Doddrell, *J. Am. Chem. Soc.*, 1969, **91**, 6879.
18. A. Allerhand, R. F. Childers, and E. Oldfield, *J. Magn. Reson.*, 1973, **11**, 272.
19. A. Allerhand, R. F. Childers, R. A. Goodman, E. Oldfield, and X. Ysern, *Am. Lab.*, 1972, **4**, 19.
20. A. Allerhand, R. F. Childers, and E. Oldfield, *Biochemistry*, 1973, **12**, 1335.

21. E. Oldfield and M. Meadows, *J. Magn. Reson.*, 1978, **31**, 327.
22. E. Oldfield and A. Allerhand, *Proc. Natl. Acad. Sci. U S A*, 1973, **70**, 3531.
23. R. K. Gupta and A. G. Redfield, *Science*, 1970, **169**, 1204.
24. E. Oldfield and R. J. Kirkpatrick, *Science*, 1985, **227**, 1537.
25. K. D. Park, et al. *Biochemistry*, 1991, **30**, 2333.
26. J. D. Augspurger, C. E. Dykstra, and E. Oldfield, *J. Am. Chem. Soc.*, 1991, **113**, 2447.
27. J. Augspurger, et al. *J. Magn. Reson.*, 1992, **100**, 342.
28. J. D. Augspurger, A. C. de Dios, E. Oldfield, and C. E. Dykstra, *Chem. Phys. Lett.*, 1993, **213**, 211.
29. A. C. de Dios and E. Oldfield, *Chem. Phys. Lett.*, 1993, **205**, 108.
30. B. D. Sykes and W. E. Hull, *Meth. Enzymol.*, 1978, **49**, 270.
31. J. G. Pearson, E. Oldfield, F. S. Lee, and A. Warshel, *J. Am. Chem. Soc.*, 1993, **115**, 6851.
32. L. A. Luck and J. J. Falke, *Biochemistry*, 1991, **30**, 4248.
33. S. Spera and A. Bax, *J. Am. Chem. Soc.*, 1991, **113**, 5490.
34. J. F. Hinton, P. Guthrie, P. Pulay, and K. Wolinski, *J. Am. Chem. Soc.*, 1992, **114**, 1604.
35. A. C. de Dios, J. G. Pearson, and E. Oldfield, *Science*, 1993, **260**, 1491.
36. D. D. Laws, A. C. de Dios, and E. Oldfield, *J. Biomol. NMR*, 1993, **3**, 607.
37. A. C. de Dios, J. G. Pearson, and E. Oldfield, *J. Am. Chem. Soc.*, 1993, **115**, 9768.
38. J. G. Pearson, et al. *J. Am. Chem. Soc.*, 1997, **119**, 11941.
39. D. D. Laws, H. Le, A. C. de Dios, R. H. Havlin, and E. Oldfield, *J. Am. Chem. Soc.*, 1995, **117**, 9542.
40. A. C. de Dios and E. Oldfield, *J. Am. Chem. Soc.*, 1994, **116**, 5307.
41. H. Le, J. G. Pearson, A. C. de Dios, and E. Oldfield, *J. Am. Chem. Soc.*, 1995, **117**, 3800.
42. J. G. Pearson, J.-F. Wang, J. L. Markley, H. Le, and E. Oldfield, *J. Am. Chem. Soc.*, 1995, **117**, 8823.
43. E. Oldfield, *J. Biomol. NMR*, 1995, **5**, 217.
44. A. Cavalli, X. Salvatella, C. M. Dobson, and M. Vendruscolo, *Proc. Natl. Acad. Sci. U S A*, 2007, **104**, 9615.
45. Y. Shen, et al. *Proc. Natl. Acad. Sci. U S A*, 2008, **105**, 4685.
46. P. Robustelli, A. Cavalli, and M. Vendruscolo, *Structure*, 2008, **16**, 1764.
47. N. Janes, S. Ganapathy, and E. Oldfield, *J. Magn. Reson.*, 1983, **54**, 111.
48. A. C. de Dios, D. D. Laws, and E. Oldfield, *J. Am. Chem. Soc.*, 1994, **116**, 7784.
49. J. Heller, et al. *J. Am. Chem. Soc.*, 1997, **119**, 7827.
50. S. Wi, H. Sun, E. Oldfield, and M. Hong, *J. Am. Chem. Soc.*, 2005, **127**, 6451.
51. R. H. Havlin, H. Le, D. D. Laws, A. C. de Dios, and E. Oldfield, *J. Am. Chem. Soc.*, 1997, **119**, 11951.
52. R. H. Havlin, et al. *J. Am. Chem. Soc.*, 2001, **123**, 10362.
53. H. Sun, L. K. Sanders, and E. Oldfield, *J. Am. Chem. Soc.*, 2002, **124**, 5486.
54. N. Tjandra and A. Bax, *J. Am. Chem. Soc.*, 1997, **119**, 9576.
55. A. E. Walling, R. E. Pargas, and A. C. de Dios, *J. Phys. Chem. A*, 1997, **101**, 7299.
56. H. B. R. Cole, S. W. Sparks, and D. A. Torchia, *Proc. Natl. Acad. Sci. U S A*, 1988, **85**, 6362.
57. J. Herzfeld and A. E. Berger, *J. Chem. Phys.*, 1980, **73**, 6021.
58. B. J. Wylie, et al. *J. Am. Chem. Soc.*, 2007, **129**, 5318.
59. J. C. C. Chan and R. Tycko, *J. Chem. Phys.*, 2003, **118**, 8378.
60. B. J. Wylie, C. D. Schwieters, E. Oldfield, and C. M. Rienstra, *J. Am. Chem. Soc.*, 2009, **131**, 985.
61. G. Cornilescu, F. Delaglio, and A. Bax, *J. Biomol. NMR*, 1999, **13**, 289.
62. H. Sun and E. Oldfield, *J. Am. Chem. Soc.*, 2004, **126**, 4726.
63. F. Cheng, H. Sun, Y. Zhang, D. Mukkamala, and E. Oldfield, *J. Am. Chem. Soc.*, 2005, **127**, 12544.
64. D. Mukkamala, Y. Zhang, and E. Oldfield, *J. Am. Chem. Soc.*, 2007, **129**, 7385.
65. H. Spiesecke and W. G. Schneider, *Tetrahedron Lett.*, 1961, **2**, 468.
66. B. J. Kimber, et al. *Nature*, 1978, **271**, 184.
67. W. D. Arnold, J. Mao, H. Sun, and E. Oldfield, *J. Am. Chem. Soc.*, 2000, **122**, 12164.
68. W. D. Arnold and E. Oldfield, *J. Am. Chem. Soc.*, 2000, **122**, 12835.
69. R. F. W. Bader, *Atoms in Molecules–A Quantum Theory*, Oxford University Press, 1990,
70. W. G. Proctor and F. C. Yu, *Phys. Rev.*, 1951, **81**, 20.
71. N. F. Ramsey, *Phys. Rev.*, 1952, **86**, 243.
72. R. Freeman, G. R. Murray, and R. E. Richards, *Proc. Roy. Soc. London A*, 1957, **242**, 455.
73. M. Bühl, *Chem. Phys. Lett.*, 1997, **267**, 251.
74. N. Godbout and E. Oldfield, *J. Am. Chem. Soc.*, 1997, **119**, 8065.
75. J. Chung, H. C. Lee, and E. Oldfield, *J. Magn. Reson.*, 1990, **90**, 148.
76. N. Godbout, R. H. Havlin, R. Salzmann, P. G. Debrunner, and E. Oldfield, *J. Phys. Chem. A*, 1998, **102**, 2342.
77. Y. Zhang, J. Mao, and E. Oldfield, *J. Am. Chem. Soc.*, 2002, **124**, 7829.
78. Y. Zhang, J. Mao, N. Godbout, and E. Oldfield, *J. Am. Chem. Soc.*, 2002, **124**, 13921.
79. R. H. Havlin, et al. *J. Am. Chem. Soc.*, 1998, **120**, 3144.
80. M. McMahon, et al. *J. Am. Chem. Soc.*, 1998, **120**, 4784.
81. G. S. Kachalova, A. N. Popov, and H. D. Bartunik, *Science*, 1999, **284**, 473.
82. B. Stec and G. N. Phillips Jr., *Acta Crystallogr. D Biol. Crystallogr.*, 2001, **57**, 751.
83. Y. Zhang, W. Gossman, and E. Oldfield, *J. Am. Chem. Soc.*, 2003, **125**, 16387.
84. S. J. Wilkens, B. Xia, F. Weinhold, J. L. Markley, and W. M. Westler, *J. Am. Chem. Soc.*, 1998, **120**, 4806.
85. J. Mao, Y. Zhang, and E. Oldfield, *J. Am. Chem. Soc.*, 2002, **124**, 13911.
86. Y. Zhang and E. Oldfield, *J. Am. Chem. Soc.*, 2008, **130**, 3814.
87. K. Liu, D. Ryan, K. Nakanishi, and A. McDermott, *J. Am. Chem. Soc.*, 1995, **117**, 6897.
88. Y. Ishii, N. P. Wickramasinghe, and S. Chimon, *J. Am. Chem. Soc.*, 2003, **125**, 3438.
89. Y. Zhang, H. Sun, and E. Oldfield, *J. Am. Chem. Soc.*, 2005, **127**, 3652.
90. J. A. Vila, et al. *Proc. Natl. Acad. Sci. U S A*, 2008, **105**, 14389.

Acknowledgments

I would like to thank Adam Allerhand, Cliff Dykstra, Angel de Dios, Peter Pulay, Nathalie Godbout, Yong Zhang, and Cynthia Jameson for their diverse contributions to the work, which was supported by the United States Public Health Service (NIH grants GM65307 and GM073216).

Biographical Sketch

Eric Oldfield was educated at Bec School, London, and received a B.Sc. in Chemistry from Bristol University, doing research with Jake Macmillan and Geoffrey Eglinton, and a Ph.D. in Biophysical Chemistry from Sheffield University, with Dennis Chapman. He then worked as an EMBO Fellow at Indiana University with Adam Allerhand and was a Visiting Scientist at MIT with John Waugh. He joined the Chemistry Department at the University of Illinois at Urbana-Champaign in 1975 and is currently the Alumni Research Scholar Professor of Chemistry, a Professor of Biophysics in the Center for Computational Biology and Biophysics, and a member of the Center for Zoonoses Research. His research interests have covered a broad range of topics, including ^{2}H NMR of lipids and proteins, ^{13}C NMR of proteins, NMR of nonintegral spin quadrupolar nuclei, electrochemical NMR, and quantum chemical investigations of NMR, EPR, and Mössbauer spectra in metalloproteins and model systems. Most recently, he has developed new approaches to the treatment of infectious diseases targeting bacterial virulence factors, the development of novel antiparasitics for tropical diseases, as well as the development of new lipophilic bisphosphonates for use in cancer therapy and immunotherapy. He has been the recipient of several awards for his work, including RSC's Meldola Medal and Prize, RSC's Spectroscopy Award, RSC's Award in Soft Matter and Biophysical Chemistry, The Biochemical Society's Colworth Medal, ACS's Award in Pure Chemistry, and the American Heart Association's Basic Science Research Prize.