Annu. Rev. Phys. Chem. 2002. 53:349–78
DOI: 10.1146/annurev.physchem.53.082201.124235

CHEMICAL SHIFTS IN AMINO ACIDS, PEPTIDES, AND PROTEINS: From Quantum Chemistry to Drug Design

Eric Oldfield
Department of Chemistry and Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801; e-mail: eo@chad.scs.uiuc.edu

Key Words electrostatics, porphyrins

■ **Abstract** This chapter discusses recent progress in the investigation and use of ^{13}C, ^{15}N, and ^{19}F nuclear magnetic resonance (NMR) chemical shifts and chemical shift tensors in proteins and model systems primarily using quantum chemical (ab initio Hartree-Fock and density functional theory) techniques. Correlations between spectra and structure are made and the techniques applied to other spectroscopic and electrostatic properties as well, including hydrogen bonding, ligand binding to heme proteins, J-couplings, electric field gradients, and atoms-in-molecules theory, together with a brief review of the use of NMR chemical shifts in drug design.

INTRODUCTION

It has been known for many years that folding a protein into its native conformation causes a large range of chemical shift nonequivalence. For reasons of sensitivity, the earliest protein nuclear magnetic resonance (NMR) investigations utilized the ^{1}H nucleus, but there were few resolved single atom resonances, typically just those of H^{ε_2} of histidine residues, which fall outside the main protein spectral region (1). With the advent of Fourier transform NMR techniques, there was considerable interest in using ^{13}C NMR spectroscopy to investigate chemical shifts in proteins because it was thought that there would be far more resolved, single atom resonances owing to the well-known large chemical shift range of ^{13}C. Unfortunately, the earliest investigations were unsuccessful, because of low signal-to-noise ratios. However, in 1972 Allerhand developed a 20-mm NMR probe that enabled the first observation of single carbon atom resonances in proteins (2). The results were of interest because they showed a very large range of chemical shift nonequivalence due to folding, for example, up to 6 ppm in the case of C^{γ} of tryptophan residues (2). In addition, it was demonstrated that there were significant chemical shift

differences between C^{α} chemical shifts in random coil versus helical polypeptides (3). These early observations suggested that ^{13}C chemical shifts might be useful probes of protein structure, but it took another decade of development in the area of heteronuclear NMR, together with isotopic labeling, to enable the routine measurement of ^{13}C (and ^{15}N) chemical shifts in proteins. Once this had been accomplished, it became clear that ^{13}C chemical shifts in proteins were sensitive to both backbone ϕ and ψ torsion angles, and Spera & Bax (4) showed that helical and sheet segments could be clearly differentiated based upon differences in their C^{α} and C^{β} chemical shifts. Because of the clearly defined dependence of the observed chemical shifts on these backbone torsion angles, their work strongly suggested that it should be possible to compute such experimental chemical shifts by using quantum chemical techniques.

Exactly how such calculations should be carried out was not clear, however. In particular, we had made a number of observations of the ^{13}C and ^{17}O NMR chemical shifts of the CO ligands in a variety of carbonmonoxyheme proteins. We had found that the ^{13}C and ^{17}O NMR chemical shifts were highly (anti) correlated and were also proportional to the CO vibrational stretch frequencies and the ^{17}O NMR-determined electric field gradients, as well as the Fe-C vibrational stretch frequencies determined from Raman spectroscopy (5). Augspurger et al. suggested that the origin of all these correlations was due to electrical polarization of the FeCO fragment in the protein, an electrostatic field effect, which appeared to be the case (6). The question then arose as to whether the ^{13}C NMR chemical shifts seen by Spera & Bax might also have a primarily electrostatic origin. It was also well-known that the ^{19}F NMR chemical shifts in proteins covered a very wide range, and in this case purely electronic structural differences caused by changes in backbone torsion angles would not reasonably be expected to have any direct effect on the fluorine chemical shifts.

Fortunately, it is now possible to compute essentially all of these chemical shifts in proteins by using quantum chemical techniques, suitably tailored to the nucleus of interest. In the case of ^{13}C and ^{15}N nuclei in amino-acid residues in peptides and proteins, the chemical shifts report on local geometric or electronic structure, in particular, the backbone and side-chain torsion angles, ϕ, ψ, and χ. However, in the case of ^{19}F nuclei in aromatic amino acids in proteins as well as in the case of ^{13}C and ^{17}O NMR chemical shifts in carbonmonoxyheme proteins, it is the electrostatic field generated by the protein that is responsible for the observed chemical shift nonequivalencies. As a result, it is necessary to use two rather different approaches to computing experimental spectra: one focusing on torsional angles and electronic effects, the second on electrostatics.

I begin this review by discussing the computation of the ^{13}C and ^{15}N NMR chemical shifts in amino acids, peptides, and proteins using both ab initio Hartree-Fock and density functional theory (DFT) techniques. This is followed by a brief description of recent progress in computing chemical shifts in paramagnetic systems. I then describe approaches to the use of chemical shifts and chemical shift tensors in predicting and refining aspects of peptide and protein structure. Next, I

consider the computation of ^{19}F NMR chemical shifts in proteins and model systems, with emphasis on the effects of electrostatic fields. I then consider the case of metalloproteins such as carbonmonoxymyoglobin, in which questions about metal-ligand geometries and electrostatic field effects can be probed by using a combination of chemical shift, chemical shift tensor, Mössbauer, infrared, and Raman spectroscopy, together with quantum chemical calculations. This leads naturally into a more detailed discussion of electrostatics, including a consideration of hydrogen bonding in proteins, the covalence of the hydrogen bond, and the observation of through-space hydrogen bond J-couplings. Finally, I briefly describe results of the use of purely experimental NMR chemical shifts in the area of drug design, an important recent development in the use of chemical shift information.

With modern codes and high-speed computer workstations or workstation clusters and accessibility to supercomputer centers, it is now possible to evaluate essentially all of the spectroscopic observables in amino acids, peptides, or proteins with good accuracy. Whereas I focus on the chemical shift or chemical shielding and its associated tensor element magnitudes (and orientations), I also consider topics such as spin densities (which control hyperfine shifts in proteins), together with J-couplings and electric field gradients, the latter being of interest both in NMR and Mössbauer spectroscopy. Consideration of the electric field gradient tensor leads into a more general consideration of electrostatics, in which the results of calculations can be validated against charge densities, electrostatic potentials, and Laplacians/Hessians of the charge densities obtained from high-resolution X-ray diffraction experiments. In every instance I make a comparison between the experimental observable and the computed property, obtained by using quantum chemistry. Moreover, where possible, I consider complete tensor representations of the observable and the computed property, including tensor orientations, which for the chemical shielding, electric field gradient, and Hessian-of-$\rho(\mathbf{r})$ tensors may be conveniently combined and described in an icosahedral representation.

CARBON-13 AND NITROGEN-15 NMR CHEMICAL SHIFTS IN AMINO ACIDS, PEPTIDES, AND PROTEINS

How does one calculate the chemical shifts of an amino acid, a peptide, or a protein? In early work we attempted to find empirical correlations between experimental ^{13}C chemical shifts and obvious aspects of protein secondary structure, such as the backbone ϕ and ψ torsion angles, with electrostatic field effects and with ring current effects. However, this met with little success, even though the combination of electrostatic field and ring current effects had found some success in reproducing ^{1}H chemical shifts in proteins (7). The reason for this lack of success can be attributed primarily to the dominance of the paramagnetic term in determining shielding of the heavier elements ^{13}C and ^{15}N. Consequently, we decided to attempt

the use of ab initio Hartree-Fock quantum chemical techniques to evaluate shielding in amino acids, peptides, and proteins. For amino acids, the use of Hartree Fock methods (or density functional theory techniques) might be expected to be successful because the molecules are quite small, although one might expect significant contributions from the electrostatic fields arising from neighboring zwitterionic amino acid molecules. However, in the case of proteins, it was unclear to what extent it would be possible to evaluate chemical shifts, because protein NMR (and crystal) structures have considerable structural uncertainties compared with the small molecule structures more typically investigated using quantum chemistry. In addition, the effects of solvation/dielectric constant/electrostatic field effects and internal motion might all make major contributions to shielding. It was also unclear just how large a structure would have to be investigated in order to adequately reproduce the experimental NMR spectra. Would ab initio or density functional theory computations of protein NMR chemical shifts be feasible? We thought so.

In early work we therefore decided to try to predict the ^{13}C NMR chemical shielding tensors in two amino acids: zwitterionic L-threonine and L-tyrosine. These molecules were chosen because they both had high-resolution X-ray structures. Additionally, we had carried out a single crystal ^{13}C NMR shielding tensor determination of each of the carbons in L-threonine (8), and Frydman and coworkers had carried out a determination of the shielding tensor elements of each of the carbons in L-tyrosine (9). When taken together, the experimental data provided 39 individual shielding tensor elements covering a 250-ppm chemical shift range, in addition to the orientations of the chemical shielding tensor elements in the molecular frame in L-threonine. This represented a sizeable body of data for investigation using quantum chemical techniques.

To carry out these calculations we used the TEXAS90 program of Pulay, Wolinski, and Hinton, which makes use of an efficient implementation of the gauge-including-atomic-orbital (GIAO) methods introduced by them in earlier work (10). Because essentially all atomic sites were of interest, we used a uniform 6-31G** basis in all calculations. The combined results for the ^{13}C shielding tensors in L-threonine and L-tyrosine were very promising, as shown in Figure 1*A* (11). The slope of the fitted line is -1.03 and the intercept is about 206 ppm [$R^2 = 0.987$, root mean squared deviation (RMSD) from the fitted line $= 12.3$ ppm]. However, several points (I-IV in Figure 1*A*) significantly deviated from the fitted line, and these points arose from the tensor components σ_{11} and σ_{22} of the carboxyl ^{13}C sites in both molecules. The σ_{11} component lies in the sp^2 plane and is perpendicular to the C^α–C^β bond, whereas σ_{22} lies perpendicular to the sp^2 plane. Upon comparison with experiment, σ_{11} is underestimated, whereas σ_{22} is overestimated, so the discrepancies are less dramatic when only the isotropic values are considered. That is to say, evaluation of the shielding tensor elements is a more rigorous test of the quality of a calculation than is a determination of solely the isotropic chemical shift.

The origins of these discrepancies were tentatively attributed to the effects of the electrostatic fields originating from nearby charge centers. We therefore

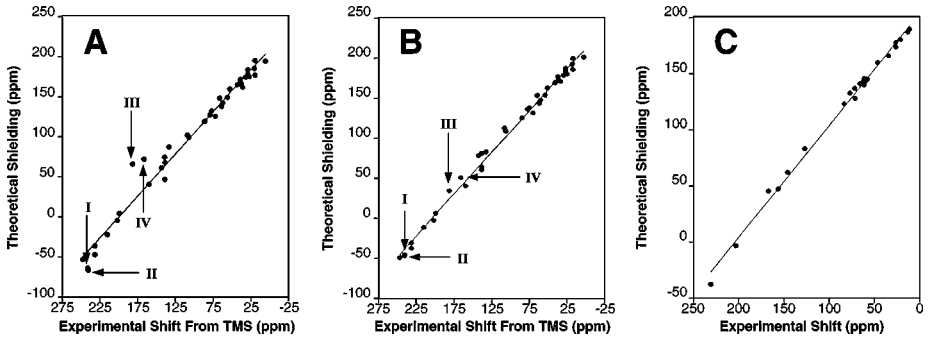


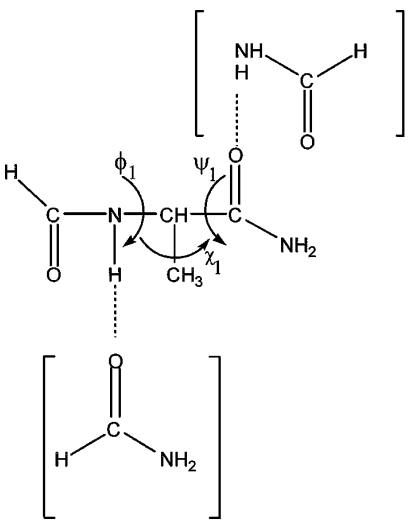
Figure 1 Graphs showing experimental shifts [ppm from tetramethylsilane (TMS)] versus theoretical shieldings (in ppm) for ^{13}C NMR shielding tensor elements in L-threonine and L-tyrosine. (*A*) Single-residue calculation; (*B*) charge-field perturbation–gauge-including-atomic-orbital (CFP-GIAO) calculation; (*C*) CFP-GIAO calculation for L-threonine only, icosahedral tensor representation. I = σ_{11} Tyr; II = σ_{11} Thr; III = σ_{22} Tyr; IV = σ_{22} Thr. Reproduced with permission from *J. Am. Chem. Soc.* 1994 116:7784–7786. Copyright 1994 Am. Chem. Soc.

carried out a second series of calculations in which unit cell translations were used to generate 28 additional molecules in the vicinity of the molecule of interest. Only the molecule of interest had basis sets assigned to its atoms, neighboring molecules being represented by point charges obtained from an AMBER force-field. Considerable improvements were achieved when the calculations were performed in the presence of point charges, as shown in Figure 1*B* (slope = −1.04, intercept = 207 ppm, R^2 = 0.996, RMSD = 6.4 ppm), where the previous outlying points for σ_{11} and σ_{22} are now much closer to the fitted line. The electrostatic contributions led to a significant increase in the value of σ_{11} for the carboxyl groups [16 ppm for L-tyrosine (I) and 11 ppm for L-threonine (II)] and a major decrease in σ_{22} [37 ppm for L-tyrosine (III) and 27 ppm for L-threonine (IV)]. These results demonstrated the ability to evaluate the magnitudes of the ^{13}C shielding tensors in amino acids (11).

Analysis of a shielding tensor involves examination not only of the principal components of the tensor but also the orientation of these components with respect to the molecular frame. Fortunately, in our previous NMR work on L-threonine we utilized a single crystal in our study, and consequently the orientations of the ^{13}C shielding tensors were well known for each site, enabling a comparison of tensor orientations. If one compares the experimental and calculated direction cosines with respect to the crystallographic axes of the ^{13}C shielding tensors, one obtains good agreement, although comparing tensor orientations in terms of direction cosines does not properly gauge the calculations because in addition to being difficult to visualize, errors in direction cosines do not translate directly into errors in terms of radians or degrees. We therefore used the icosahedral representation

described previously by Alderman and coworkers (12), in which the three principal components of the shielding tensor together with the three direction cosines are transformed into six icosahedral shielding tensor elements. For a 220 ppm overall chemical shielding tensor element range, the so-called charge-field perturbation–gauge-including-atomic-orbitals (CFP-GIAO) method yielded excellent results, as shown in Figure 1*C*, in which there is an R^2 value of 0.997 and an RMSD of 4.4 ppm. Thus, the results of these early calculations clearly indicated, at least with amino acids, that it was possible to compute the shielding tensor elements, as well as their orientations, with good accuracy.

The next question that arises is, to what extent can the chemical shifts of peptides and in particular the much larger proteins of principal interest be predicted by using quantum chemical techniques? To investigate this question we first began to investigate the C^{α} and C^{β} chemical shifts for the helical and beta sheet residues in the database of Spera & Bax (4). These workers had shown that a clear distinction could be made between the chemical shifts of residues with these two distinct structure types, which suggested (but did not prove) a major, direct effect of the ϕ and ψ torsion angles on electronic structure and thence on shielding. However, other effects such as hydrogen bonding/electric field effects might also be important, as might solvation and dynamics. We therefore began to investigate the C^{α} and C^{β} chemical shieldings in a series of alanine molecular fragments, as shown in the following structure:



To begin, we used a relatively small and uniformly dense (6-311G**) basis for all atoms but neglected the formamide hydrogen bond partners shown above. Figure 2*A* shows histograms of the experimental C^{α} and C^{β} chemical shifts reported in the database of Spera & Bax, and Figure 2*B* shows the theoretical chemical shielding results computed by using ab initio Hartree-Fock GIAO calculations

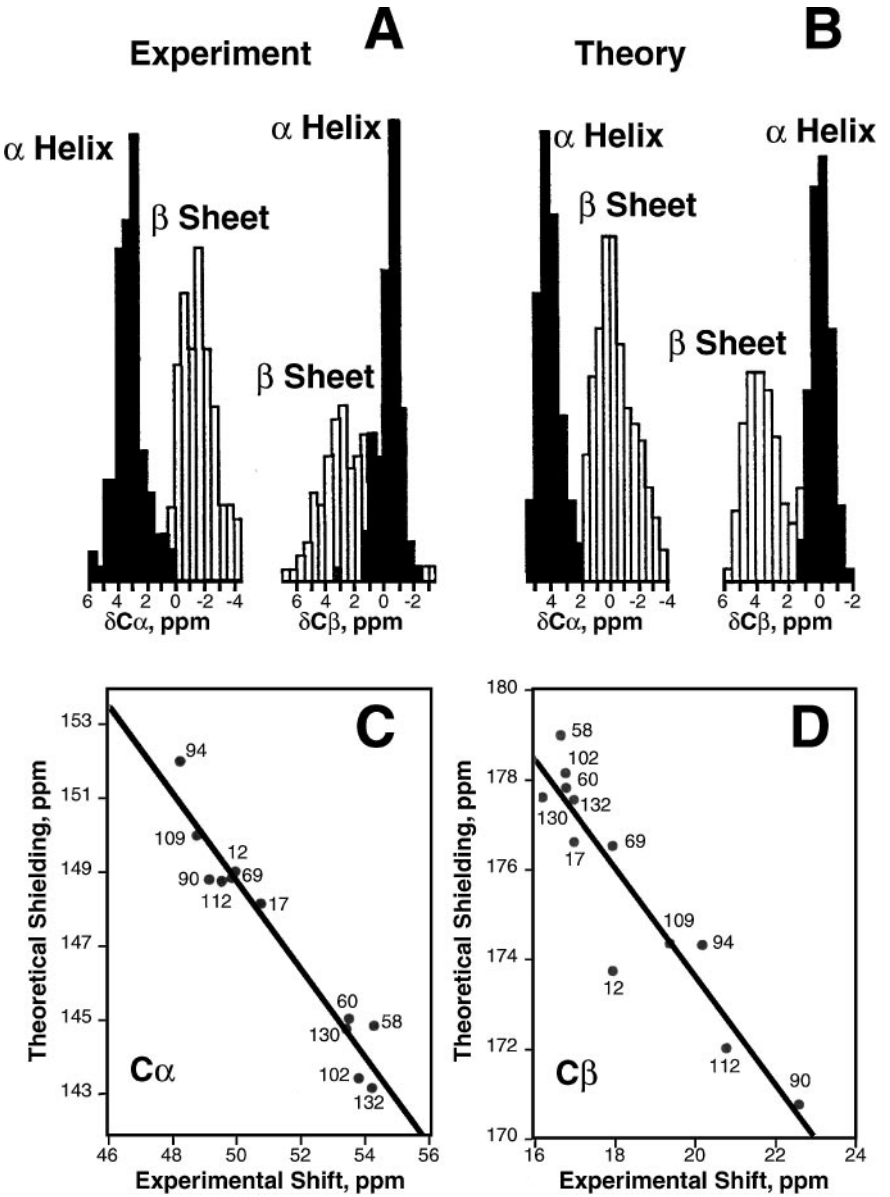


Figure 2 ^{13}C NMR shielding results for C^{α}, C^{β} sites in protein. (*A*) C^{α}/C^{β} shift separation histograms from Spera & Bax (4); (*B*) computed C^{α}/C^{β} shift histogram for model alanine fragments; (*C*) experimental shift/theoretical shielding results for 12 Ala C^{α} sites in SNase; and (*D*) same as (*C*) but with C^{β} sites. Reprinted with permission from *Science* 260:1491–96 (1993). Copyright 1993 Am. Assoc. Adv. Sci.

(13). Clearly, the general ~5 ppm chemical shift separation seen experimentally between the helical and sheet residues for both C^{α} and the C^{β} were reproduced in the calculations, but we did not have a one-to-one comparison between specific protein chemical shifts and experimentally computed shieldings, only a general trend. We therefore extended our calculations to investigate 12 specific alanine residues in a nuclease from *Staphylococcus aureus* [Staphylococcal nuclease (SNase)]. To get the best agreement between theory and experiment, we used a locally dense basis set approach proposed originally by Chesnut & Moore (14), in which a larger number of basis functions is given to the nucleus whose chemical shielding is to be calculated. In this case we used a 6-311++G(2d,2p) basis for C^{α}, C^{β}, C^{o}, N^{H}, and H^{α}, but only a 6-31G basis for the other atoms in the fragment.

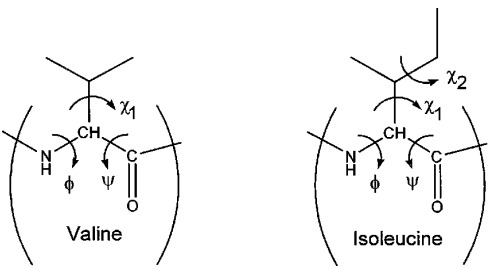
Given a general concern about the possible effects of solvation, electrostatic field effects, hydrogen bonding, dynamics, and magnetic susceptibility effects, the experimentally observed slope of 0.85 and regression coefficient r = 0.94 represented a rather promising initial agreement between theory and experiment (13). We then investigated the effects of hydrogen bonding on shielding using the formamide hydrogen bond partners shown above, together with the effects of the protein's charge field, using the CFP-GIAO approach used with the amino acids. For both C^{α} and C^{β}, we found excellent agreement between theory and experiment, as shown in Figures 2*C*,*D*. Interestingly, however, there was only moderate accord between theory and experiment when fragment geometries based directly on the crystallographic structures were utilized, and we had to carry out an energy minimization of the protein (crystal) structure (suggested by C. Jameson) by using molecular mechanics techniques, in order to regularize all of the bond lengths and three-atom bond angles to those of an AMBER forcefield. The fact that the ^{13}C NMR chemical shifts are highly sensitive to bond lengths then led to an investigation of the bond-length and bond-angle shielding derivatives in glycine, alanine, and valine residues in which chemical shieldings were calculated for both C^{α} and C^{β} at various geometries by systematically varying one structural parameter (a bond length or a bond angle) while simultaneously fixing all other geometric parameters (15). The shielding traces were expressed in terms of Taylor series expansions of the shielding as a function of bond-length (r) or bond-angle (θ):

$$\sigma(\mathrm{r}) = \sigma(\mathrm{r}_0) + \left(\frac{\partial\sigma}{\partial \mathrm{r}}\right)_{\mathrm{r}_0}(\mathrm{r}-\mathrm{r}_0) + \frac{1}{2}\left(\frac{\partial^2\sigma}{\partial \mathrm{r}^2}\right)_{\mathrm{r}_0}(\mathrm{r}-\mathrm{r}_0)^2 + \frac{1}{6}\left(\frac{\partial^3\sigma}{\partial \mathrm{r}^3}\right)_{\mathrm{r}_0}(\mathrm{r}-\mathrm{r}_0)^3 + \cdots \quad 1.$$

$$\sigma(\theta) = \sigma(\theta_0) + \left(\frac{\partial\sigma}{\partial \mathrm{r}}\right)_{\mathrm{r}_0}(\theta-\theta_0) + \frac{1}{2}\left(\frac{\partial^2\sigma}{\partial \theta^2}\right)_{\theta_0}(\theta-\theta_0)^2 + \frac{1}{6}\left(\frac{\partial^3\sigma}{\partial \theta^3}\right)_{\theta_0}(\theta-\theta_0)^3 + \cdots \quad 2.$$

with first, second, and third-order derivatives being reported. The results of these calculations clearly showed that the effects of individual bond-length and bond-angle changes were additive, making it possible to predict chemical shifts for a given set of structural parameters without having to perform further ab initio calculations, once a basic Ramachandran shielding surface [defined here as $\sigma(\phi, \psi)$] has been evaluated. Ramachandran shielding surfaces for all of the naturally occurring amino acids in their most popular side-chain conformations have now been evaluated and are available at http://feh.scs.uiuc.edu.

Of course, as one increases the complexity of the amino acids, from glycine to alanine to valine to isoleucine, for example, the number of structures that need to be computed increases. In the case of alanine there is an additional torsional angle, χ_1, but in this case the methyl group can simply be fixed at its fully staggered conformation. However, in the case of valine residues, there are three clearly defined χ_1 conformers, two of which occur frequently in proteins. In the case of isoleucine both χ_1 and χ_2 conformers exist, as shown below:



leading in principle to nine side-chain conformers, although once again only two of these occur with high frequency in proteins. For example, in calculations of valine shifts in proteins, we used an N-formyl-L-valine amide structure to evaluate complete Ramachandran shielding surfaces for all three χ_1 conformers, and shielding results for C^α and C^β are shown in Figure 3 (16). If the ϕ, ψ, and χ values in a protein are known, then it should be possible to use such shielding surfaces to compute the experimental C^α and C^β chemical shifts from the theoretical shielding surfaces. However, when we attempted to use the theoretical shielding surfaces to predict the ^{13}C NMR chemical shifts for valine residues in calmodulin using torsion angles determined from X-ray crystallography, we found only a modest agreement between theory and experiment. The results were particularly puzzling for helical C^β shifts because five out of seven valine residues in calmodulin have (solution NMR) C^β shifts in the 31.3–31.7 ppm range, even though the X-ray structures showed vastly different χ_1 angles, which would be expected to result in a much wider chemical shift distribution. However, assuming that all of the valine residues in calmodulin in solution were present as the *trans* conformer ($\chi_1 = 180^\circ$) resulted in excellent agreement between experimental and predicted C^α and C^β chemical shifts (16). This result was consistent with J-coupling constant information obtained from Ikura, which showed that six out of seven valine residues existed in the $\chi_1 = 180^\circ$ conformation in calmodulin in solution.

This success then led to the desire to calculate a complete valine chemical shielding hypersurface, the chemical shift as a function of ϕ, ψ, and χ, which involved the evaluation of 1728 points uniformly distributed in ϕ, ψ, and χ (17). The chemical shifts predicted using this shielding hypersurface were in good agreement with the chemical shift values determined experimentally. For example, the C^{α} chemical shifts in a Staphylococcal nuclease and an invertebrate calmodulin were predicted with an RMSD = 1 ppm and an $R^2 = 0.842$ (17).

But what about ^{13}C shielding tensors and their orientations? In order to answer this question, we evaluated the ^{13}C NMR shielding tensors in glycine, valine, isoleucine, serine, and threonine, comparing the tensor magnitudes and orientations for helical and sheet residues, including the effects of χ_1 on shielding (18). Results for glycine and alanine showed the expected ~4–5 ppm increase in the isotropic chemical shielding of sheet over helical residue geometries and generally similar overall breadths of shielding tensors for both helical and sheet fragments. However, for each of the C^{β}-branched amino acids (valine, isoleucine, and threonine) we found not only an increase in shielding of sheet over helical fragments, but we also noted a large increase in the overall tensor breadths for the sheet residues. On average, C^{α} sheet tensor breadths were ~50% larger than the overall tensor breadths found for helical residues. In addition, the C^{α} shielding tensor elements were all sensitive to χ_1, and the orientations of the shielding tensors were quite different in helical and sheet geometries (18).

We then began to investigate experimentally the C^{α} and C^{β} shielding tensor magnitudes in alanine containing tripeptides with a wide range of geometries, encompassing both helical and sheet regions of Ramachandran space. Using three alanine $^{13}C^{\alpha}$–labeled tripeptides and slow magic-angle sample-spinning, the individual components of the C^{α} shielding tensor were evaluated and compared with predictions made by using the theoretical shielding surface (19). For the nine tensor elements, an R^2 value of 0.99 and RMSD of 2.15 ppm was found, although the slope of the correlation was only -0.67, suggesting a small systematic error in the theoretical calculations. For a $^{13}C^{\beta}$-alanine labeled tripeptide, an R^2 value of 0.99 was found (20), together with a slope of -1.07 (compare with an ideal value of -1.0). When taken together, these results indicate that there is good to excellent agreement between the experimental and theoretical shielding tensor element magnitudes, with the helical/sheet tensor differences seen in the calculations being reproduced experimentally.

In other recent experimental work, Tjandra & Bax have shown that it is possible to determine the C^{α} chemical shift anisotropy (defined as the value of the C^{α} shielding along the C^{α}–H^{α} bond minus the average perpendicular to this bond) for proteins in solution by using dipole-dipole/chemical shift anisotropy interference techniques (21). Sitkoff & Case have compared the results of these experimental C^{α} "chemical shift anisotropies" with their calculations for ubiquitin and a calmodulin/M13 complex (22). They computed shielding surfaces for an alanine dipeptide using structures that were optimized from an all-atom CHARMM forcefield, using a 30° grid for their Ramachandran shielding surface. The anisotropies

were approximated from the surface by using a Fourier series. The correlation between computed and experimental values was quite good, although the computed results appeared to be systematically somewhat larger than the experimental results, possibly owing to motional averaging about the mean structure (22). We obtained similar results using a density functional theory (DFT)–computed alanine C^{α} shielding surface, with a slight improvement when using an ab initio Hartree-Fock surface, for all residue types (20). Even better agreement was obtained when specific glycine, alanine, valine, serine, threonine, isoleucine, and phenylalanine surfaces were used, with the best correlations being found with ubiquitin, presumably because it is a smaller protein and has a higher resolution (X-ray crystallographic) structure.

At present, there are no direct determinations of ^{13}C shielding tensor element orientations in proteins or peptides, only the amino acid results discussed previously and what may be inferred from the relaxation interference experiments. Nevertheless, the fact that C^{α} chemical shift tensor breadths or chemical shift anisotropies in sheet structures are often considerably larger than those found in helical structures has led to the development of a novel chemical shift anisotropy "filtering" experiment by Hong and coworkers (23). This experiment enables the acquisition of solid-state NMR spectra of primarily helical spectra (in this case in the protein ubiquitin), together with a determination of chemical shift anisotropy (CSA) values, which were found to be in good agreement with the results of the quantum chemical calculations. Spectra of sheet residues can be obtained by using a second filtering experiment (24), and are of use in spectral assignments.

The ability to obtain resolved, high-resolution solid-state NMR spectra of proteins in the crystalline solid state is also of interest because it enables a rather direct comparison between crystal and solution NMR structures, based on an analysis of NMR chemical shifts. In early work, Cole and coworkers utilized one-dimensional magic-angle sample-spinning NMR to investigate selectively ^{13}C- and ^{15}N-labeled SNase and concluded that the chemical shift patterns seen in the solid state were similar to those observed in solution (25). More recently, using high-field, high-speed magic-angle sample-spinning and high-power proton decoupling, two-dimensional NMR spectra of microcrystalline bovine pancreatic trypsin inhibitor (BPTI) have been obtained and a number of specific spectral assignments have been made (26). There was once again good agreement between solution and crystal NMR shifts. In the case of BPTI, we have computed the chemical shifts for C^{α}, C^{β}, $C^{\gamma 1}$, $C^{\gamma 2}$, and C^{δ} for two isoleucine residues using the X-ray crystal structure and found good agreement between theory and experimental chemical shifts (27). Thus, it is possible to compute more than just C^{α} and C^{β} shieldings in proteins, and it seems likely that quantum chemical calculations may have a role to play in identifying side-chain conformations in proteins, both in solution and in the solid state.

In addition to investigating the chemical shifts of aliphatic carbons in proteins, there is considerable interest in understanding the chemical shifts of the carbon

and nitrogen atoms in the peptide bond. As had been found for C^{α} chemical shifts, carbonyl carbons (C^{o}) in helical residues were known to resonate ~5 ppm downfield of carbonyl carbons in sheet residues in proteins. It would therefore seem reasonable that it should be possible to compute this chemical shift separation by using quantum chemical techniques, but early calculations showed the opposite trend to that observed experimentally, with the helical carbonyl sites being more shielded. Because it seemed likely that hydrogen bonding (as opposed to backbone torsion angles) might be responsible for the deshielding of helical residues, this was also investigated, but agreement with experiment could only be obtained at artificially short hydrogen bond distances. Based on single molecule geometry optimizations, it also appeared possible that the peptide nitrogen might be somewhat pyramidal, but this is contrary to recent experimental observations. However, when shielding computations were performed on SCF-optimized helical and β-turn N-formylpentaalanine amide structures, in which the hydrogen bond structure is geometry optimized (28), a helix-sheet chemical shift difference of 4.9 ppm was found (29), with the helical site being deshielded—as is found experimentally, where carbonyl nuclei in helical alanine sites are typically deshielded by ~4.6 ppm when compared with sheet-like residues.

Based on quantum chemical calculations, the ϕ, ψ contributions are about −1.3 ppm (29), so the overall hydrogen bond contribution to deshielding is about 6.2 ppm (29). The chemical shift separation seen for carbonyl groups appears to be due to the changes in C–O bond length that occur on hydrogen bonding, consistent with a very large first-order shielding derivative in the carbonyl group. More recently, Walling and coworkers have made a detailed investigation of the carbonyl shielding tensor in peptides and concluded that the principal axis system of the C^{o} shielding tensor follows the directions given by the amide plane and the C=O bond and that hydrogen bonding does not have a major effect on the orientation of this principal axis system (30).

Within the peptide group, the other hetero-atom for which there exists an extensive database of structural and chemical shift information is nitrogen. In the case of ^{15}N, however, it is not solely the torsion angles ϕ and ψ that dominate shielding, as is the case with C^{α} residues, owing of course to the fact that the peptide group bridges two different amino acid residues. As a result, it is the torsion angles ϕ and ψ_{i-1}, where i refers to the residue whose ^{15}N chemical shift is of interest, that dominate shielding. In early work, it was found that two-dimensional chemical shielding surfaces $\delta(\phi_i, \psi_{i-1})$ could be constructed that gave moderate predictions of the ^{15}N chemical shifts for a given structure (31). On average, the rms error between experiment and prediction was ~3.5 ppm for a chemical shift range of ~15–20 ppm. For β-branched residues, such as valine, the experimental chemical shift range is larger (up to ~35 ppm) and variations in χ_1 values (not reflected in the empirical shielding surfaces) resulted in somewhat larger rms errors of ~4.8 ppm. The results of ab initio quantum chemical calculations confirmed that indeed it is the two backbone dihedral angles closest to the peptide group, ϕ_{i-1} and ϕ_i, that contribute most (~20 ppm) to the experimental shielding, whereas the adjacent

torsion angles ϕ_{i-1} and ψ_i make a contribution of up to ~8 ppm (32). Variations in side-chain conformation produced up to ~4 ppm chemical shift effects. In addition, as expected from earlier work, hydrogen bonding effects are also important contributors to ^{15}N shielding in proteins. Both ab initio Hartree-Fock and density functional theory calculations of ^{15}N shieldings gave a reasonably good description of experimental chemical shifts in a small series of proteins. For example, in the case of 13 alanine residues in a cytochrome c_{551}, DFT shieldings were found to have a correlation line of slope -1.04 and an R^2 value of 0.97, whereas the Hartree-Fock results had a slope of -1.11 and an R^2 value all 0.96, with the DFT results giving a closer absolute shielding intercept (251.3 ppm) to the experimental value of 244.6 ppm (32).

As for the chemical shift anisotropy of ^{15}N in proteins, there have been extensive investigations into the measurement and use of the ^{15}N CSA in the determination of peptide/protein structures in membranes using oriented samples (33, 34), together with numerous investigations of the ^{15}N CSA for proteins in solution using dipole-dipole/CSA cross-correlation techniques (35, 36). Results with ubiquitin show a larger number of statistically significant values far from the mean CSA, a result not observed in ribonuclease H. It may be possible to better understand the shielding tensor results by using DFT techniques, as described above, although such calculations have not yet been reported.

PARAMAGNETIC SYSTEMS

All of the systems discussed above have been diamagnetic. However, there are many proteins of interest that contain paramagnetic centers, and it is perhaps surprising that there has been very little work done on the computation of spectroscopic properties in these systems. Evaluating the chemical or purely orbital shifts in paramagnetic systems is generally not feasible because of the breakdown of the Ramsey formulation. Fortunately, however, we are not concerned with computing purely orbital shifts in paramagnetic systems, because the observed overall shift can to a good first approximation be broken down into the purely orbital (or diamagnetic) shift of a reference (diamagnetic) model compound and a hyperfine shift, made up of a Fermi contact contribution together with a dipolar or pseudo-contact shift (37). The purely orbital or diamagnetic shift can be computed using standard quantum chemical methods or can be estimated from the chemical shifts of model compounds, and in some cases from temperature dependence studies. The pseudo-contact shift has already been used to carry out protein structure refinements (38). There has, however, been much less use of the Fermi contact shift. Nevertheless, in recent work, Wilkens et al. used DFT to compute ^{2}H, ^{13}C, and ^{15}N (protein) Fermi contact shifts for an iron-sulfur protein (39) and found very good correlation between theory and experiment. Their results clearly open the way to using Fermi contact shifts without any extensive semi-empirical parameterizations, as well as dipolar or pseudo-contact shifts, as structure refinement tools.

DFT methods appear to be particularly attractive for use in computing the Fermi contact shift, as well as Mössbauer quadrupole splittings (even in paramagnetic systems) because all are ground-state properties and, as is well known, the Hohenberg/Kohn formulation of DFT states that all ground-state properties are functions of the charge density, $\rho(\mathrm{r})$, where:

$$\rho(\mathbf{r}) = \sum_{\mathrm{i}} [\phi_{\mathrm{i}}(\mathbf{r})]^2 \qquad 3.$$

and ϕ_{i} represents a molecular orbital. The molecular orbitals can be occupied by either spin up (α) or spin down (β) electrons, so it is possible to form two different charge densities. Putting both α and β spins into the same molecular orbital (MO) is called a spin-restricted calculation, whereas putting α and β spins into different MOs is called a spin-unrestricted calculation. In the unrestricted case, it is possible to form two sets of charge densities, one for the α and one for the β MOs. Their sums give the overall charge density $\rho(\mathrm{r})$, whereas their differences give the spin density, $\rho_{\alpha\beta}$:

$$\rho_{\alpha\beta} = \sum_{\mathrm{i}} \left[\left|\Psi_{\mathrm{i}}^{\beta}(0)\right|^2 - \left|\Psi_{\mathrm{i}}^{\alpha}(0)\right|^2 \right] \qquad 4.$$

and the Fermi contact shift, δ^{FC}, is directly proportional to the spin density, $\rho_{\alpha\beta}$:

$$\delta^{\mathrm{FC}} = \frac{\mu_0 \mu_{\mathrm{B}}^2 \mathrm{g}_{\mathrm{e}}^2 (\mathrm{S}+1)}{9\mathrm{kT}} \bullet \rho_{\alpha\beta}. \qquad 5.$$

For open shell molecules (radicals and other paramagnetic systems), DFT methods appear to be the most promising ones for handling spin polarization without significant spin contamination (the appearance of electronic states of higher multiplicity), as demonstrated by the remarkable success of Wilkens et al. in computing ^{2}H, ^{13}C, and ^{15}N Fermi contact shifts in rubredoxins from *Clostridium pasteurianum* (39, 40).

Wilkens et al. utilized a 104-atom model of the iron-sulphur cluster and neighboring amino acid residues in a rubredoxin from *C. pastuerium* as a basis for DFT calculations of the spin densities on the ^{1}H, ^{2}H, ^{13}C, and ^{15}N nuclei of threonine, valine, glycine, leucine, and the metal-cluster-coordinated cysteine residues in oxidized and reduced protein forms. They used the Gaussian-94 program with a B3LYP hybrid exchange-correlation functional and a uniform 6-311G** basis set to evaluate the Fermi contact spin densities and clearly showed that the experimental hyperfine shifts were dominated by the Fermi contact interaction. The computed spin densities were all found to correlate linearly with the experimental isotropic hyperfine shifts for Fe(III) rubredoxin, with R^2 values of 0.93–0.96 for 12 experimental ^{2}H NMR resonances (depending on the exact geometric model used) and 0.85–0.96 for 12 experimental ^{15}N resonances. Using the same hybrid exchange correlation functional, but with a smaller basis set (3-21G*), these workers carried out a geometry optimization of the H^{N} atom positions in their models,

which resulted in an improvement in R^2 (from 0.85 to 0.94) for the ^{15}N shifts for the ferric protein.

There was much worse agreement, however, for Fe(II) shifts using a hypothetical model for the ferrous protein structure derived from X-ray crystallographic structures of the ferric protein, suggesting that the ferrous protein undergoes structural changes upon reduction. Nevertheless, given a ^2H shift range of $\sim$800 ppm and an ^{15}N chemical shift range of $>$600 ppm in the ferric protein, these results are very promising. Our group has been investigating the hyperfine shifts in paramagnetic heme proteins, and as with the work described above on rubredoxin, we have found good agreement between the experimentally observed hyperfine shifts (for ^{13}C) and those computed with density functional theory, a hybrid exchange-correlation functional, and moderately sized basis sets (J. Mao, Y. Zhang, E. Oldfield, unpublished data). It seems clear that in future work DFT techniques will prove to be very helpful in predicting hyperfine shifts in paramagnetic proteins, in addition to being used in structure prediction and refinement, as are semi-empirical correlations currently.

PREDICTING/REFINING ASPECTS OF PEPTIDE AND PROTEIN STRUCTURE

Simply being able to use quantum chemical techniques to predict chemical shifts and chemical shift tensors in amino acids, peptides, and proteins is, in and of itself, of some interest. However, it might be more useful to use experimental chemical shifts or chemical shift tensor information to determine or at least to refine aspects of peptide and protein structure. Such a capability might be particularly useful for noncrystalline materials, such as membrane-bound proteins, in addition of course to playing a role in structure determination for proteins in solution.

In a recent study, Van Alsenoy et al. used ab initio Hartree-Fock methods to carry out the first quantum chemical geometry optimization of a small protein (41), and with the availability of larger computational resources and refined codes, it is entirely possible that geometry optimizations of small proteins may become more generally accessible over the next few years. We found that energy minimization/geometry optimization appears to be quite useful in evaluating ^{13}C NMR chemical shifts, at least for valine residues in proteins (42). We first carried out both molecular mechanics (an AMBER forcefield in Discover, combined with a conjugate gradient minimization) and Hartree-Fock geometry optimization of 19 valine residues in 3 proteins. We then carried out chemical shift calculations on these structures; the results showed that the most successful prediction of both C^α and C^β chemical shifts was obtained for all 19 residues by utilizing the lowest energy χ_1 conformer obtained from these geometry optimizations, together with the X-ray ϕ, ψ values (42). This strongly suggests that the X-ray–determined backbone torsion angles are accurate and reflect the solution values, whereas the side-chain conformations appear to be less so, or at least they are more likely

to differ between proteins in solution and in the crystalline solid state. This suggests that geometry optimization using quantum chemical techniques can play a role in structure determination, but faster computers are needed for more routine application.

We have also begun to more directly predict backbone torsion angles for amino acids in proteins. We developed a Bayesian probability technique that utilizes C^{α}, C^{β}, and H^{α} chemical shifts to predict the likelihood that a particular ϕ, ψ torsion angle combination represents the experimental backbone torsion angles. In its simplest form, one would have a parameter, P, which is a function of a single angle, α, such that

$$P = f(\alpha). \qquad 6.$$

We define the probability that the experimental value of P, P_{expt}, corresponds to a given angle, α, given by the unnormalized probability, Z, as (43):

$$Z = e^{-\frac{(P_{expt} - f(\alpha))^2}{W}}, \qquad 7.$$

where W represents a search width (to take into account any computational inadequacies or experimental uncertainties). The approach can be readily applied to a two-dimensional chemical shift surface in which ϕ and ψ are variables:

$$Z(\phi, \psi) = \exp\left[-(\delta_{expt} - \delta(\phi, \psi)^2/W\right]. \qquad 8.$$

As additional properties, P_1, $P_2 \ldots P_n$, are investigated, the overall conditional or Bayesian probability simply becomes the cumulative probability ${}^{n}Z = Z_1^{\times} Z_2^{\times} \ldots Z_n$. Using empirically determined parameter (chemical shift and J-coupling) surfaces, the backbone ϕ, ψ error between prediction and experiment was about 17°, but for 22 out of 24 ϕ, ψ torsion angles for alanine residues in Staphylococcal nuclease, this reduced to ~10° when quantum mechanically computed ^{13}C shielding surfaces were utilized (43).

Using C^{α} and C^{β} chemical shift restraints for 12 alanine and 8 valine residues in the nuclease from *S. aureus*, we also used a direct approach to refining structure (43) and found that the ϕ, ψ, and χ values obtained from structures that included chemical shift restraints were much closer to the X-ray structure than were the solution NMR structures obtained without chemical shift restraints, yielding back-calculated chemical shifts in good accord with experiment and with only very minor nuclear Overhauser effect (NOE) violations (43). It should be noted that dipolar coupling information (44) may be even more useful in structure refinement (45) and is currently being employed in conjunction with chemical shift information by some groups.

Cornilescu and coworkers have reported a somewhat different approach to using multiple chemical shifts in structure determination. In their TALOS (torsion angle likelihood obtained from shifts) program (46), these workers used a database of C^{α}, C^{β}, C^{o}, H^{α}, and N^{H} chemical shifts for 20 proteins with highly resolved X-ray crystal structures and known solution NMR chemical shifts. The idea is to search

the database for sequences of adjacent residues, having sequence and chemical shift similarity, that give the closest matches to a (triplet) sequence of interest. This work is a useful advance over earlier work by Wishart and coworkers (47), which utilizes solely (secondary) chemical shift information, because both chemical shift and sequence homology information is used to predict the most likely backbone torsion angles for a residue in a sequence of interest. The results are impressive and indicate a $\sim 15^{\circ}$ RMSD between the output of the TALOS program and the X-ray–derived backbone torsion angles (46). It may be possible to extend this approach even further, and progress is being made in using solely chemical shift information in generating three-dimensional protein structures (48).

FLUORINE-19 NMR AND ELECTROSTATIC FIELD EFFECTS

Why might one be interested in fluorine-19 NMR spectroscopic studies of fluorine-labeled proteins? Historically, fluorine-19 NMR studies of protein structure, or at least spectroscopy began in the early 1970s. In these early days high-field, NMR spectrometers were not available and as noted above, ^{1}H NMR spectra were poorly resolved because of the small chemical shift range of the ^{1}H nucleus. However, ^{19}F was known to have a much larger chemical shift range. Consequently, a number of groups began to chemically modify proteins with ^{19}F probes (49) or to incorporate fluorine-labeled amino acids biosynthetically into proteins (50) in order to effect large spectral simplifications. The observed chemical shift range can indeed be very large. For example, we found a 16.8 ppm shift range for a [4-F]-Trp-labeled hen egg-white lysozyme (51) using a *Gallus domesticus* expression system. The ^{19}F chemical shifts in this enzyme are very sensitive to inhibitor binding, and such shift effects may in the future be of use in, for example, drug screening. However, the question arises: What is the origin of the large ^{19}F chemical shift nonequivalencies seen in native proteins? In earlier work we had estimated that the electric field–induced shifts in proteins would be very large, on the order of 10 ppm for fluoroaromatic groups, owing to the low dielectric constants expected in protein interiors and the large polarizability of the C–F bond, and consequently they might dominate the ^{19}F shift nonequivalencies seen experimentally. However, before we discuss electrostatic field effects in ^{19}F NMR spectra of proteins, we need to be confident that theoretical techniques can adequately describe ^{19}F NMR spectra of small molecules. We therefore investigated the isotropic chemical shifts for a series of 13 fluorobenzenes whose isotropic chemical shifts covered a 63-ppm chemical shift range, as well as the shielding tensor elements (derived from solid-state NMR), which covered a 237-ppm range. We used geometry optimization and a locally dense basis gauge-including-atomic-orbital (GIAO) approach to calculate shieldings and found that for the isotropic, liquid state chemical shifts, the slope was -0.94, $R^2 = 0.975$, and the RMSD was 3.1 ppm over the 63 ppm chemical shift range (52). For the solid-state shielding tensor elements, the slope was 0.94, $R^2 = 0.989$, and the RMSD was 6.5 ppm over the 237 ppm chemical shielding

range, essentially the uncertainty in the experimental measurements (52). This gives considerable confidence in the quality of ^{19}F NMR shielding calculations, at least in the absence of an electrical perturbation.

We then investigated the effects of electrostatic field perturbations on ^{19}F NMR chemical shifts. Unfortunately, there were no "model compounds" that had experimental chemical shifts with which the calculations could be validated. Consequently we decided to validate the calculations in a different way, by carrying out a series of different computations of the electrostatic contributions to shielding, using fluorobenzene as the model target compound. In the first series of calculations (53) we investigated how the addition of hydrogen fluoride molecules influenced the ^{19}F NMR shielding in fluorobenzene, using "super-molecule" cluster calculations. These clusters contained one fluorobenzene molecule and up to five hydrogen fluoride molecules, oriented in different ways (HF or FH) along the molecular x, y, and z coordinates. The results of these calculations, on C_6H_5F-$(HF)_n$ where $n = 1–5$, clearly demonstrated the additivity of the intramolecular contribution to shielding, as expected for a purely electrical perturbation. We also investigated the effects of using point charges (no basis functions) to model the long-range purely electrostatic field effects on shielding. The results showed good accord between the super-molecule calculations and those using point charges. This is consistent with a purely electrostatic (or polarization) effect of the hydrogen fluoride molecules on the shielding in fluorobenzene, and is also consistent with the additivity of the intramolecular shift.

We then used a second approach, which involved determining the multipole shielding polarizabilities (MSP) of fluorobenzene: how the tensor elements change with electrical polarization (54). Dykstra used derivative Hartree-Fock methods to calculate the MSP tensor elements for the ^{19}F nucleus in fluorobenzene with respect to a uniform field, a field gradient, and a field-hypergradient. These shielding polarizability tensor elements, when multiplied by the appropriate field, field gradient, or hypergradient terms (which can be readily calculated from the known charge distributions, at least in the H-F molecule), yield the electrical contributions to shielding. A comparison of the multipole shielding polarizability expansion truncated at the point of linear field, field gradient, and hypergradient terms was found to be in excellent accord with benchmark ab initio results, both for a dipole and for a point-charge perturbation (54). On this basis, we concluded that super-molecule, point charge, and low-order MSP expansions permitted the accurate determination of the electrical contributions to shielding in model systems and should therefore be applicable to proteins as well.

To test these methods, we first investigated the ^{19}F NMR spectrum of a [5-F]-Trp-labeled galactose binding protein (in the presence of Ca^{2+} and galactose) from *Escherichia coli*, because specific assignments had previously been reported by Luck & Falke (55). Galactose binding protein contains five tryptophan residues at positions 127, 133, 183, 195, and 284. To compute the ^{19}F NMR chemical shifts in galactose binding protein, we used two techniques: the charge field perturbation-gauge including atomic orbital (CFP-GIAO) method (13) and

the MSP-local reaction field approach (56). In the first approach, we used ab initio Hartree-Fock GIAO methods to compute shieldings using (*a*) a radial cut-off with atomic resolution, (*b*) a radial cut-off with all residue resolution, and (*c*) a radial cut-off with all electroneutral residue resolution.

We first investigated shielding results for Trp-127. Using a cut-off radius with atomic resolution (only atoms within the cut-off being included in the calculation), there were large oscillations as a function of cut-off distance, owing to formal charge imbalances, as shown in Figure 4*A*. We therefore next investigated how shielding varied with radial cut-off, using whole residue resolution, which avoided artificially truncating molecules and allowed the presence of charged sidechains. The oscillations were much less apparent (Figure 4*A*). However, based on experimental results in which we had chemically converted the charged lysine sidechains in lysozyme to electroneutral N-acetyl groups and had found that there were essentially no chemical shift changes (51), it appeared that inclusion of full formal charges was unnecessary and might indeed contribute to the small residual oscillations seen experimentally. We therefore carried out a third series of calculations, using whole electroneutral residue resolution, and as shown in Figure 4*A* the oscillation was completely damped. We then used this approach to compute the chemical shifts of the four buried fluorotryptophan residues in galactose binding protein, as a function of the cut-off radius. These theoretical results successfully reproduced the full 10-ppm chemical shift nonequivalence seen experimentally (Figure 4*B*) together with the correct ordering and position of the four buried fluorotryptophan residues. However, the CFP-GIAO technique appeared incapable of successfully reproducing the chemical shift of the exposed tryptophan residue, Trp-284.

We therefore turned our attention to the use of the MSP–local reaction field technique, which includes the effects of molecular dynamics and solvation. We calculated a series of molecular dynamics trajectories and sampled the fields and field gradients at the fluorine nuclei at selected points along each trajectory. Multiplying the dipole and quadrupole shielding polarizabilities (computed previously) by the uniform field and electric field gradient tensor components yielded the purely electrical contributions to shielding, as a function of the time along the molecular dynamics trajectory. Selected results are shown in Figure 4*C*. The average electrical contributions to shielding (the theoretical chemical shifts or shieldings) are plotted as a function of the experimental chemical shifts in Figure 4*D*. This comparison reveals an excellent correlation and slope between predicted and experimental ^{19}F NMR chemical shifts for each of the five fluorotryptophan residues in galactose binding protein, including the solvent-exposed Trp-284. The CFP-GIAO calculations also show a good correlation for the four buried fluorotryptophan residues. However, the slope of the theory-versus-experiment correlation in that case is about 1.6, which we attributed to a dielectric constant effect for these buried residues (13).

To confirm that these results were not purely fortuitous, we investigated the ^{19}F NMR chemical shifts in a series of [4-F]-Trp-labeled myoglobins and

hemoglobins. In these systems the maximum observed chemical shift range induced by folding was smaller than in galactose binding protein, only 6.4 ppm. Using the MSP–local reaction field approach once again, we predicted the ^{19}F NMR chemical shifts and found that for fluorine atoms that did not have close contacts with neighboring groups there was a $\sim$1 ppm mean square deviation between experimental and theoretical chemical shifts (57): good agreement once again.

In the case of ^{19}F-labeled aliphatic amino acids in proteins the situation is more complex. Feeney and coworkers (58) found that a (2S,4S)-5-fluoroleucine-labeled dihydrofolate reductase from *Lactobacillus casei* exhibited a well resolved spectrum containing 12 peaks for the 13 leucine residues, covering a 15-ppm chemical shift range. For fluorine-labeled aliphatic carbons, the dipole and quadrupole shielding polarizabilities are expected to be smaller than in the more highly polarizable aromatic residues. The large range of chemical shifts seen experimentally could therefore not be explained solely in terms of electrical contributions to shielding. However, in the case of fluoroleucine it is clear that side-chain conformational differences (such as γ–gauche effects) would be expected to make major contributions to shielding . We therefore performed ^{19}F NMR chemical shift calculations of simple model systems and found that side-chain conformational differences can contribute up to 10 ppm to fluorine shielding. Similar calculations on a [5-F]-leucine-amide model indicated a 12-ppm difference in chemical shift between the two most likely leucine rotamers. For such aliphatic amino acids, the theoretical calculations predict about a 5–8 ppm maximal shift range owing to electrostatics, but up to a 12-ppm contribution to the chemical shift range owing to conformational effects. Both effects can be expected to contribute to fluorine NMR chemical shifts in fluoro-aliphatic amino acids in proteins, a somewhat more complex situation than is observed with fluoro-aromatic amino acid residues.

Interestingly, the same dihydrofolate reductase enzyme had previously been investigated by Feeney and coworkers, who used fluorotryptophan-labeled amino acids and found the very unusual result that a well resolved J-coupling was observable between two of the fluorotryptophan sites (59). These workers postulated, based on the magnitude of the J-coupling, that this must indicate that two of the Trp residues were close to each other in space, even though they were not close in the primary sequence. Such observations of long-range or through-space fluorine-fluorine J-couplings have been observed in the literature for many years in small organic molecules, and in recent work we have used DFT to try to predict them. We first investigated several of these small molecule systems and found excellent agreement between long-range J-couplings computed with DFT and those determined experimentally (60). In addition, we found that DFT techniques could also be used to estimate the J-coupling determined experimentally in dihydrofolate reductase (60). These results are of interest in the context of hydrogen bond J-couplings, because both are through-space interactions, as discussed in more detail below.

HEMEPROTEINS AND MODEL SYSTEMS

As described in the Introduction, in early experimental work we found strong correlations between the ^{13}C and ^{17}O NMR chemical shifts of the CO ligands in a wide range of carbonmonoxyheme proteins; these chemical shifts were also correlated with the ^{17}O electric field gradients and with the C–O and Fe–C vibrational stretch frequencies determined by infrared and Raman spectroscopy (5). These correlations are what would be expected if electrostatic polarization effects dominated the frequencies and electric field gradients seen experimentally, and in early work using only CO as a model it was found to be possible to reproduce all the experimental trends seen in proteins, with only electrical polarization (6). These correlations would almost certainly be expected to break down if there were large geometric distortions in Fe–C–O bonding geometry between the different proteins, but an examination of the literature revealed a that a wide range of such bonding geometries had actually been proposed, over the years, for different hemeproteins. Of particular interest, the bonding in carbonmonoxymyoglobin had been reported, based on X-ray crystallographic determinations, to be highly distorted, and this distortion had been proposed to be of particular importance in controlling the relative affinities of CO and O_2 molecule-binding to hemoglobin and myoglobin. Indeed, two modern biochemistry textbooks discuss this topic in some depth (61, 62). But is this view correct?

To investigate the question of Fe–C–O binding geometry in more detail, we undertook the synthesis of a broad range of heme model compounds and determined their structures using high-resolution small-molecule X-ray crystallography (63, 64). In particular, we synthesized carbonyl complexes of iron, ruthenium, and osmium tetraphenylporphyrinates and octaethylporphyrinates, complexed with pyridine and 1-methylimidazole. In all cases the metal-C–O geometry was very close to linear and was untilted. We carried out solid-state NMR determinations of the ^{13}C and ^{17}O NMR chemical shift tensors, together with an estimation of the ^{17}O electric field gradient (using nutation spectroscopy) and then used DFT to evaluate these properties (63, 64). As might now be expected, we found good agreement between theory and experiment. In addition, we used Mössbauer spectroscopy to determine the quadrupole splittings, and hence the electric field gradient, at the iron nuclei. Remarkably, the ^{13}C NMR shifts and shift anisotropy, the ^{17}O NMR chemical shift, the ^{17}O NMR electric field gradient, and the ^{57}Fe Mössbauer quadrupole splitting were all virtually identical to those found in the protein carbonmonoxymyoglobin. This strongly suggested the possibility that the ligand binding geometry in the carbonmonoxyheme proteins was also linear and untilted, just as it is in the model compound.

However, to investigate this question further it was necessary to determine just how all the spectroscopic observables would be influenced by changes in metal-ligand geometry. Perhaps they would not be sensitive to ligand distortions such as tilt and bend. We therefore computed ^{13}C and ^{17}O NMR chemical shielding surfaces as a function of ligand tilt (τ) and bend (β) (65), in much the same way that we

had previously computed Ramachandran shielding surfaces for amino acids as a function of the backbone torsion angles ϕ and ψ. These results clearly indicated that the experimentally observed correlations between carbon and oxygen spectra could not be generated by variations in ligand tilt and bend (65).

We next began to investigate the possible utility of ^{57}Fe NMR chemical shifts (66) and Mössbauer electric field gradients (67) as probes of hemeprotein structure, using DFT techniques. In earlier work, we had labeled myoglobin with ^{57}Fe and had obtained ^{57}Fe NMR chemical shifts for carbonmonoxymyoglobin and a series of alkylisocyanide-labeled myoglobins (68). Because the ^{57}Fe NMR chemical shifts and electric field gradients in ferrocytochrome c and in bis-pyridine and bis-trimethylphosphine Fe(II) porphyrins were also known, this provided a moderate database for chemical shift and Mössbauer quadrupole splitting/electric field gradient tensor calculations, especially when supplemented with additional Mössbauer and chemical shift results on much smaller molecule model compounds (67). We used DFT and a locally dense basis approach to evaluate the ^{57}Fe NMR shielding and electric field gradients at iron and compared these results with the experimentally measured chemical shifts and quadrupole splittings. We found good accord between the experimental and theoretical iron-57 chemical shifts ($R^2 = 0.992$, slope $= -0.984$), as well as for the Mössbauer quadrupole splittings ($R^2 = 0.975$, slope $= 1.04$, RMSD $= 0.18$ mm sec^{-1}), using a Wachters' all electron representation for iron, a 6-311++ G (2d) basis for all atoms directly attached to iron, 6-31G* for the second shell of atoms, and a 3-21G* basis for more distant atoms, together with a B3LYP hybrid exchange-correlation functional. However, we found extremely poor accord between theory and experiment for the ^{57}Fe NMR chemical shifts and electric field gradients when using carbonmonoxymyoglobin models with the highly distorted structures proposed in earlier X-ray crystallographic investigations (66, 67), clearly suggesting that the Fe–C–O bond geometry is very close to linear and untilted, both in solution and in the solid state.

We then investigated to what extent we could actually predict the experimental tilt and bend angles, using the Bayesian probability or Z-surface approach described above for the amino acid residues in peptides and proteins. In the case of carbonmonoxyhemeproteins, we had six observables: the ^{13}C NMR chemical shift, the ^{13}C NMR chemical shift anisotropy, the ^{17}O NMR chemical shift, the ^{17}O electric field gradient, the ^{57}Fe NMR chemical shift, and the ^{57}Fe Mössbauer quadrupole splitting (or electric field gradient). We therefore computed six property surfaces as a function of tilt and bend, $P(\tau, \beta)$, and used them together with the experimentally determined property values to predict a ^{6}Z probability surface. For clarity, we show only two such property and probability results in Figure 5. The results of adding additional property surfaces had little effect on τ, β predictions and confirmed that the ligand tilt and bend angles are very close to zero. In the case of the so-called A_o substate of carbonmonoxymyoglobin, we determined most probable tilt and bend angles of 0° and 1°, and in the A_1 substate of carbonmonoxymyoglobin we determined $\tau = 4^\circ$ and $\beta = 7^\circ$ (65).

A year after our publication, Kachalova and coworkers (69) reported a reinvestigation of the X-ray crystallographic structure of carbonmonoxymyoglobin,

using a synchrotron source. They reported results only 0.6° different from those we had determined by using the combination of NMR spectroscopy, Mössbauer spectroscopy, and quantum chemical calculations, confirming the correctness of our approach. In addition, deDios & Earle (70) have recently shown by using DFT methods on a charge balanced Fe(II)CO fragment that the ^{13}C NMR chemical shift/^{17}O NMR chemical shift/^{17}O electric field gradient/IR vibrational stretch frequency correlations observed previously can now all be very accurately reproduced by using the charge field perturbation technique, supporting the electrostatic origin of these correlations. In further recent work, Phillips and coworkers (71) have used Delphi (a finite-difference Poisson-Boltzmann algorithm) to compute the electrostatic potentials on CO in a variety of carbonmonoxymyoglobin mutants. They found a linear relationship between the infrared vibrational frequency and the potential, confirming once again the electrostatic origins of the vibrational frequency differences seen between different carbonmonoxyhemeproteins.

How then, are CO and O_2 differentiated in their binding to hemeproteins? Whereas there are undoubtedly a number of factors that contribute to the stabilization of oxygen binding (see e.g., Ref. 72 for detailed discussion), the results of density functional theory calculations on carbonmonoxy- and oxy-heme model compounds clearly indicate a much more negative electrostatic potential on the terminal oxygen in oxyheme when compared with carbonmonoxy-heme model compounds (73). That is to say, there is a much larger negative charge build-up on the terminal oxygen in oxyheme model compounds or oxyheme proteins than in the corresponding CO proteins, which can reasonably be thought to facilitate an electrostatic/hydrogen bonding interaction with the distal histidine residues found in hemoglobin and myoglobin. But how much confidence can be placed in such computations of charge density or electrostatic potential? That is my next topic.

HYDROGEN BONDING AND ELECTROSTATICS

The question of hydrogen bonding and electrostatics in proteins is of great importance from the standpoint of how the three-dimensional structure of a protein is held together, how enzymes function, how ligands such as CO and O_2 bind to proteins, how drugs bind, and so forth, and is clearly a topic worthy of a multivolume series. Nevertheless, when carrying out either ab initio Hartree-Fock or density functional theory (DFT) calculations of NMR chemical shifts, one can obtain, essentially as a by-product of the calculations, extensive details about electrostatics. The question then arises as to how much confidence should be placed in the accuracy of these results, because unlike the spectroscopic observables, electrostatic properties are almost always much more difficult to determine experimentally. We therefore wish to briefly review some recent results of the investigation of electrostatic properties in amino acids, including hydrogen bonding, that are of interest and relevant in the context of the observation of hydrogen bond J-couplings in proteins, as well as the observation of highly deshielded proton resonances in the active sites of many enzymes.

We have recently investigated the charge density, $\rho(\mathbf{r})$, its curvatures, $\partial^2\rho/\partial\mathbf{r}_{ij}$, the dipole moments, μ, the electrostatic potential, $\Phi(\mathbf{r})$, and the electric field gradients at the nitrogen sites, ∇E, in L-asparagine monohydrate, using ultra-high-resolution X-ray crystallography (a synchrotron source, area detector, 20°K sample temperature) and quantum chemical techniques (74). We used both ab initio Hartree-Fock and DFT methods with a variety of basis sets to calculate the electrostatic properties and used Bader's atoms in molecules (AIM) theory (75) to help analyze some of the experimental results. We found that the charge density, the dipole moment, and the molecular electrostatic potential, as well as the curvatures of the charge density at the bond critical points and at the hydrogen bond critical points, extracted from the high resolution X-ray diffraction results were in extremely good agreement with the values computed theoretically, using both Hartree-Fock and DFT methods (74). This gives us some confidence that the electrostatic properties, which can be readily obtained from the calculation, are accurate, a conclusion also reached by a number of workers investigating other amino acids (76, 77). For the electric field gradient, the results obtained from the crystallographic study were highly correlated with the theoretical results but had a poor slope. However, when using electric field gradient tensor elements determined from NMR spectroscopy, we found excellent agreement between calculation and experiment. We also used the icosahedral representation (mentioned above in the context of the chemical shielding tensor) to analyze the electric field gradient tensor as well as the Hessian-of-ρ(r) tensor, which provided a very rigorous evaluation of experimental versus theoretical-electrostatic property comparisons. Once again, we found excellent agreement between theory and experiment (74).

This initial investigation into hydrogen bonding and electrostatics clearly indicated that hydrogen bond charge densities at hydrogen bond critical points, hydrogen bond principal curvatures, and hydrogen bond Hessian tensors in the icosahedral representation could all be extracted from high resolution crystallographic data and that the values of these properties could all be accurately predicted by using either Hartree-Fock or DFT methods. This is an important observation because of recent debate of the question of the chemical nature of biological hydrogen bonds that has followed the observation of NMR scalar couplings across protein backbone hydrogen bonds. New ways of looking at hydrogen bonds are of interest because the existence of such $^{3h}J_{NC'}$ couplings has led some workers to conclude that even weak biological hydrogen bonds must be partially covalent. This assertion might appear surprising, given that through-space scalar couplings in simple organic molecules have long been recognized. Furthermore, as outlined above, we have been able to show via DFT calculations that through-space couplings between fluorine nuclei, J_{FF}, do indeed occur in the absence of a bond network, covalent or otherwise. We therefore began an investigation of protein hydrogen bonds in terms of both calculated Hessian-of-$\rho(\mathbf{r})$ tensors and experimental hydrogen bond scalar couplings and ^{1}H chemical shifts.

By calculating ρ(r) and $\partial^2\rho/\partial r_i r_j$ for protein backbone hydrogen bonds, we showed that the magnitude of $^{3h}J_{NC'}$ is an exponential function of the mutual

penetration of the nonbonding van der Waals shells of the isolated hydrogen bond donor and acceptor fragments, and not the partial covalence invoked previously (78). Using nonbonded fluoromethane dimer models for the calculation of J_{FF}, our results also showed that the magnitude of these through-space couplings exhibits the same exponential dependence upon the penetration of nonbonding monomer charge densities (Figure 6*A*). These results suggested that both the hydrogen bond $^{19}F–^{19}F$ J-coupling and the through-space J-coupling are subject to the same inductive mechanism.

Atoms-in-molecules theory is also useful in classifying (hydrogen-) bonding. It is straightforward to extract the relative contributions of both kinetic and potential energies to a bond from the calculation of $\partial^2\rho/\partial r_i r_j$ (75, 78). A local expression of the virial theorem relates the trace of the bond critical point Hessian tensor to the electronic kinetic energy density, G(r), and the electronic potential energy, V(r) (75):

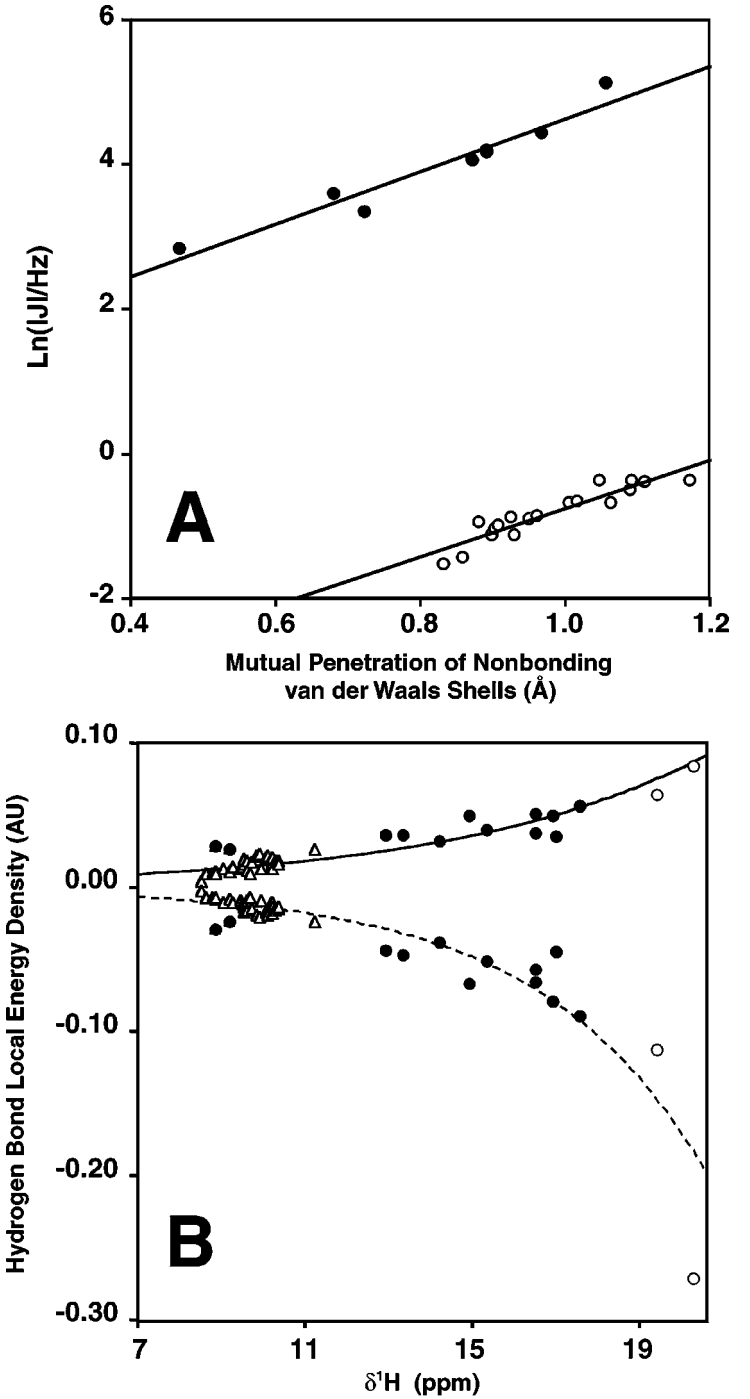
$$\frac{1}{4}\mathrm{Tr(Hessian)} = 2\mathrm{G(r)} + \mathrm{V(r)} \qquad 9.$$

This means one can determine which energy density is in excess of the 2:1 kinetic:potential average and thus determine the character of the bond. Stabilization resulting from a concentration of electronic charge in the bonding region is reflected by an excess of potential energy. Thus, closed-shell (electrostatic) bonds have a virial excess of kinetic energy, $|2G(r)| > |V(r)|$, whereas shared-electron (covalent) bonds have a virial excess of potential energy, $|V(r)| > |2G(r)|$. Furthermore, one can characterize a closed-shell situation in which $|2G(r)| > |V(r)|$, but $|V(r)| > |G(r)|$, as partially covalent. Using these definitions, we found that protein backbone hydrogen bonds are purely closed-shell, electrostatic interactions: $|2G(r)| > |V(r)|$ and $|G(r)| > |V(r)|$.

The chemical shift is also a very useful probe of the hydrogen bond, especially when combined with a quantum chemical/AIM analysis (78). Peptide backbone hydrogen bonds appear at high-field and correlate with a purely closed shell or electrostatic interaction (no covalence). Between 12 and 14 ppm partial covalent character begins to develop in the hydrogen bonds as $|V(r)| > |G(r)|$ and is clearly seen in enzyme short-strong hydrogen bonds and in some acids (78) (Figure 6*B*). This covalent character increases exponentially as the ^1H nucleus becomes more deshielded, and at ~21 ppm, the chemical shift region of experimentally observed low barrier hydrogen bonds, the hydrogen bond becomes a fully covalent, shared-electron interaction, with $|V(r)| > |2G(r)|$, as shown in Figure 6*B* (78).

TOWARDS DRUG DESIGN

Chemical shifts have also recently been used with great success as a tool with which to probe ligand binding to proteins in the drug discovery process. The idea is to use chemical shift changes on ligand binding to carry out high throughput



screening of small molecule libraries to discover lead compounds (typically with micromolar affinity constants) and then to use NMR or X-ray crystallographic structural information to help appropriately link two (or more) such compounds to provide inhibitors with nanomolar binding constants.

In its earliest embodiment (79), ^{15}N and ^{1}H amide chemical shifts obtained from two-dimensional spectra were used to find an initial lead compound, based on chemical shift changes observed on ligand binding. Ten small ligand molecules were added at a time to a protein solution, which enabled the screening of ~1000 compounds in a reasonable time period. Once a promising lead compound was identified, this compound could be optimized. A second series of screenings was carried out (possibly in the presence of the first, optimized lead compound) to detect a second ligand binding site, again based solely on changes in chemical shift owing to ligand binding. Optimization of the second lead compound, using high-resolution solution NMR or X-ray crystallographic studies, was then carried out to determine the location of the two binding ligands, a strategy that facilitates the relatively precise engineering of a suitable tether to chemically link the two individual ligands. When two micromolar binding ligands are tethered together a typically nanomolar binding inhibitor is obtained. For example, a tight binding inhibitor ($K_d = 19$ nM) of the FK506 binding protein was discovered by linking two weak binding inhibitors ($K_d = 2\ \mu$M, 100 μM) that had been identified and optimized in less than two months, and it was only necessary to synthesize five tethered molecules to obtain the nanomolar inhibitor. In another example, a nanomolar nonpeptide inhibitor of stomelysin, a matrix metalloproteinase, was produced by linking together 17-mM and 20-μM inhibitors, and it has been pointed out that conventional high throughput screening of more than 100,000 compounds against stomelysin failed to produce a single inhibitor with better than a 10-μM K_d (80).

In early reports, this approach was limited to screening about 1000 compounds per day, a rather low throughput by high throughput screening standards, because relatively high protein concentrations were required to obtain NMR spectra in reasonable time periods. In addition, high protein concentrations limited the total number of compounds that could be screened in a given spectrum, based on solubility considerations. More recently, the use of cryoprobe technology has decreased typical sample concentrations required for data acquisition (81) and it has been reported that mixtures containing 100 compounds can be screened in a single experiment

←

Figure 6 (*A*) Dependence of the magnitude of through-space J-couplings upon the mutual penetration of nonbonding van der Waals shells, $\Sigma\Delta r$. Filled circles, J_{FF}; open circles, $^{3h}J_{NC}$. (*B*) Dependence of the hydrogen bond local energy densities upon the proton chemical shift. Open circles, low barrier hydrogen bonds; filled circles, short-strong hydrogen bonds; triangles, peptide backbone hydrogen bonds; solid line, $G(r) = 0.0066 \exp[0.12]$, $R^2 = 0.87$; dashed line, $V(r) = -0.0012 \exp[0.24]$, $R^2 = 0.82$. Based on data in Ref. 78.

at protein concentrations of 50 μM, enabling libraries of more than 200,000 compounds to be investigated in less than one month. This chemical shift–based approach to drug design also has considerable advantages over more standard combinatorial chemistry approaches because far fewer compounds need to be synthesized, because linker compounds can be specifically tailored based on known NMR or X-ray crystallographic structural information. Moreover, the technique has the advantage that no specific high throughput screening assays need to be developed.

This approach has recently been extended to the use of ^{13}C-labeled proteins. The method involves ^{13}C-labeling of the methyl groups of valine, leucine, and isoleucine (δ_1) residues from inexpensively produced ^{13}C α-ketoisovalerate and α-ketobutyrate and results in a sensitivity increase of almost a factor of three when compared with the ^{15}N/^{1}H chemical shift screening approach (82), owing to the favorable relaxation properties of methyl groups and the fact that three protons are attached to each ^{13}C-labeled methyl group. Whereas fewer resonances are monitored in the ^{1}H/^{13}C-experiment, an analysis of 191 crystal structures of proteins with bound ligands showed that 92% of the ligands were within 6 Å of at least one methyl carbon in valine, leucine, or isoleucine-δ_1, but only 82% of the ligands were within 6 Å of a backbone nitrogen. Moreover, even high molecular weight protein targets, such as dihydroneopterin aldolase (MW = 110 kDa), could be investigated using this chemical shift approach (82).

ACKNOWLEDGMENTS

I thank A.C. de Dios, C.J. Jameson, and P. Pulay for their contributions to the work reported here, and W. Arnold for help with preparing this article. This work was supported by the U.S. National Institutes of Health (Grants GM-50694, HL-19481) and by the National Computational Science Alliance (Grant MCB000018N).

LITERATURE CITED

1. Meadows DH, Markley JL, Cohen JS, Jardetzky O. 1967. *Proc. Natl. Acad. Sci. USA* 58:1307–13
2. Allerhand A, Childers RF, Oldfield E. 1973. *Biochemistry* 12:1335–41
3. Allerhand A, Oldfield E. 1973. *Biochemistry* 12:3428–33
4. Spera S, Bax A. 1991. *J. Am. Chem. Soc.* 113:5490–92
5. Park KD, Guo K, Adebodun F, Chiu ML, Sligar SG, Oldfield E. 1991. *Biochemistry* 30:2333–47
6. Augspurger JD, Dykstra CE, Oldfield E. 1991. *J. Am. Chem. Soc.* 113:2447–51
7. Osapay K, Case DA. 1991. *J. Am. Chem. Soc.* 113:9436–44
8. Janes N, Ganapathy S, Oldfield E. 1983. *J. Magn. Reson.* 54:111–21
9. Frydman L, Chingas GC, Lee YK, Grandinetti PJ, Eastman MA, et al. 1992. *Isr. J. Chem.* 32:161–64
10. Wolinski K, Hinton JF, Pulay P. 1990. *J. Am. Chem. Soc.* 112:8251–60
11. deDios AC, Laws DD, Oldfield E. 1994. *J. Am. Chem. Soc.* 116:7784–86

12. Alderman DW, Sherwood MH, Grant DM. 1993. *J. Magn. Reson.* 101:188–97
13. deDios AC, Pearson JG, Oldfield E. 1993. *Science* 260:1491–96
14. Chesnut DB, Moore KD. 1989. *J. Comput. Chem.* 10:648–59
15. deDios AC, Pearson JG, Oldfield E. 1993. *J. Am. Chem. Soc.* 115:9768–73
16. deDios AC, Oldfield E. 1994. *J. Am. Chem. Soc.* 116:5307–14
17. Laws DD, Le H, deDios AC, Havlin RH, Oldfield E. 1995. *J. Am. Chem. Soc.* 117:9542–46
18. Havlin RH, Le H, Laws DD, deDios AC, Oldfield E. 1997. *J. Am. Chem. Soc.* 119:11951–58
19. Heller J, Laws DD, King DS, Wemmer DE, Pines A, et al. 1997. *J. Am. Chem. Soc.* 119:7827–31
20. Szabo CM, Sanders LK, Arnold W, Grimley JS, Godbout N, et al. 1999. In *ACS Monograph on Chemical Shielding*, ed. A deDios, J Facelli, pp. 40–62
21. Tjandra N, Bax A. 1997. *J. Am. Chem. Soc.* 119:9576–77
22. Sitkoff D, Case DA. 1998. *Prog. NMR Spectrosc.* 32:165–90
23. Hong M. 2000. *J. Am. Chem. Soc.* 122:3762–70
24. Huster D, Yamaguchi S, Hong M. 2000. *J. Am. Chem. Soc.* 122:11320–27
25. Cole HB, Sparks SW, Torchia DA. 1988. *Proc. Natl. Acad. Sci. USA* 85:6362–65
26. McDermott A, Polenova T, Bockmann A, Zilm KW, Paulsen EK, et al. 2000. *J. Biomol. NMR* 16:209–19
27. Sanders L, Oldfield E. Unpublished results
28. Schafer L, Newton SQ, Cao M, Peeters A, Van Alsenoy C, et al. 1993. *J. Am. Chem. Soc.* 115:272–80
29. deDios AC, Oldfield E. 1994. *J. Am. Chem. Soc.* 116:11485–88
30. Walling AE, Pargas RE, deDios AC. 1997. *J. Phys. Chem. A* 101:7299–303
31. Le H, Oldfield E. 1994. *J. Biomol. NMR* 4:341–48
32. Le H, Oldfield E. 1996. *J. Phys. Chem.* 100:16423–28
33. Marassi FM, Opella SJ. 2000. *J. Magn. Reson.* 144:150–55
34. Wang J, Denny J, Tian C, Kim S, Mo Y, et al. 2000. *J. Magn. Reson.* 144: 162–67
35. Fushman D, Tjandra N, Cowburn D. 1998. *J. Am. Chem. Soc.* 120:10947–52
36. Kroenke CD, Rance M, Palmer AG III. 1999. *J. Am. Chem. Soc.* 121:10119–25
37. Kurland RJ, McGarvey BR. 1970. *J. Magn. Reson.* 2:286–301
38. Qi PX, Beckman RA, Wand AJ. 1996. *Biochemistry* 35:12275–86
39. Wilkens SJ, Xia B, Weinhold F, Markley JL, Westler WM. 1998. *J. Am. Chem. Soc.* 120:4806–14
40. Wilkins SJ, Westler WH, Markley JL, Weinhold F. 1999. *J. Inorg. Biochem.* 74: 338
41. Van Alsenoy C, Yu C-H, Peeters A, Martin JML, Schäfer L. 1998. *J. Phys. Chem. A* 102:2246–51
42. Pearson JG, Le H, Sanders L, Godbout N, Havlin RH, Oldfield E. 1997. *J. Am. Chem. Soc.* 119:11,941–50
43. Pearson JG, Wang J-F, Markley JL, Le H, Oldfield E. 1995. *J. Am. Chem. Soc.* 117:8823–29
44. Tjandra N, Bax A. 1997. *Science* 278: 1111–14
45. Chou JJ, Li S, Bax A. 2000. *J. Biomol. NMR* 18:217–27
46. Cornilescu G, Delaglio F, Bax A. 1999. *J. Biomol. NMR* 13:289–302
47. Wishart DS, Sykes BD, Richards FM. 1992. *Biochemistry* 31:1647–51
48. Wishart DS, Nip AM. 1998. *Biochem. Cell Biol.* 76:153–63
49. Bode J, Blumenstein M, Raftery MA. 1975. *Biochemistry* 14:1153–60
50. Sykes BD, Weingarten HI, Schlesinger MJ. 1974. *Proc. Natl. Acad. Sci. USA* 71:469–73
51. Lian C, Le H, Montez B, Patterson J, Harrell S, et al. 1994. *Biochemistry* 33: 5238–45
52. deDios AC, Oldfield E. 1994. *J. Am. Chem. Soc.* 116:7453–54

53. deDios AC, Oldfield E. 1993. *Chem. Phys. Lett.* 205:108–16
54. Augspurger JD, deDios AC, Oldfield E, Dykstra CE. 1993. *Chem. Phys. Lett.* 213: 211–16
55. Luck LA, Falke JJ. 1991. *Biochemistry* 30:4248–56
56. Pearson JG, Oldfield E, Lee FS, Warshel A. 1993. *J. Am. Chem. Soc.* 115:6851–62
57. Pearson JG, Montez B, Le H, Oldfield E, Chien EYT, Sligar SG. 1997. *Biochemistry* 36:3590–99
58. Feeney J, McCormick JE, Bauer CJ, Birdsall B, Moody CM, et al. 1996. *J. Am. Chem. Soc.* 118:8700–6
59. Kimber BJ, Feeney J, Roberts GCK, Birdsall B, Griffiths DV, et al. 1978. *Nature* 271:184–85
60. Arnold WD, Mao J, Sun H, Oldfield E. 2000. *J. Am. Chem. Soc.* 122:12164–68
61. Stryer L. 1995. *Biochemistry*. New York: Freeman. 4th ed.
62. Garrett RH, Grisham CM. 1999. *Biochemistry*. Orlando, FL: Harcourt Brace. 2nd ed.
63. Salzmann R, Ziegler CJ, Godbout N, McMahon MT, Suslick KS, Oldfield E. 1998. *J. Am. Chem. Soc.* 120:11323–34
64. Salzmann R, McMahon MT, Godbout N, Sanders LK, Wojdelski M, Oldfield E. 1999. *J. Am. Chem. Soc.* 121:3818–28
65. McMahon MT, deDios AC, Godbout N, Salzmann R, Laws DD, et al. 1998. *J. Am. Chem. Soc.* 120:4784–97
66. Godbout N, Havlin R, Salzmann R, Debrunner PG, Oldfield E. 1998. *J. Phys. Chem. A* 102:2342–50
67. Havlin RH, Godbout N, Salzmann R, Wojdelski M, Arnold W, et al. 1998. *J. Am. Chem. Soc.* 120:3144–51
68. Chung J, Lee HC, Oldfield E. 1990. *J. Magn. Reson.* 90:148–57
69. Kachalova GS, Popov AN, Bartunik HD. 1999. *Science* 284:473–76
70. deDios AC, Earle EM. 1997. *J. Phys. Chem. A* 101:8132–34
71. Phillips GN Jr, Teodoro M, Li TS, Smith B, Olson JS. 1999. *J. Phys. Chem. B* 103: 8817–29
72. Springer BA, Sligar SG, Olson JS, Phillips GN Jr. 1994. *Chem. Rev.* 94:699–714
73. Godbout N, Sanders LK, Salzmann R, Havlin RH, Wojdelski M, Oldfield E. 1999. *J. Am. Chem. Soc.* 121:3829–44
74. Arnold WD, Sanders LK, McMahon MT, Volkov AV, Wu G, et al. 2000. *J. Am. Chem. Soc.* 122:4708–17
75. Bader RFW. 1990. *Atoms in Molecules—A Quantum Theory*. Oxford: Clarendon
76. Koritsanszky T, Flaig R, Zobel D, Krane H-G, Morgenroth W, Luger P. 1998. *Science* 279:356–58
77. Flaig R, Koritsanszky T, Zobel D, Luger P. 1998. *J. Am. Chem. Soc.* 120:2227–38
78. Arnold WD, Oldfield E. 2000. *J. Am. Chem. Soc.* 122:12835–41
79. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. 1996. *Science* 274:1531–34
80. Hajduk PJ, Meadows RP, Fesik SW. 1997. *Science* 278:497–99
81. Hajduk PJ, Gerfin T, Boehlen J-M, Häberli M, Marek D, Fesik SW. 1999. *J. Med. Chem.* 42:2315–17
82. Hajduk PJ, Augeri DJ, Mack J, Mendoza R, Yang J, et al. 2000. *J. Am. Chem. Soc.* 122:7898–904

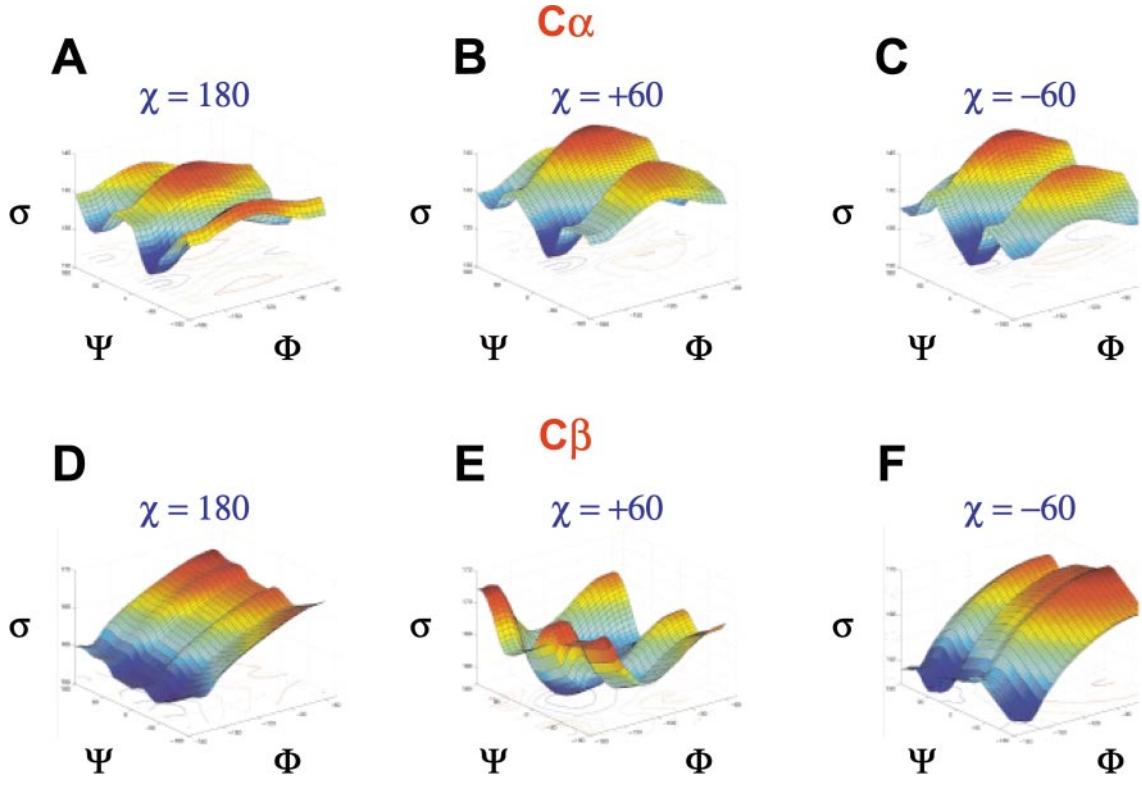


Figure 3 Calculated $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shieldings in N-formylvaline amide as a function of ϕ and ψ (at various χ^1 values). (*A*) C^{α}, $\chi^1 = 180^{\circ}$; (*B*) C^{α}, $\chi^1 = 60^{\circ}$; (*C*) C^{α}, $\chi^1 = -60^{\circ}$; (*D*) C^{β}, $\chi^1 = 180^{\circ}$; (*E*) C^{β}, $\chi^1 = 60^{\circ}$; (*F*) C^{β}, $\chi^1 = -60^{\circ}$. Reproduced with permission from *J. Am. Chem. Soc*. 1995. 117:9542–46. Copyright 1995 Am. Chem. Soc.

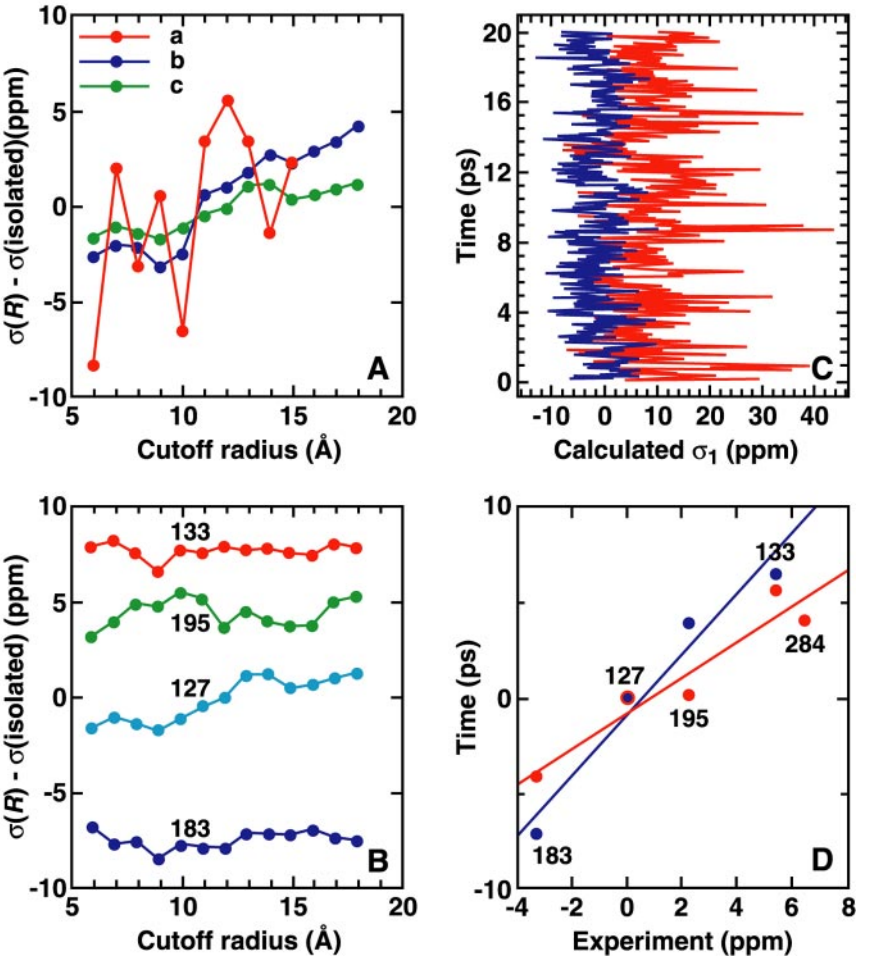


Figure 4 Fluorine shielding results for the Trp residues in the *Escherichia coli* galactose binding protein (GBP). (*A*) Ab initio [gauge-including-atomic-orbital (GIAO)] results with the use of a radial cutoff with atomic resolution (*a*, *red*), a radial cutoff with whole residue resolution (*b*, *blue*), and a radial cutoff with whole, electroneutral residue resolution (*c*, *green*). (*B*) GIAO shielding results for Trp residues 127, 133, 183, and 195 with the use of whole electroneutral residue resolution as a function of radial cutoff. (*C*) Results for 20-ps shielding trajectories for [5-F]Trp^{183} (*blue*) and Trp^{284} (*red*) residues in GBP. (*D*) Graph showing experimental versus theoretical shielding results in GBP: charge-field perturbation—GIAO (*blue*) and *red*, multipole shielding polarizability—local reaction field method. Reprinted with permission from *Science*, 260:1491–96 (1993). Copyright 1993 Am. Assoc. Adv. Sci.

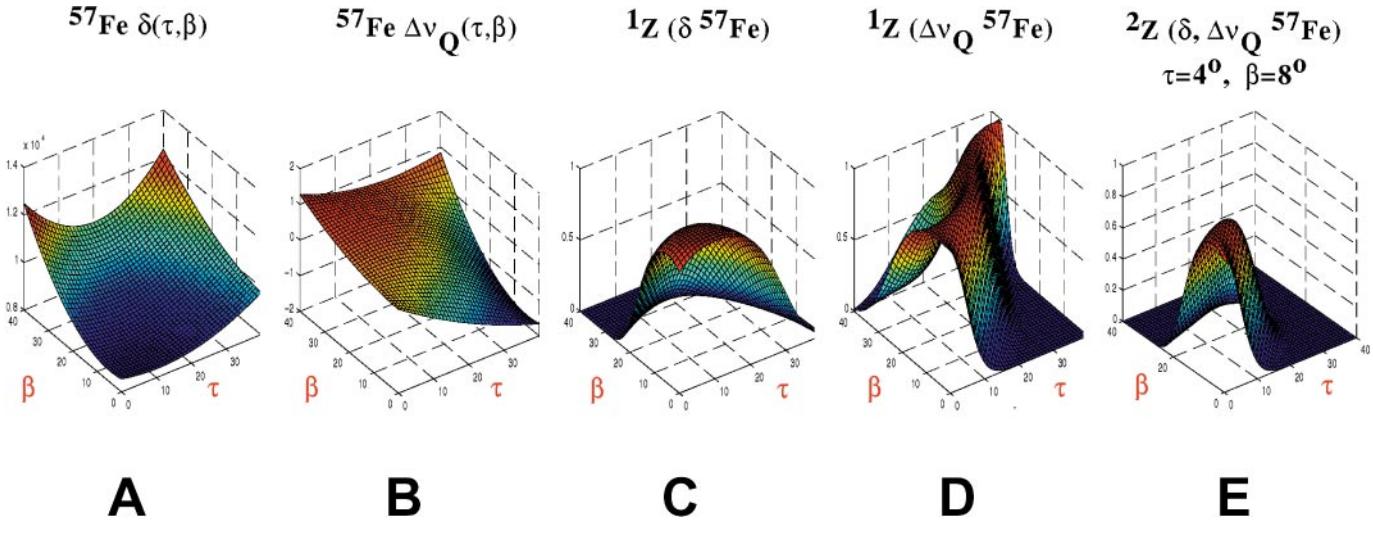


Figure 5 Property and τ,β probability surfaces for CO heme model system. (*A*) ^{57}Fe NMR chemical shift τ,β surface; (*B*) ^{57}Fe Mössbauer quadrupole splitting surface; (*C*) ^{1}Z chemical shift probability surface; (*D*) ^{1}Z quadrupole splitting probability surface; (*E*) ^{2}Z probability surface for tilt and bend, the product of (*C*) and (*D*) surfaces. Based on data in Reference 65.

Annual Review of Physical Chemistry
Volume 53, 2002

CONTENTS

Frontispiece—*Ignacio Tinoco, Jr.* xiv

PHYSICAL CHEMISTRY OF NUCLEIC ACIDS, *Ignacio Tinoco, Jr.* 1

HIGHER-ORDER OPTICAL CORRELATION SPECTROSCOPY IN LIQUIDS, *John T. Fourkas* 17

TIME-RESOLVED PHOTOELECTRON ANGULAR DISTRIBUTIONS: CONCEPTS, APPLICATIONS, AND DIRECTIONS, *Tamar Seideman* 41

SCATTERING RESONANCES IN THE SIMPLEST CHEMICAL REACTION, *Félix Fernández-Alonso and Richard N. Zare* 67

VACUUM ULTRAVIOLET SPECTROSCOPY AND CHEMISTRY BY PHOTOIONIZATION AND PHOTOELECTRON METHODS, *Cheuk-Yiu Ng* 101

THE MOLECULAR HAMILTONIAN, *Henning Meyer* 141

REVERSIBLE POLYMERIZATIONS AND AGGREGATIONS, *Sandra C. Greer* 173

SCANNING TUNNELING MICROSCOPY STUDIES OF THE ONE-DIMENSIONAL ELECTRONIC PROPERTIES OF SINGLE-WALLED CARBON NANOTUBES, *Min Ouyang, Jin-Lin Huang, and Charles M. Lieber* 201

ELECTRON TRANSFER AT MOLECULE-METAL INTERFACES: A TWO-PHOTON PHOTOEMISSION STUDY, *X.-Y. Zhu* 221

AB INITIO MOLECULAR DYNAMICS WITH DENSITY FUNCTIONAL THEORY, *John S. Tse* 249

TRANSITION PATH SAMPLING: THROWING ROPES OVER ROUGH MOUNTAIN PASSES, IN THE DARK, *Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler* 291

ELECTRONIC STRUCTURE AND CATALYSIS ON METAL SURFACES, *Jeff Greeley, Jens K. Nørskov, and Manos Mavrikakis* 319

CHEMICAL SHIFTS IN AMINO ACIDS, PEPTIDES, AND PROTEINS: FROM QUANTUM CHEMISTRY TO DRUG DESIGN, *Eric Oldfield* 349

REACTIVE COLLISIONS OF HYPERTHERMAL ENERGY MOLECULAR IONS WITH SOLID SURFACES, *Dennis C. Jacobs* 379

MOLECULAR THEORY OF HYDROPHOBIC EFFECTS: "SHE IS TOO MEAN TO HAVE HER NAME REPEATED," *Lawrence R. Pratt* 409

STUDIES OF POLYMER SURFACES BY SUM FREQUENCY GENERATION VIBRATIONAL SPECTROSCOPY, *Zhan Chen, Y. R. Shen, and Gabor A. Somorjai* 437

QUANTUM MECHANICAL METHODS FOR ENZYME KINETICS, *Jiali Gao and Donald G. Truhlar* 467

SURFACE FEMTOCHEMISTRY: OBSERVATION AND QUANTUM CONTROL OF FRUSTRATED DESORPTION OF ALKALI ATOMS FROM NOBLE METALS, *Hrvoje Petek and Susumu Ogawa* 507

CONNECTING LOCAL STRUCTURE TO INTERFACE FORMATION: A MOLECULAR SCALE VAN DER WAALS THEORY OF NONUNIFORM LIQUIDS, *John D. Weeks* 533

INDEXES

Author Index 563
Subject Index 591
Cumulative Index of Contributing Authors, Volumes 49–53 623
Cumulative Index of Chapter Titles, Volumes 49–53 625

ERRATA

An online log of corrections to *Annual Review of Physical Chemistry* chapters may be found at http://physchem.annualreviews.org/errata.shtml