Journal of Biomolecular NMR, 5 (1995) 217–225
ESCOM

J-Bio NMR 280

Chemical shifts and three-dimensional protein structures

Eric Oldfield

Department of Chemistry, University of Illinois at Urbana-Champaign, 505 South Mathews Avenue, Urbana, IL 61801, U.S.A.

Received 24 February 1995
Accepted 24 March 1995

Keywords: Chemical shifts; Chemical shift tensors; Ab initio calculations; Structure refinement and prediction; Electrostatics

Summary

During the past three years it has become possible to compute ab initio the ^{13}C, ^{15}N and ^{19}F NMR chemical shifts of many sites in native proteins. Chemical shifts are beginning to become a useful supplement to more established methods of solution structure determination, and may find utility in solid-state analysis as well. From ^{13}C NMR, information on ϕ, ψ and χ torsions can be obtained, permitting both assignment verification, and structure refinement and prediction. For ^{15}N, both torsional and hydrogen-bonding effects are important, while for ^{19}F, chemical shifts are primarily indicators of the local charge field. Chemical shift calculations are still slow, but shielding hypersurfaces – the shift as a function of the dihedral angles that define the molecular conformation – are becoming accessible. Over the next few years, theoretical and computer hardware improvements will enable more routine use of chemical shifts in structural studies, including the study of metal–ligand interactions, the analysis of drug and substrate binding and catalysis, the study of folding/unfolding pathways, as well as the characterization of conformational substates. Rather than simply being a necessary prerequisite for multidimensional NMR, chemical shifts and chemical shift non-equivalence due to folding are now beginning to be useful for structural characterization.

Introduction

The chemical shift is unique in the sense that if a resolved and assigned NMR spectrum can be obtained, then chemical shift information is available, and in principle structural information should follow. In a recent Perspective, Wagner (1993) has discussed the prospects for NMR of large proteins, where line broadening and spin diffusion become major problems, restricting the types of pulse sequence which can be employed, and assignments also become more difficult as size increases. While understanding chemical shifts does not solve the line-broadening problem, there is the possibility that at least some structural information can be gleaned from what is available, and this may help partially offset problems in other areas. In the solid or semisolid state, linebroadening and spin-diffusion effects are even more severe, so again, understanding chemical shifts could make a useful contribution to structure determination, especially if it were possible to accurately calculate shielding tensors for use with oriented sample NMR experiments.

Understanding chemical shifts could also hold considerable promise for drug design (Fesik, 1993). For example, many lead compounds bind tightly to their receptors, and if well-defined shift–structure correlations could be developed, then useful conformational information could be readily obtained, either in solution or in the solid state. Likewise, many ligands cause very discrete, localized changes in protein ^{13}C and ^{15}N chemical shifts on binding to their target. These '$\Delta\delta$' values are typically displayed as a function of sequence position, but have never been fully analyzed. Understanding the relationships between shift and structure may offer a rapid way of analyzing what '$\Delta\delta$' means.

Thus, there appears to be an essentially unlimited scope for using chemical shifts in protein (as well as nucleic acid and carbohydrate) structural studies in biology, biochemistry, pharmacology and biomedical research. What is needed is a workable theory of chemical shifts.

Shift prediction

For ^{1}H NMR, over the years empirical methods have been quite successful, and have permitted the first chemical shift refinements of structure (Ösapay et al., 1994), but for the heavier elements ^{13}C, ^{15}N and ^{19}F, there has until recently been much less success in predicting chemical shifts. The reason for this is that the paramagnetic term in the Ramsey equation (Ramsey, 1950,1952) dominates shielding, and this is difficult to evaluate since it involves a consideration of excited states, and has to be evaluated ab initio. Nevertheless, empirical approaches are of use in detecting helical and sheet secondary structural elements in proteins, using for example the Wishart Chemical Shift Index (Wishart et al., 1992), and clear trends in helix and sheet C^{α} and C^{β} shifts have been shown in the work of Spera and Bax (1991), leading to the idea that it should be possible to relate chemical shifts and three-dimensional protein structure – the theme of this Perspective. The question is, then, how can this be done? One approach is to use quantum chemistry.

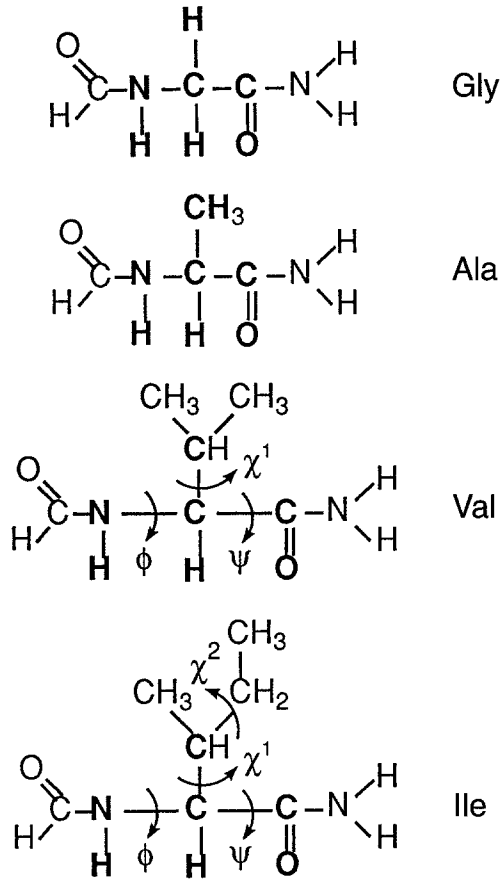
The methods used by theoreticians to compute chemical shifts have developed steadily in speed and accuracy over the years. Most of them are based upon the self-consistent field (SCF) approximation, which is invoked to first evaluate the total molecular eigenenergy, E, and then the second-derivative property, the chemical shielding (σ), using either a gauge-including atomic orbital (London, 1937; Ditchfield, 1972; Wolinski et al., 1990) or an individual gauge for localized orbitals method (Kutzelnigg et al., 1993). Basically similar approaches are in principle available for evaluating J-couplings and magnetic susceptibilities, χ:

$$\sigma_{\alpha\beta} = \left.\frac{\partial^2 E}{\partial \mu_\alpha \partial B_\beta}\right|_{\mu_\alpha = B_\beta = 0} \quad (1)$$

$$J_{\alpha\beta} = \left.\frac{\partial^2 E}{\partial \mu_\alpha \partial \mu_\beta}\right|_{\mu_\alpha = \mu_\beta = 0} \quad (2)$$

$$\chi_{\alpha\beta} = \left.\frac{\partial^2 E}{\partial B_\alpha \partial B_\beta}\right|_{B_\alpha = B_\beta = 0} \quad (3)$$

where $\sigma_{\alpha\beta}$, $J_{\alpha\beta}$ and $\chi_{\alpha\beta}$ are the $\alpha\beta$ components of the chemical shielding tensor, the spin–spin coupling tensor and the magnetic susceptibility tensor, respectively, and μ_α,μ_β and B_α,B_β are the α,β components of the magnetic moment μ and the magnetic field, B. Suitable programs are available from P. Pulay (University of Arkansas) and W. Kutzelnigg (Ruhr Universität, Bochum), and typically require a 20 Mflop or faster CPU, 3 GB of disc space, and 32–64 MB of memory. These programs compute absolute shielding from the bare nucleus, σ, while



Scheme 1. Small fragments used to evaluate chemical shifts in proteins.

what is experimentally measured is the shift, δ, from a given reference. Relationships between the two are well known.

At present, the chemical shifts of complete proteins cannot be evaluated, since the computational times scale approximately as $N^{3.5}$, where N is the number of basis functions used to construct the atomic orbitals. Thus, small fragments, such as those depicted in Scheme 1, need to be constructed. Fortunately, such fragments are sufficiently large that they do carry structural information – for example ϕ, ψ and χ^1 shift dependencies for the valine fragment – but they are sufficiently small that ab initio shielding calculations can be carried out in acceptable periods of time, with the results comparing well with experiment. For example, an excellent correlation is found between theoretically predicted shielding and experimental chemical shifts for C^{α} and C^{β} sites in the alanine and valine residues in calmodulin (de Dios and Oldfield, 1994a), or for the alanines in staphylococcal nuclease (SNase) (de Dios et al., 1993b).

In general, one is interested in obtaining multidimensional shielding surfaces $\sigma(\phi,\psi)$, $\sigma(\phi,\psi,\chi)$ and $\sigma(\phi,\psi,\chi^1,\chi^2)$ in order to first validate assignments based on structural knowledge or, as described below, to begin to extract structural information from observed chemical shifts. This

requires improved computational capabilities, or alternative approaches to shielding calculations. At present, a large basis calculation on an alanine fragment takes about 6 h on a 20 Mflop machine, for a single chemical shift (on an IBM RISC/6000 Model 365). Recent Power-2/Power Indigo[2] CPUs promise about a factor of 4–6 speed improvement, but it will be several more years before full 3D/4D surfaces can be evaluated for all amino acids. However, one must question whether large basis sets are essential. For accurate *absolute* shieldings (i.e. relative to the bare nucleus), they *are* essential, but simply for deducing e.g. ϕ,ψ information, this is unlikely to be the case, since a slope/offset error due to a basis deficiency can be easily corrected. That is to say, it is primarily the *shape* of a Ramachandran shielding surface which encodes the structural information. For example, a large offset in absolute shielding in a very small basis can usually be corrected for by a large basis calculation, or via comparison with experiment when a large body of data is available. For alanine, use of a small, locally dense basis (6–31G++ (2d,2p) on C^{α} and N; 4–31G elsewhere) reduces the computational time from 6 h to ~1 h, and this value should drop to ~10 min within a year or two (using R10 000 MIPS processors). When comparing alanine surfaces evaluated with large versus small bases, the rmsd is ~0.4 ppm, approaching the accuracy of measurement, especially when considering the current idiosyncrasies in protein chemical shift referencing.

With the presently available CPUs and code, evaluation of two- and three-dimensional shielding surfaces is already possible, albeit slow. Figure 1 shows recent results for C^{α} and C^{β} for three valine fragments ($\chi^1 = \pm 60°, 180°$). These results clearly indicate why empirical approaches will be rather limited, since the $\sigma(\phi,\psi,\chi)$ surfaces are quite different, depending on which χ^1 is chosen. For longer side chains, $\sigma(\phi,\psi,\chi^1,\chi^2)$ hypersurfaces are in principle required for structure prediction and refinement. Since generally only limited ranges of χ^1,χ^2 values are found in proteins (Ponder and Richards, 1987), at least good initial progress can be expected using a much more limited series of surfaces, representing the common χ^i conformers.

Correlation and accuracy

In essentially all of the cases of ^{13}C chemical shift calculations in proteins that have been investigated, absolute shieldings are systematically in error by up to 5–10 ppm. For work with, say, a valine C^{α}, this is unimportant since data can be offset to fit a large database of known structure/shift correlations. This essentially corresponds to using a reference shift such as a random coil shift; in fact, the shielding of TMS can be evaluated ab initio, and the error in absolute shielding with a given basis used to deduce the correct absolute shielding. Errors in absolute shielding can arise either from the lack of basis saturation (not enough functions), or from the neglect of electron correlation, the instantaneous electron–electron interaction which is omitted in the SCF approximation. Post-Hartree–Fock methods, such as MP2-GIAO (Gauss, 1992) or MC-IGLO (Kutzelnigg et al., 1993), do include correlation, but scale as N^5 rather than N^4, where N is the number of basis functions – making their use impractical for protein work.

However, electron correlation can be handled in an

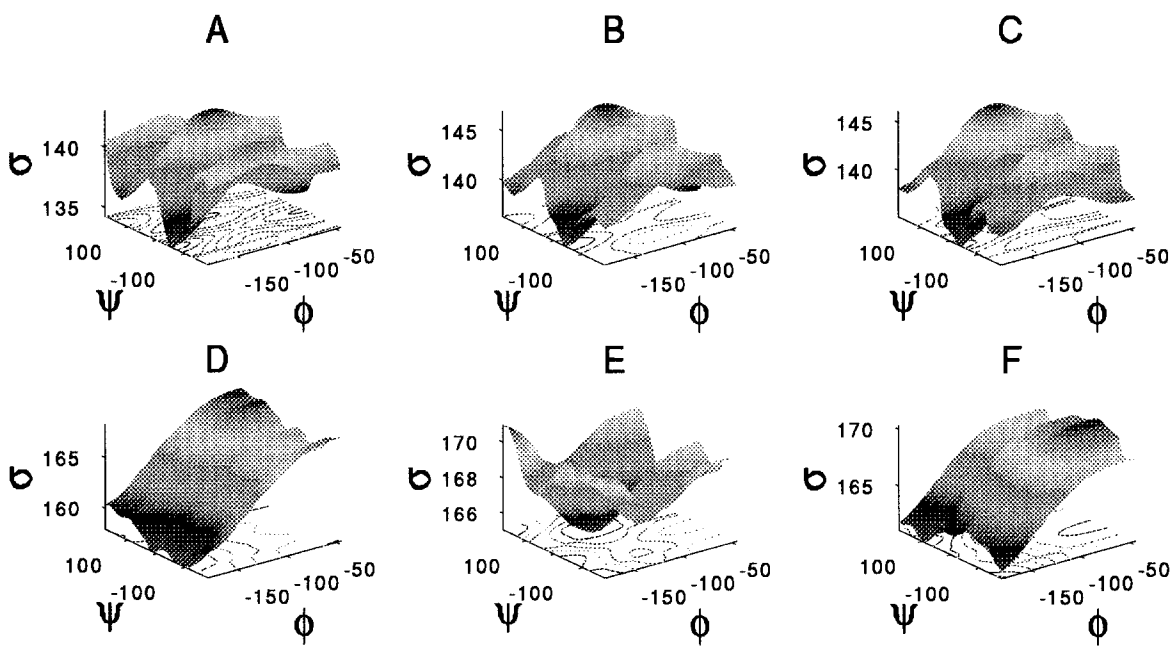


Fig. 1. Calculated $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shieldings in formylvaline amide as a function of ϕ and ψ (at various χ^1 values): (A) C^{α}, $\chi^1 = 180°$; (B) C^{α}, $\chi^1 = 60°$; (C) C^{α}, $\chi^1 = -60°$; (D) C^{β}, $\chi^1 = 180°$; (E) C^{β}, $\chi^1 = 60°$; (F) C^{β}, $\chi^1 = -60°$.

alternative way, by using Kohn–Sham density functional theory (DFT) (Kohn and Sham, 1965), where speed scales as ~ N^3 (or, in practice, $N^{2.2}$). DFT programs for structure calculations are commercially available, and DFT-based chemical shielding calculations will complement their SCF partners over the next few years (Malkin et al., 1994). The advantage will be a major speeding up of shielding calculations – including the elusive correlation aspect, and the ability to compute the shielding of much heavier elements, such as ^{57}Fe and ^{113}Cd. Obtaining accurate absolute shieldings becomes particularly important when studying 'single-point' chemical shifts, i.e., when a whole range of known conformational / shift data is absent. Examples would be the shielding of a bound drug molecule whose conformation is sought, a metal-bound CO or O_2 ligand, or a metal ion shift. In such circumstances, the desirability of additional solid-state data, in the form of shielding tensors, is quite evident. Of course, even though DFT methods are fast, analysis of very large (e.g. 1000 basis functions) systems can be expected to be slowed down because of convergence problems. Nevertheless, analysis of very large species, such as hemes, is now tractable on fast (100 Mflop) machines.

^{1}H Ab initio

To date, essentially all ^{1}H shift analyses have been empirical. This is due in part to the fact that the simple empirical fitting procedures work well, at least for H^α. Also, accurate ^{1}H shieldings are difficult to compute ab initio, due to the need for many functions on H as well as the other atoms in its vicinity, and the importance of additional long-range ring-current, heme-current, peptide susceptibility, and weak electrical interactions. Quantum chemistry can be expected to make two important contributions in this area. First, it will put the magnetic susceptibility contributions on a more rigorous basis by permitting evaluation of the appropriate susceptibilities, χ, at a high level. Second, the very difficult H^N shift problem will become tractable. With H^N, the problems encountered using the empirical methods must be largely due to strong hydrogen bonding in the peptide group. This can be handled ab initio, but the problem as always is that an accurate geometry is required. The H^N shift will be highly sensitive not only to the N-H and N-O bond lengths, but also to the relative orientation of the N-H and C-O bond vectors. For good measure, these are all likely to vary with solvent exposure, and will change on ab initio geometry optimization (as we have found for C^o (de Dios and Oldfield, 1994b)). All is not hopeless, however, since given a reasonable starting structure, it should be possible to refine it using ab initio geometry optimization. That is, the starting structure will be energy minimized not by using an empirical force field, but by means of high-level quantum chemical calculation.

Structure prediction

In the early work of Spera and Bax (1991) and Wishart et al. (1992) it was clearly shown for C^α, C^β, C^o and H^α that well-defined distinctions between helical and sheet chemical shifts exist, culminating in the Chemical Shift Index and Consensus Index approaches for secondary structure prediction (Wishart and Sykes, 1994). These are essentially digital or tristate approaches to structure, and for more accurate structure prediction, more continuous functions are required in order to accurately predict e.g. ϕ,ψ torsion angles in sheet regions, as well as χ^i values. One approach to solving this problem is to evaluate quantum mechanically full shielding surfaces or hypersurfaces: $\sigma(\phi,\psi)$; $\sigma(\phi,\psi,\chi^1)$; $\sigma(\phi,\psi,\chi^1,\chi^2, ...)$ etc. and use these to deduce probable structure.

In order to demonstrate one approach, consider the alanine C^α and C^β shielding surfaces described previously (de Dios et al., 1993a). These shielding surfaces resemble the global experimental secondary chemical shift surfaces of Spera and Bax, but the quantum chemical approach permits evaluation of individual surfaces for each amino acid, and all regions of ϕ,ψ space are accessible, not just those represented in the relatively limited experimental database.

If a particular residue has a known set of ϕ,ψ values, then the shielding surfaces presented previously can be used to predict the experimental chemical shifts with quite good accuracy (Le et al., 1995). For 57 glycine, alanine and valine C^α sites in the nuclease from *Staphylococcus aureus* and a vertebrate calmodulin, the rmsd between solution-state experiment and (X-ray) theoretical prediction is ~ 1.6 ppm (Le et al., 1995) over the full 24 ppm C^α chemical shift range. This then leads immediately to the idea that if C^α and C^β chemical shifts are known, then at least some estimates can be made (by reference to such shielding surfaces) about the regions of ϕ,ψ space which permit prediction of, or are consistent with, the experimental chemical shifts. One quantitative measure of the likelihood that a given ϕ,ψ set represents the actual ϕ,ψ values can be made simply by evaluating the Gaussian:

$$Z = e^{-\left(\frac{\delta_{expt} - \delta(\phi,\psi)}{W}\right)^2} \qquad (4)$$

where Z is the probability that a particular ϕ,ψ set gives a chemical shift, $\delta(\phi,\psi)$, identical to the experimental value, δ_{expt}. W is a search-width parameter (to take into account errors in the surface and errors in δ_{expt}), and can be defined as the standard deviation between predicted and known experimental shifts from a series of shifts in proteins of known structure. Figures 2A and B show probability surfaces from C^α and C^β shifts of a helical residue (Ala^{60} in SNase). The Z-surfaces are clearly quite different, but both encompass the correct helical solution

region. Since both probability surfaces are independent of one another, they can be multiplied to yield a Bayesian 'consensus' or ^2Z surface, and H^α data can be incorporated in basically the same way. Figures 2C and D show such ^2Z and ^3Z surfaces for a typical helical residue in SNase (Le et al., 1995).

For structure prediction purposes, we can take the maximum value of the ^3Z surface to define a ϕ,ψ solution, and for 22 of the 24 alanine ϕ,ψ values in SNase, we find an rmsd between the X-ray values and those predicted from the isotropic chemical shifts of ~10°. Two residues are ~30° from the X-ray solution, but without crystal chemical shifts it is not possible to deduce the origins of these small differences. While the predicted ϕ,ψ values are very close to those expected, a 10–15° error propagated over all residues would not give a reasonable structure. The frequent inter-residue nature of NOE distance constraints would therefore appear to make them much more suitable for structure determination, but at present it is not known just what improvements in structure will be obtained when all chemical shift ϕ,ψ,χ predictions in a protein are used. Here, it is worth noting that solid-state NMR must play a pivotal role in testing and refining all theories of protein chemical shifts, and solid-state isotropic (and anisotropic) chemical shifts in proteins are urgently required. C^α, C^β, C^o, ^{15}N etc. chemical shifts in crystalline proteins can readily be measured (Cole et al., 1988), and the chemical shifts determined can be used to make structure predictions which can be *directly* compared with X-ray results. This approach seems the one most likely to lead to resolution of the puzzling observation that NMR structures from different groups may more closely resemble X-ray structures than each other (Smith et al., 1994).

Structure refinement

A more realistic short-term goal of chemical shift theory is to be able to *refine* solution and crystal structures by providing additional restraints on ϕ, ψ and χ from chemical shifts (Laws et al., 1993). Progress in this direction has been reported using ring currents in heme proteins, and other weak magnetic susceptibility anisotropy and electrical interactions, with H^α as the probe (Ösapay et al., 1994). More recently, Celda et al. (1995) and Kuszewski et al. (1995) have used C^α and C^β shifts and the empirical shift surfaces of Spera and Bax (1991) in refinement.

The most computationally simple method for chemical shift refinement using ^{13}C shifts is simply to take the most

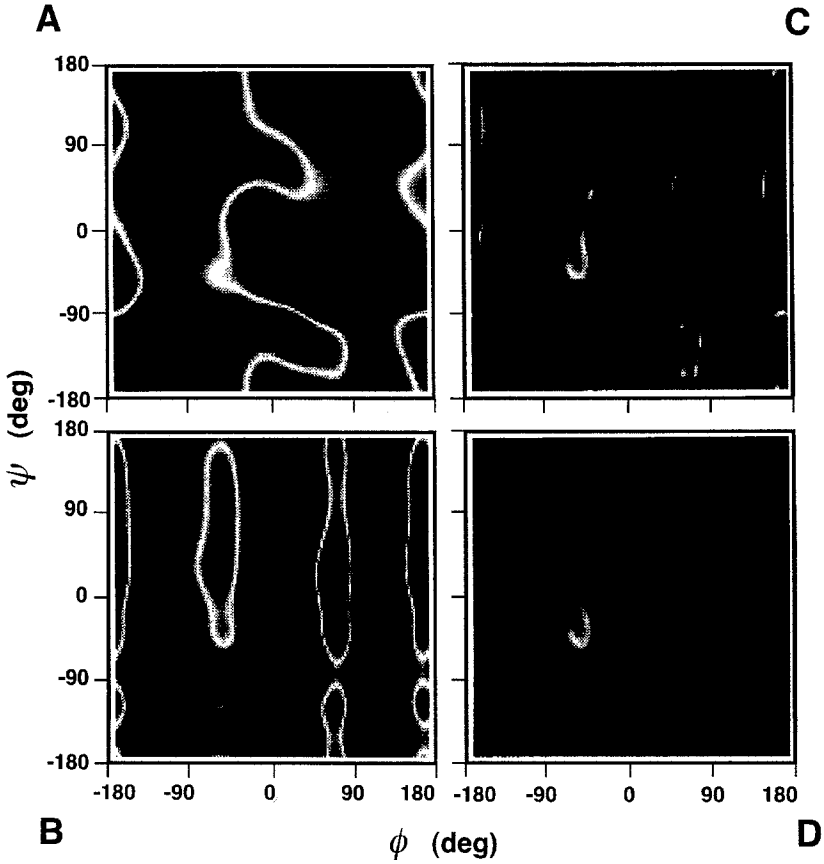


Fig. 2. Chemical shift surface prediction of peptide backbone ϕ,ψ torsion angles for Ala^{60} in SNase. (A) $^1Z\delta(C^\alpha)$; (B) $^1Z\delta(C^\beta)$; (C) $^2Z\delta(C^\alpha)\delta(C^\beta)$; (D) $^3Z\delta(C^\alpha)\delta(C^\beta)\delta(H^\alpha)$. Based on work described by Le et al. (1995).

likely ϕ,ψ solution from a Z-surface prediction and incorporate this as a torsional restraint. However, as can be seen from Fig. 2, this approach will only be successful when there is a very small region of allowed conformational space. This is often not the case, since other mathematically correct solutions are also frequently observed. A more appropriate approach is to incorporate chemical shifts, chemical shift surfaces, and their derivatives, in a direct chemical shift structure refinement. Working with SNase data from Markley and Wang, we have incorporated alanine and valine shifts and shift surfaces *directly* into X-PLOR (Brünger, 1992). The rmsd results for ϕ, ψ and χ for a set of 10 structures are within about 10–15° of the X-ray results, to be compared with a ~15–20° rmsd for a similar set of structures which do not contain chemical shift restraints. Both sets have about the same number of minor NOE violations (~0.2 per structure).

In the future, it will be possible to incorporate additional restraints by consideration of the shielding of other atoms, e.g. C^{γ}, C^{o} and ^{15}N. We have recently computed a full three-dimensional chemical shift hypersurface for C^{α},C^{β} of valine, $\sigma(\phi,\psi,\chi)$, and at least for sp^3 carbons it will be possible to correlate structure and shielding using hypersurfaces. For C^{o} and ^{15}N, the situation is much more complex. C^{o} shifts exhibit an apparently quite well defined correlation with ϕ,ψ structure, and C^{o} is very useful in the consensus CSI approach (Wishart and Sykes, 1994). However, early ab initio results showed the opposite shift behavior to that seen experimentally (Jiao et al., 1993). The origin of these anomalous theoretical results appears to be that there are differential bond length changes on hydrogen bonding in helices and sheets, a result deduced from ab initio geometry optimization (de Dios and Oldfield, 1994b). The apparent correlation of C^{o} shifts with ϕ,ψ or secondary structure must therefore be an indirect effect due to geometric changes (primarily a C^{o}-O bond length change). The real ϕ,ψ dependence is in fact very small, only about 1 ppm, and in the wrong direction with respect to the ~5 ppm shift difference seen experimentally. In the future, this small effect may still be of use, although construction of geometry-optimized surfaces will be very time-consuming, so C^{o} shifts will initially be used in an empirical fashion. Indeed, it now seems logical to incorporate H^{α}, C^{α} and C^{β} into all refinements, initially using global empirical surfaces, and then replacing each of these as individual ab initio surfaces become available.

For ^{15}N, good progress has already been made in predicting experimental chemical shifts from known structure, but the type of inversion from shift to structure used for C^{α},C^{β}, or direct use of ^{15}N shifts, is much more complex. This is because there are not only ϕ,ψ contributions to shielding, but also ψ_{i-1}, ϕ_{i+1}, χ, hydrogen bonding, and potentially even longer range electrostatic field contributions, as well as local sequence effects. At present ^{15}N chemical shifts are best used in validation, rather than prediction or refinement.

Electrostatics and dynamics

As noted previously (de Dios et al., 1993b), an operational categorization of the contributions to total shielding, σ_t, is helpful in deciding how to go about carrying out a protein chemical shift calculation:

$$\langle\sigma_t\rangle = \langle\sigma_s\rangle + \langle\sigma_e\rangle + \langle\sigma_m\rangle \tag{5}$$

$\langle\sigma_s\rangle$ is the short-range contribution to total shielding (or shift) and, for the heavy elements ^{13}C, ^{15}N and ^{17}O, needs to be evaluated ab initio. This term dominates the folding contributions to shielding for C^{α}, C^{β}, C^{γ},..., ^{15}N, and probably ^{17}O. The second term, $\langle\sigma_e\rangle$, represents the electrostatic contributions to shielding, and the third the magnetic contributions. There are at least two approaches to evaluating $\langle\sigma_e\rangle$. In one method, a charge field in a protein is created by substituting atoms with basis functions by their partial atomic charges – the charge field perturbation gauge-including atomic orbital (CFP-GIAO) approach (de Dios et al., 1993b; de Dios and Oldfield, 1993). In a second method, which closely follows that of Stephen and Buckingham (Stephen, 1957; Buckingham, 1960), derivatives of the shielding are evaluated with respect to the components of a series expansion of the potential:

$$\begin{aligned} A_{\alpha\beta,\gamma} &= \frac{\partial\sigma_{\alpha\beta}}{\partial V_{\gamma}}\cdots \\ B_{\alpha\beta,\gamma,\gamma} &= \frac{\partial^2\sigma_{\alpha\beta}}{\partial V_{\gamma}^2}\cdots \\ C_{\alpha\beta,\gamma\gamma} &= \frac{\partial\sigma_{\alpha\beta}}{\partial V_{\gamma\gamma}}\cdots \end{aligned} \tag{6}$$

where A, B and C are so-called shielding polarizability tensors, V_{γ} is the γth element of the uniform field, $V_{\gamma\gamma}$ is the $\gamma\gamma$ element of the field gradient, and so on (Augspurger et al., 1991). Several of these tensors have been calculated by Buckingham, Raynes and more recently by Dykstra et al. (Augspurger and Dykstra, 1991; Augspurger et al., 1992). When multiplied by the appropriate component of the charge field (field, hypergradient, etc.), the corresponding electrical contributions to shielding are obtained. There is a close correspondence between the shielding polarizability and CFP-GIAO methods in model systems (Augspurger et al., 1993), and with supermolecule calculations in which basis functions are used everywhere (de Dios and Oldfield, 1993). Moreover, GIAO methods themselves give a very good account of the ^{19}F isotropic

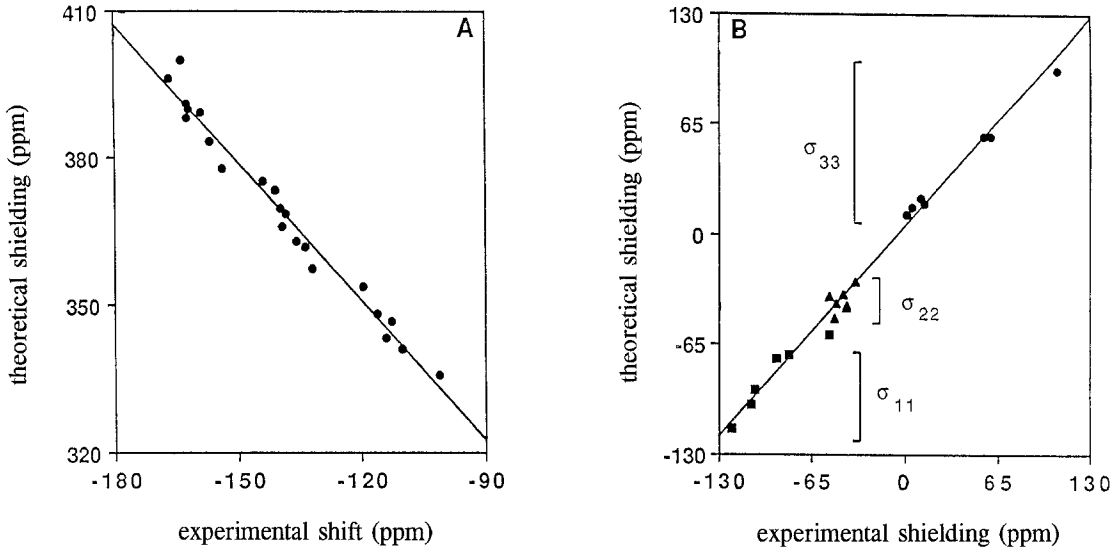


Fig. 3. Experimental versus theoretical ^{19}F chemical shift/shielding results for fluorobenzenes. (A) Liquid-state isotropic chemical shifts (in ppm from external $CFCl_3$) plotted versus computed shielding values; slope = −0.94, $R^2 = 0.975$. Molecules studied: C_6H_5F; 1,2-$C_6H_4F_2$; 1,3-$C_6H_4F_2$; 1,4-$C_6H_4F_2$; 1,2,3-$C_6H_3F_3$; 1,2,4-$C_6H_3F_3$; 1,3,5-$C_6H_3F_3$; 1,2,3,4-$C_6H_2F_4$; 1,2,3,5-$C_6H_2F_4$; 1,2,4,5-$C_6H_2F_4$; C_6HF_5; C_6F_6; and C_6F_5Cl. (B) Solid-state shielding tensor components (in ppm from C_6F_6) plotted versus computed shielding tensor elements; slope = 0.954, $R^2 = 0.989$. Molecules studied: C_6H_5F; 1,2-$C_6H_4F_2$; 1,3-$C_6H_4F_2$; 1,4-$C_6H_4F_2$; 1,3,5-$C_6H_3F_3$; 1,2,4,5-$C_6H_2F_4$; and C_6F_6. The 'ortho-effect', the shielding of σ_{22} by ~50 ppm on each *o*-F substitution, is clearly seen in (B), as is the invariance of σ_{33} with substitution. Chemical shielding calculations were carried out on geometry-optimized structures.

shifts and shielding tensors for many fluorobenzenes, as shown in Fig. 3, without recourse to use of any adjustable parameters. In particular, we find no support for the notion that van der Waals dispersion dominates ^{19}F shielding in these or any other systems.

As noted in the references given above, in simple model systems such as fluorobenzene perturbed by HF molecules, GIAO supermolecule calculations (all atoms have basis functions), CFP-GIAO and shielding polarizability approaches give the same results for the charge field perturbation of fluorobenzene by HF molecules. Precisely the same methods can be applied to proteins, either using a simple charge field approach in which the local charge field is represented by e.g. AMBER point charges, or evaluating the field and field gradient using a suitable program such as Polaris/Enzymix (Pearson et al., 1993 and references cited therein); these elements of the field are then multiplied by the appropriate coefficients A, B etc. to obtain the electrical contributions to shielding. Good accord with experiment has been obtained with the five [5-F]Trp residues in the galactose binding protein from *Escherichia coli* (Pearson et al., 1993). Of course, computing even the field in a protein is a challenging proposition, but the fact remains that ^{19}F shifts can be predicted, so long as an accurate initial structure is known.

Electrostatic fields also contribute to ^{13}C and ^{15}N shielding, but the electrical contributions in most cases appear to be small. For ^{13}C, most atoms are sp^3, and the shielding polarizabilities for sp^3 carbons are small (Augspurger et al., 1992). For nitrogen, much of the charge field effect on shielding is usually contained in the hydrogen-bond interaction evaluated as σ_s, with the weaker longer range interactions being minor when compared with the geometric (ϕ,ψ,χ) contributions. One case in which electrostatic fields are important in ^{13}C shielding is seen in the case of charged sp^2 groups, specifically the C° carbons of zwitterionic threonine and tyrosine (de Dios et al., 1994). Since these carbons are sp^2 hybridized, they are more highly polarizable and, being charged, counterion charge field effects make very large contributions to shielding. The CFP-GIAO method nevertheless gives a good account of the individual shielding tensor elements, and similar effects are likely to occur in proteins also, e.g. in $CO_2^- \cdots NH_3^+$ salt bridges.

Dynamical effects will also contribute to ^{13}C and ^{15}N chemical shifts as they do with ^{19}F, although to date we have found no improvements in predicted ^{13}C chemical shifts when using (admittedly modest) trajectory lengths (5×50 ps). For the most accurate chemical shift predictions possible, it is clear that trajectories will need to be evaluated over extremely long time scales. Some force fields can and do degrade chemical shift predictions faster than do other (good) force fields, but improvements over equilibrium (minimized) X-ray structure predictions have yet to be realized for ^{13}C shifts.

Electrostatics/enzyme mechanisms

Electrostatic fields, as well as ligand–receptor interactions, are thought by many workers to play a key role in enzyme catalysis (Warshel, 1991). There are few, if any,

reliable spectroscopic probes of electrostatic fields in proteins, although ^{19}F and some ^{13}C chemical shifts, as we have noted, do appear to be influenced by electric fields. In the future, it can therefore be expected that NMR chemical shifts will begin to play a role in probing protein electrostatics. Already, simple charge field models have enabled moderately good prediction of ^{13}C shielding tensors in crystalline amino acids, and of F-Trp shifts in the *E. coli* galactose binding protein (Luck and Falke, 1991a; Pearson et al., 1993). Since there are many descriptions of protein electrostatics, it should be possible to test which ones allow the most accurate predictions of protein chemical shifts, and by inference the best description of protein electrostatics. To date this is one of the least developed areas of chemical shift research. There are three main problems at present. First, finite difference (FD) Poisson–Boltzmann methods use a grid for the FD calculation, and in our hands at least this has resulted in relatively imprecise fields (at ^{19}F). Field gradients are also necessary and have to be evaluated by further FD. Second, the exact structures of ^{19}F-labeled proteins are not known, and there is always the possibility of small local changes on F substitution – although these must be very local (Luck and Falke, 1991a,b). Third, there are difficulties of a gauge-like nature in evaluating the quadrupole shielding polarizability (Augspurger et al., 1993). Resolution of these problems will permit even more detailed testing of the different models of protein electrostatics.

Folding/unfolding

On addition of denaturants, some proteins undergo significant chemical shift changes prior to 'full' denaturation, and analyzing chemical shifts may help define the nature of the local structural changes occurring prior to unfolding. In the case of hen egg white lysozyme, folding appears to occur in more than one step (Miranker et al., 1991); relatively well defined folding domains have been identified in the α-region of the protein. Conversely, guanidinium chloride denaturation of [4-F]Trp lysozyme reveals differential 'unfolding', as monitored by chemical shift changes, due possibly to differential hydration and charge field collapse. A combined ^{1}H, ^{13}C, ^{15}N and ^{19}F NMR study of HEWL chemical shifts, as well as those of other proteins, can be expected to contribute to our understanding of unfolding. In the longer term, chemical shift information can also be expected to contribute to a better understanding of the structures of folding intermediates such as molten globule states, although dynamical averaging using shielding surfaces will be required. At the bottom of the 'folding funnel', the nature of the major conformational substates present, in for example CO–myoglobin (Frauenfelder et al., 1990), should also be accessible. Solid-state studies (of ^{13}C, ^{17}O and potentially ^{57}Fe) at very low temperatures should be particularly useful here, since different substates can be 'trapped'. Future studies of freeze-trapped intermediates in enzyme reactions, especially those involving unusual geometries, will also be accessible by using ab initio methods.

Future directions and needs

Much of the progress in determining the relationships between the three-dimensional structures of proteins and their experimental chemical shifts has been made in the past three years, especially for the heavier elements. The general notion that ^{13}C, ^{15}N and ^{19}F shifts in proteins are too difficult to calculate seems to have been overestimated, although the process is still a very slow one. Faster RISC machines (or a lot of slower ones) are required in order to make more rapid progress. Density functional methods which scale formally as N^3 (where N is the number of basis functions) rather than N^4 or N^5 (for Hartree–Fock and post-Hartree–Fock methods), and also include electron correlation, appear attractive as well and offer hope for the possibility to evaluate heavy-atom shifts, such as those of Fe. On the horizon, so-called order-N methods, of current interest in the solid-state physics community, may also play a role.

What can reasonably be expected in the next couple of years is the availability of additional shielding surfaces, and their incorporation into structure refinement, both in solution and in the crystalline and not-so-crystalline solid state. Empirical surfaces will initially play a role, then they will be replaced as individual ab initio shielding surfaces and hypersurfaces become available. In favorable cases, it will be possible to use tensor information to predict or refine structure, using data from both solid- and liquid-state experiments. Using DFT methods in particular, the shifts of metal ions, such as those of ^{57}Fe or ^{113}Cd, should also be accessible, as will the shifts of ligands bound to these metals. Many of the quantum chemical methods described above are also clearly applicable to studies of other macromolecules, such as lipids, nucleic acids and carbohydrates, either alone or when bound to proteins.

Finally, further development and testing of protein chemical shift theories requires new experiments to be performed, in particular those aimed at providing solid-state chemical shifts of proteins having well-defined crystallographic structures, so that more precise tests of computed shieldings and experimental shifts (and shift tensors) can be made, both for the lighter elements (^{13}C, ^{15}N, ^{17}O) and for more challenging cases, such as metal (e.g. ^{57}Fe) and metal–ligand (e.g. Fe–CO, Cu–histidine, Cd–S) interactions. Once this is achieved, then the geometric, electronic and dynamic structures of unknown systems can be confidently addressed using chemical shifts.

Acknowledgements

I am most grateful to J. Augspurger, A.C. de Dios, C. Dykstra, D. Grant, J.F. Hinton, C. Jameson, W. Kutzelnigg, D.D. Laws, H. Le, F. Lee, J.G. Pearson, P. Pulay, D. Salahub, A. Warshel and K. Wolinski for their invaluable contributions to this project. This work was supported by the United States Public Health Service (Grants HL-19481, GM-40426 and GM-50694), by the American Heart Association with support from the AHA Illinois affiliate, Inc. (Grant 92-013340), and in part by an IBM Shared University Research Equipment grant and by the Graduate Research Board of the University of Illinois.

References

Augspurger, J.D. and Dykstra, C.E. (1991) *J. Phys. Chem.*, **95**, 9230–9238.

Augspurger, J.D., Dykstra, C.E. and Oldfield, E. (1991) *J. Am. Chem. Soc.*, **113**, 2447–2451.

Augspurger, J., Pearson, J., Oldfield, E., Dykstra, C.E., Park, K.D. and Schwartz, D. (1992) *J. Magn. Reson.*, **100**, 342–357.

Augspurger, J.D., de Dios, A.C., Oldfield, E. and Dykstra, C.E. (1993) *Chem. Phys. Lett.*, **213**, 211–216.

Brünger, A.T. (1992) *X-PLOR*, Yale University Press, New Haven, CT.

Buckingham, A.D. (1960) *Can. J. Chem.*, **38**, 300–307.

Celda, B., Biamonti, C., Arnau, M.J., Tejero, R. and Montelione, G.T. (1995) *J. Biomol. NMR*, **5**, 161–172.

Cole, H.B.R., Sparks, S.W. and Torchia, D.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6362–6365.

de Dios, A.C. and Oldfield, E. (1993) *Chem. Phys. Lett.*, **205**, 108–116.

de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993a) *J. Am. Chem. Soc.*, **115**, 9768–9773.

de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993b) *Science*, **260**, 1491–1496.

de Dios, A.C., Laws, D.D. and Oldfield, E. (1994) *J. Am. Chem. Soc.*, **116**, 7784–7786.

de Dios, A.C. and Oldfield, E. (1994a) *J. Am. Chem. Soc.*, **116**, 5307–5314.

de Dios, A.C. and Oldfield, E. (1994b) *J. Am. Chem. Soc.*, **116**, 11485–11488.

Ditchfield, R. (1972) *J. Chem. Phys.*, **56**, 5688–5691.

Fesik, S.W. (1993) *J. Biomol. NMR*, **3**, 261–269.

Frauenfelder, H., Alberding, N.A., Ansari, A., Braunstein, D., Cowen, B.R., Hong, M.K., Iben, I.E.T., Johnson, J.B., Luck, S., Marden, M.C., Mourant, J.R., Ormos, P., Reinisch, L., School, R., Schulte, A., Shyamsunder, E., Sorensen, L.B., Steinback, P.J., Xie, A., Young, R.D. and Yue, K.T. (1990) *J. Phys. Chem.*, **94**, 1024–1037.

Gauss, J. (1992) *Chem. Phys. Lett.*, **191**, 614–620.

Jiao, D., Barfield, M. and Hruby, J.M. (1993) *J. Am. Chem. Soc.*, **115**, 10883–10887.

Kohn, W. and Sham, L.J. (1965) *Phys. Rev.*, **140A**, 1133–1138.

Kuszewski, J., Qin, J., Gronenborn, A.M. and Clore, G.M. (1995) *J. Magn. Reson. Ser. B*, **106**, 92–96.

Kutzelnigg, W., Van Wullen, C., Fleischer, U., Franke, R. and Van Mourik, T. (1993) In *Nuclear Magnetic Shieldings and Molecular Structure* (Ed., Tossell, J.A.) Kluwer, Dordrecht, pp. 141–162.

Laws, D.D., de Dios, A.C. and Oldfield, E. (1993) *J. Biomol. NMR*, **3**, 607–612.

Le, H., Pearson, J.G., de Dios, A.C. and Oldfield, E. (1995) *J. Am. Chem. Soc.*, in press.

London, F. (1937) *J. Phys. Radium*, **8**, 397–409.

Luck, L.A. and Falke, J.J. (1991a) *Biochemistry*, **30**, 4248–4256.

Luck, L.A. and Falke, J.J. (1991b) *Biochemistry*, **30**, 4257–4261.

Malkin, V.G., Malkina, O.L., Casida, M.E. and Salahub, D.R. (1994) *J. Am. Chem. Soc.*, **116**, 5898–5908.

Miranker, A., Radford, S.E., Karplus, M. and Dobson, C.M. (1991) *Nature*, **349**, 633–636.

Ösapay, K., Theriault, Y., Wright, P.E. and Case, D.A. (1994) *J. Mol. Biol.*, **244**, 183–197.

Pearson, J.G., Oldfield, E., Lee, F.S. and Warshel, A. (1993) *J. Am. Chem. Soc.*, **115**, 6851–6862.

Ponder, J.W. and Richards, F.M. (1987) *J. Mol. Biol.*, **193**, 775–791.

Ramsey, N.F. (1950) *Phys. Rev.*, **78**, 699–703.

Ramsey, N.F. (1952) *Phys. Rev.*, **86**, 243–246.

Smith, L.J., Redfield, C., Smith, R.A.G., Dobson, C.M., Clore, G.M., Gronenborn, A.M., Walter, M.R., Naganbushan, T.L. and Wlodawer, A. (1994) *Nature Struct. Biol.*, **1**, 301–310.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Stephen, M.J. (1957) *Mol. Phys.*, **1**, 223–232.

Wagner, G. (1993) *J. Biomol. NMR*, **3**, 375–385.

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

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.

Wolinski, K., Hinton, J.F. and Pulay, P. (1990) *J. Am. Chem. Soc.*, **112**, 8251–8260.