PMSE POLYMERIC MATERIALS SCIENCE AND ENGINEERING

VOLUME 71

FALL MEETING 1994
WASHINGTON, D.C.

PROCEEDINGS OF THE AMERICAN CHEMICAL SOCIETY
DIVISION OF POLYMERIC MATERIALS: SCIENCE AND ENGINEERING

NMR STUDIES OF BIOPOLYMERS: ANALYZING CHEMICAL SHIFTS, Eric Oldfield, Angel C. de Dios, John G. Pearson and Hongbiao Le, Department of Chemistry, University of Illinois at Urbana-Champaign, 505 South Mathews Avenue, Urbana, IL 61801, USA

Introduction. It has been known for more than 20 years (1,2) that the folding of a protein into its native conformation causes large ranges of nuclear magnetic resonance (NMR) chemical shift non-equivalencies to be introduced - about 5-10 ppm for ^{13}C (2,3) 30 ppm for ^{15}N (4), and 15 ppm for ^{17}O (5) and ^{19}F (6). However, surprisingly little progress in computing such chemical shifts from known solid or liquid state structures has been made, for without these non-equivalencies, modern multidimensional NMR studies of protein structure would not be possible. An understanding of the origins of chemical shifts is expected to lead to new ways of predicting, refining and validating protein structure.

In principle, chemical shifts can be computed by using *ab initio* techniques, but full *ab initio* computations on 1000 atom-plus structures are not currently feasible. However, it seemed unreasonable to us to suppose that the effects of *all* atoms would need to be incorporated into a chemical shielding calculation, because nuclear shielding is fundamentally a local phenomenon. We show below and elsewhere (7-13) that good values for ^{1}H, ^{13}C, ^{15}N, and ^{19}F folding-induced shielding can now be obtained using quantum chemical methods.

The idea of using chemical shift information from proteins to deduce structure can, in a sense be traced back to very early observations - in the 1960s and 70s, of characteristically different chemical shifts for α-helical, β-sheet and "random coil" shifts in proteins in peptides (14). Much later, a number of groups began to parameterize shifts, especially H^{α}, using peptide magnetic anisotropy, ring current, and electrostatic field effects (15,16). These semi-empirical ideas have now begun to find some utility in refining protein structures obtained by using standard NOE methods (17). In addition, Wishart, Sykes and Richards have presented a structure prediction protocol, the so-called chemical shift index (CSI). The idea here is simple and is that for C^{α}, H^{α}, C', and to a lesser extent C^{β}, there are clear positive or negative secondary chemical shifts from the random coil chemical shift position which are structure (helix, sheet) dependent. The CSI method uses a tri-state model in which e.g. for C^{α}, a "1" is assigned to a helical shift, a "0" to a random coil shift, or a "-1" to a β-sheet shift. When a certain number of consecutive residues exhibit e.g. helical behavior, then the presence of helical secondary structure is inferred. The method is useful when C^{α}, H^{α} and C' indices are combined into a consensus CSI pattern, but is limited to helix, sheet separations.

The success of the empirical chemical shift index method immediately leads one to ask - why it is successful, and how can it be improved? The method is successful because chemical shifts are conformation dependent. Since many such chemical shifts can now be predicted using *ab initio* methods, it is thus logical to see if additional, more detailed structural information can be obtained, since as shown by Spera and Bax (18), there are considerable chemical shift differences *within* a particular structure type region, even in a global amino-acid data base. That is, helical and sheet residues do not have just two chemical shifts, rather, a more or less continuous distribution. How then can more detailed structural information be derived from these distribution functions?

The answer lies in being able to predict *exactly* how e.g. C^{α}, C^{β} vary with φ,ψ, since this knowledge of precise C^{α}, C^{β} chemical shifts for a given amino-acid, would, together with the availability of each amino-acid shielding surface, permit a prediction of an allowed region of φ,ψ space consistent with the experimental shifts. For alanine, we have already shown that this approach permits prediction of φ,ψ values for *Staphylococcal* nuclease (8) to a ~10° rmsd with the x-ray structure, and the bulk of our current research is concerned with evaluating similar shielding surfaces for each of the other amino-acids. Our approach is thus a high-resolution "analog" version of the low-resolution, digital tri-state chemical shift index approach, and instead of simply predicting "helix" or "sheet" solutions, we predict accurate φ,ψ values.

Carbon-13 NMR. We first consider ^{13}C NMR shielding in proteins because the ability to compute backbone $^{13}C^{\alpha,\beta}$ shieldings should be a useful complement to other distance - based or more theoretical approaches to protein folding. Clear-cut correlations of ^{13}C chemical shifts with structure in polymers and proteins have only been reported fairly recently, but as Spera and Bax (18) state: "^{13}C chemical shifts remain poorly understood and significant deviations from expected secondary shifts can occur". We have thus investigated C^{α}, C^{β} shieldings of some model fragments containing Ala residues as well as the Ala residues in *Staphylococcal* nuclease to see if our understanding of ^{13}C shielding in proteins can be improved.

By way of introduction, we show in Figure 1A histograms of C^{α} and C^{β} shieldings for α-helical and β-sheet residues from the global database of *all* amino acids of Spera and Bax (18).

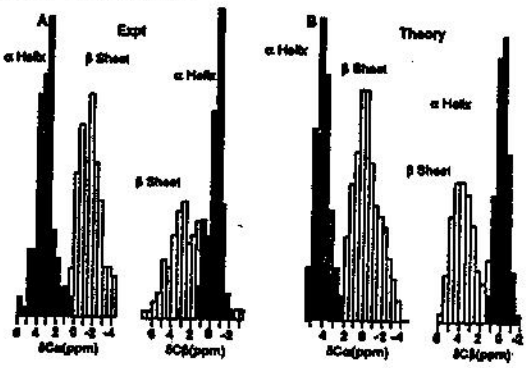
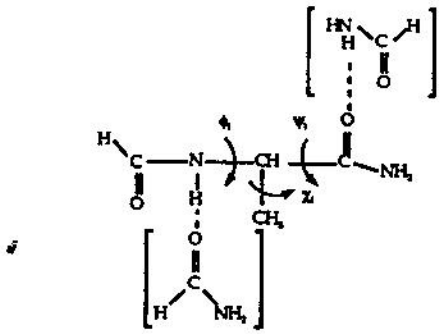


Figure 1: ^{13}C shielding in proteins. . **(A)** Histograms showing separation of α-helical and β-sheet chemical shifts for C^{α} and C^{β} sites in proteins [based on global data base of all amino-acids, Spera and Bax (18)]. **(B)** Histogram showing C^{α} and C^{β} shifts for model α-helical and β-sheet (alanine only) fragments, based on *ab initio* shielding surface calculations.

There is a clear separation between the shieldings of the two structure types, suggesting an important effect of φ, ψ torsion angles on shielding, but of course the experimental results could in principle be dominated by other interactions, such as hydrogen bonding and long-range electrostatic field effects. To investigate the effects of φ and ψ alone, we thus computed the ^{13}C NMR shieldings for C^{α} and C^{β} in a series of alanine molecular fragments:



We first used a 6-311G** basis set for all atoms (without the formamide hydrogen-bond partners). As shown in Fig. 1, there is a good general similarity between the experimental and theoretical shielding results, with the overall widths and ~5 ppm separations between helical and sheet residues being reproduced in the calculations. Of course, *there cannot be exact agreement*, since Spera and Bax used a data base of essentially all amino-acids,

and we have found that, C^α, C^β shielding surfaces are *very different* for each amino-acid - even when expressed as secondary shifts (8). ϕ and ψ torsion angles nevertheless both appear intimately correlated with C^α and C^β shieldings.

We then computed the theoretical shieldings for the 12 Ala C^α groups in SNase, (Fig. 2A) with the fragment shown above, and an attenuated basis set following the locally dense scheme proposed by Chesnut (19). We used a 6-311++G (2d,2p) basis for C^α, C^β, C^o, N, H^N and H^α (set I) and a 6-31G basis for other atoms in the fragment (set II), together with the known x-ray structure (7), and find a slope of -0.85 and a correlation coefficient $R^2 = 0.94$ (a measure of the 'goodness of fit') - a respectable agreement between theory and experiment (7). We then investigated the effects of hydrogen bonding and the proteins' charge field on C^α shielding with the CFP-GIAO (electroneutral) fragment approach used below for ^{19}F, but with the additional incorporation of formamide molecules (to represent hydrogen-bond partners) as pictured above. Results for a smaller basis set (6-311G**/I/6-31G/II) gave m = -1.2, R = 0.97, whereas a larger basis set (6-311++G(2d,2p)/I/6-31G/II) - including explicit hydrogen bonding and charge field perturbation, gave m= -1.18, R = 0.97, indicating basis set insensitivity. Generally similar results were obtained for C^β, Figure 2B (m = -0.78, R = 0.89).

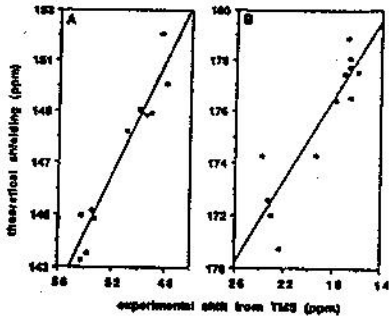


Figure 2: ^{13}C NMR shieldings for Ala residues in *Staphylococcal* nuclease (1). (A) Experimental C^α chemical shifts for the 12 Ala sites in SNase, *versus* computed shieldings. (B) Experimental C^β chemical shifts for the 12 Ala sites in SNase, *versus* computed shieldings.

Thus, ϕ,ψ effects are overwhelmingly dominant for ^{13}C, which enables us to use ^{13}C shifts as structure probes.

Here, we introduce the actual global data base results of Spera and Bax, since this allows us to introduce in a graphic way the concept of the Z-surface approach, which I will describe in detail in the accompanying lecture. Spera and Bax showed that the secondary shifts for C^α, C^β for a global data base of most amino-acids could be represented on a shielding or secondary chemical shift surface, $\delta(\phi,\psi)$:

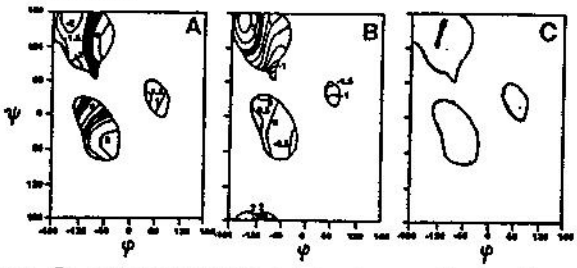


Figure 3: General approach to ϕ,ψ prediction by use of isotropic chemical shifts, using the shielding surfaces of Spera and Bax (Ref. 18), together with examples of allowed C^α, C^β shift regions. A: allowed δC^α; B: allowed δC^β; C: allowed $\delta C^\alpha + \delta C^\beta$ ϕ,ψ solution region.

where A and B represent the shift contours for C^α, C^β. The idea of the Z-surface approach is that if C^α, C^β are known from experiment, then the regions of ϕ,ψ space permitted are limited by C^α, C^β - as shown for example by the shaded regions in Figure 3A,B. Since both restraints must be obeyed, then only the region of ϕ,ψ space shown in C is an allowed ϕ,ψ. Using specific, quantum chemically computed shielding surfaces for alanine, we obtain ~10° rmsd values versus x-ray results - a powerful new approach to structure determination (8).

Nitrogen-15 NMR. Next, we will consider ^{15}N shielding in the peptide group. Nitrogen shielding might be expected to be more difficult to predict, since well defined ϕ,ψ correlations have not been observed, and hydrogen bonding/electrostatics could also be very important. Fortunately, however, Glushka et al. previously (20) noted an interesting correlation between ^{15}N secondary shifts and ψ_{i-1}, and we have reproduced these trends using empirical data bases, and more importantly, using GIAO. We discuss our detailed results for ^{15}N-alanine residues in the lecture, but here we show some results on valine (SNase) and the polar residue, threonine (in T4-phage lysozyme). ^{15}N chemical shifts can now be quite accurately predicted using GIAO methods, and Figure 4 shows results on ^{15}N-valine in SNase, as well as results on a polar amino-acid, threonine, in T4-phage lysozyme:

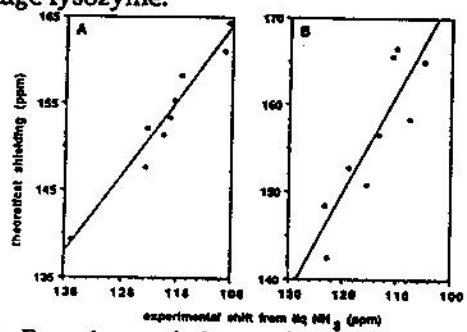


Figure 4: Experimental chemical shifts *versus* computed shieldings for ^{15}N NMR of proteins. A, ^{15}N-valine sites in SNase. B, ^{15}N-threonine sites in T4-phage lysozyme. Crystal structures used in both cases.

There is good accord in both cases, indicating that ^{15}N NMR is a potentially powerful new technique for *structure validation*. In addition, in more recent work, we have used standard fragments for alanine, and we find excellent correlations between experimental ^{15}N chemical shifts of all sheet residues and theoretical shielding, without use of any electrostatic/H-bonding contributions.

Thus, ^{15}N NMR will have utility in validation, but it is more difficult to use in the Z-surface method since ϕ_{i+1}, ψ_i also have on effect on $^{15}N_i$ shielding.

For ^{15}N, the key observations or preliminary results are: 1), in β-sheet regions, ϕ,ψ,χ torsions dominate shielding, and can thus be used in structure prediction, refinement and validation. 2) for α-helical residues, strong hydrogen bonding and helix-dipole charge fields, cause deshielding. Thus, ^{15}N chemical shifts are much more useful restraints for sheet residues - as shown for example in the excellent correlation (rmsd ~1 ppm) for the sheet residues in dihydrofolate reductase and cytochrome c_{551} (21). These results put on a firm footing the empirical ϕ_i, ψ_{i-1} shift correlations seen previously (20), and promise to be of considerable help in validation.

Finally, we consider the case of ^{19}F NMR in proteins. Fluorine chemical shifts have been studied in proteins for over 25 years, but only with our recent work have they been analyzed. They thus provide a final rigorous test case for our ability to predict essentially all observable heavy atom chemical shifts in proteins.

Fluorine NMR and Electrostatics. Fluorine is a particularly sensitive probe of electrostatic field effects in proteins and fluorine chemical shifts can now be reliably evaluated. The way this is best done is to first evaluate the derivatives of the shielding tensor elements with respect to the uniform field, the field gradient and the field hypergradient, obtaining what are called the multipole shielding polarizabilities (22):

Calculated multipole shielding polarizability tensor elements for the ^{19}F nucleus in fluorobenzene with respect to a uniform field, field gradient, and field hypergradient

Tensor elements and values (ppm/au) [a]			
$P_{xx,x}$	1483	$P_{yy,xx}$	−52907
$P_{yy,z}$	3577	$P_{yy,yy}$	−3165
$P_{xx,z}$	593.6	$P_{yy,zz}$	−6019
$P_{xz,xz}$	2224	$P_{xx,xx}$	−16356
$P_{xz,yy}$	−716.6	$P_{xx,yy}$	−6483
$P_{xz,zz}$	−488.4	$P_{xx,zz}$	−2244
$P_{yy,xx}$	6175	$P_{xx,xxxx}$	7259
$P_{yy,yy}$	−2717	$P_{xx,xxyy}$	−141.9
$P_{yy,zz}$	−652.9	$P_{xx,zzzz}$	422.5
$P_{xx,xz}$	1319	$P_{yy,xxxx}$	17087
$P_{xx,yy}$	−796.6	$P_{yy,xxyy}$	−1428
$P_{yy,zz}$	−283.8	$P_{yy,zzzz}$	1449
$P_{xx,x,x}$	−23025	$P_{xx,xxxz}$	3325
$P_{xx,yyy}$	1105	$P_{xx,xyyy}$	−694.5
$P_{xx,x,z}$	−2917	$P_{xx,xzzz}$	−21.1

then these coefficients are multiplied by the actual values of the field, the field gradient and the field hypergradient, obtained from a local reaction field-molecular dynamics program, containing a multipole description of the charge field (i.e. Enzymix, Ref. 23). The following two Figures show, first, how the chemical shifts for five [5-F]Trp sites in the galactose binding protein (GBP) from *Escherichia coli* vary with trajectory time (Figure 5):

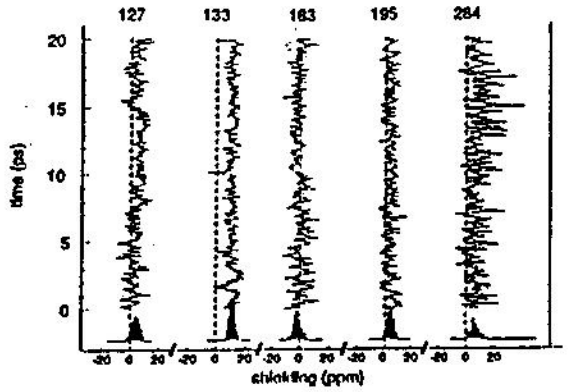


Figure 5: Individual 20-ps shielding trajectories for each of the five [5-F]Trp residues in GBP.

and second, Figure 6 shows the excellent agreement between theory and experiment for the five [5-F]Trp sites in GBP, using the shielding polarizability-molecular dynamics approach (7,11):

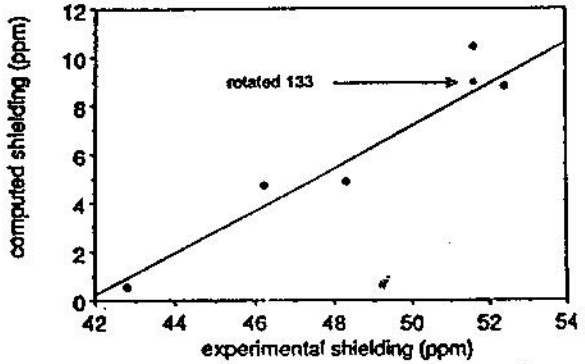


Figure 6: Plot of experimental *versus* theoretical shielding for each [5-F]Trp site in GBP. The experimental shieldings are those cited in ref 11. The computed values are referenced to a field-free value of 0 ppm. The open circle represents the shielding calculated for a 180° C^{β}-C^{γ} flip of Trp 133.

In addition, we have also been able to use the CFP-GIAO method to compute ^{19}F shifts in GBP, the results being in accord with the shielding polarizability approach (7). Thus, even the previously intractable problem of ^{19}F shifts in proteins has now been solved, using local reaction field, shielding polarizability and MD methods (7,11).

1. C. C. McDonald and W. D. Phillips, *J. Am. Chem. Soc.* **89**, 6332 (1967).
2. A. Allerhand, R. F. Childers and E. Oldfield, *Biochemistry* 12, 1335 (1973).
3. E. Oldfield, R. S. Norton and A. Allerhand, *J. Biol. Chem.*, **250**, 6381 (1975).
4. D. A. Torchia, S. W. Sparks and A. Bax, *Biochemistry* **28**, 5509 (1989).
5. K. D. Park, K. Guo, F. Adebodun, M. L. Chiu, S. G. Sligar and E. Oldfield, *Biochemistry*, 30, 2333 (1991).
6. C. Lian, H. Le, B. Montez, J. Patterson, S. Harrell, D. Laws, I. Matsumura, J. Pearson and E. Oldfield, *Biochemistry*, in press.
7. A. C. deDios, J. G. Pearson and E. Oldfield, *Science* 260, 1491 (1993).
8. J. G. Pearson, H. Le, A. C. de Dios and E. Oldfield, to be submitted to *J. Am. Chem. Soc.*
9. J. G. Pearson, J. F. Wang, J. L. Markley and E. Oldfield, submitted to *J. Am. Chem. Soc.*
10. D. D. Laws, A. C. deDios and E. Oldfield, *J. Biomol. NMR*, 3, 607 (1993).
11. J. G. Pearson, E. Oldfield, F. S. Lee and A. Warshel, *J. Am. Chem. Soc.*, **115**, 6851 (1993).
12. A. C. de Dios, J. G. Pearson and E. Oldfield, *J. Am. Chem. Soc.*, 115, 9768 (1993).
13. A. C. de Dios and E. Oldfield, *J. Am. Chem. Soc.*, in press.
14. H. Sternlicht and D. Wilson, *Biochemistry*, **6**, 2881 (1967).
15. M. P. Williamson, T. Asakura, E. Nakamura and M. Demura, J. *Biomol. NMR* 2, 83 (1992).
16. K. Osapay and D. A. Case, In Nuclear Magnetic Shielding and Molecular Structure (J. A. Tossell, Ed.), NATO ASI, Kluwer Academic Publishers, Dordrecht, p. 572 (1993).
17. D. A. Case and P. E. Wright, *NMR of Proteins*, eds G. M. Clore and A. M. Gronenborn, MacMillan, 1993, pgs. 53-91.
18. S. Spera and A. Bax, *J. Am. Chem. Soc.* 113, 5490 (1991).
19. D. B. Chesnut and K. D. Moore, *J. Comput. Chem.* **10**, 648 (1989).
20. J. Glushka, M. Lee, S. Coffin and D. Cowburn, *J. Am. Chem. Soc.*, 111, 7716 (1989); J. Glushka, M. Lee, S. Coffin and D. Cowburn, *J. Am. Chem. Soc.*, 112, 2843 (1989).
21. H. B. Le and E. Oldfield, unpublished results.
22. J. D. Augspurger, A. C. deDios, E. Oldfield and C. E. Dykstra, *Chem. Phys. Lett*, 213, 211 (1993).
23. F. S. Lee, Z. Chu and A. Warshel, *J. Comput. Chem.*, 14, 161 (1993).