

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/230753911>

# Protein Structure Refinement and Prediction via NMR Chemical Shifts and Quantum Chemistry

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · APRIL 1995

Impact Factor: 12.11 · DOI: 10.1021/ja00118a016

---

CITATIONS

88

---

READS

21

## 4 AUTHORS, INCLUDING:



Angel C de Dios

Georgetown University

103 PUBLICATIONS 2,892 CITATIONS

SEE PROFILE

# Protein Structure Refinement and Prediction via NMR Chemical Shifts and Quantum Chemistry<sup>†</sup>

**Hong-biao Le, John G. Pearson, Angel C. de Dios, and Eric Oldfield\***

*Contribution from the Department of Chemistry, University of Illinois at Urbana-Champaign,  
505 South Mathews Avenue, Urbana, Illinois 61801*

*Received April 28, 1994<sup>‡</sup>*

**Abstract:** An approach utilizing Bayesian probability and NMR chemical shifts to derive structural information about proteins is presented. The method is based on measurement of a spectroscopic parameter,  $P$  (such as a chemical shift or a coupling constant), which is then transformed *via* use of a corresponding parameter surface,  $P(\alpha, \beta)$ , into an unnormalized torsion angle probability or Z surface,  $Z(\alpha, \beta)$ . Using empirically determined parameter surfaces, the backbone  $\phi, \psi$  error between prediction and experiment is about  $17^\circ$ , but for 10 Ala residues in *Staphylococcal* nuclease, this reduces to  $\sim 10^\circ$  when quantum mechanically computed  $^{13}\text{C}$  shielding surfaces are utilized. The Z-surface approach permits unique combination of a wide variety of spectroscopic observables for refinement and prediction of protein structure in both solution- or solid-state systems.

## Introduction

During the past few years, nuclear magnetic resonance (NMR) spectroscopy has been shown to be a powerful technique with which to investigate the structures of proteins and peptides, either in solution<sup>1</sup> or in the solid or semisolid state.<sup>2</sup> In solution, structures are deduced primarily from distance information derived from nuclear Overhauser effects,<sup>1</sup> while in the solid state, both distance<sup>3,4</sup> and orientational approaches have been utilized.<sup>2,5,6</sup> In most cases, considerable additional spectroscopic information in the form of isotropic chemical shifts is usually available (for example, there are  $\sim 130\,000$  reported protein chemical shifts<sup>7</sup>), but there has been relatively little progress in the use of chemical shifts for structural analysis since the origins of folding-induced chemical shifts in proteins have been poorly understood.

Recently, however, we have shown<sup>8</sup> that NMR spectra of the heavier nuclei,  $^{13}\text{C}$  and  $^{15}\text{N}$  (and  $^{19}\text{F}$ ), in proteins can be satisfactorily predicted *via* use of quantum chemical methods,<sup>9</sup> which has led us to believe that it might be possible to deduce the detailed structural information encoded in the observed ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ) chemical shifts, a logical extension of the empirical chemical shift index method.<sup>10</sup> We show below that chemical

shifts do indeed encode a wealth of structural data, and we propose a method based upon a Bayesian treatment of the chemical shift, which we call the Z-surface approach, for extracting this information. We also show how additional (e.g.  $J$  coupling) spectroscopic information can be combined with chemical shift data to provide a new approach to protein structure refinement or determination. The Z-surface method is equally applicable to solid- and solution-state systems.

## Results and Discussion

The idea behind the Z-surface method is very simply to be able to predict the probability that a set of spectroscopic parameters,  $\{P_i\}$ , such as chemical shifts or spin–spin couplings, originates from a particular angular or orientational arrangement of atoms, such as a peptide  $\phi$  torsion angle, a **C–D** vector orientation in a membrane, or a more complex set of angles, such as  $\phi, \psi$  or  $\chi^1, \chi^2$  in a protein. For a parameter  $P$  which is a function of a single angle,  $\alpha$ ,

$$P = f(\alpha) \quad (1)$$

we define the probability that the experimental value of  $P$ ,  $P_{\text{expt}}$ , corresponds to a given angle  $\alpha$  as an unnormalized probability or Z surface

$$Z = e^{-\frac{(P_{\text{expt}} - f(\alpha))^2}{W}} \quad (2)$$

where  $W$  represents a search width (to take into account any computational inadequacies or experimental uncertainties). The Z surface is not normalized over all  $\phi$  and  $\psi$  in order to retain an absolute measure of how closely given  $\phi, \psi$  pairs match the experimental data set,  $\{P_i\}$ . To better illustrate the idea, Figure 1A represents a hypothetical spectroscopic parameter surface  $P = f(\alpha) = \cos^2 \alpha$ . For, say,  $P_{\text{expt}} = 0.5$  and  $W = 0.02$ , there are four likely  $\alpha$  solutions where  $Z = 1$  ( $-135^\circ$ ,  $-45^\circ$ ,  $45^\circ$ , and  $135^\circ$ ) as shown by the light regions in Figure 1B. This general type of “one-dimensional” Z surface can be generated using solution NMR parameters, e.g.  $^3J$  spin–spin couplings<sup>11</sup> or dipolar or quadrupolar splittings in membranes or crystals, where

<sup>†</sup> This work was supported by the U.S. National Institutes of Health (Grant HL-19481), in part by the American Heart Association with funds provided in part by the American Heart Association, Illinois Affiliate, Inc., (Grant AHA 92-013340) and by equipment grants from the International Business Machines Corp., the Shared University Research Instrumentation Program, and the Graduate Research Board of the University of Illinois. J.G.P. is a U.S. Public Health Service Postdoctoral Fellow (Grant GM-14545). H.L. is a University of Illinois Graduate Fellow. A.C.D. is an American Heart Association Inc. (Illinois Affiliate) Postdoctoral Research Fellow (Grant FW-01).

<sup>‡</sup> Abstract published in *Advance ACS Abstracts*, March 1, 1995.

(1) Billeter, M.; Qian, Y. Q.; Otting, G.; Müller, M.; Gehring, W.; Wüthrich, K. *J. Mol. Biol.* 1993, 234, 1084.

(2) Ketcham, R. R.; Hu, W.; Cross, T. A. *Science* 1993, 261, 1457.

(3) Hing, A. H.; Schaefer, J. *Biochemistry* 1993, 32, 7035.

(4) Peersen, O. B.; Yoshimura, S.; Hojo, H.; Armoto, S.; Smith, S. O. *J. Am. Chem. Soc.* 1993, 115, 4332.

(5) Rothgeb, T. M.; Oldfield, E. *J. Biol. Chem.* 1981, 256, 1432.

(6) Ulrich, A. S.; Heyn, M. P.; Watts, A. *Biochemistry* 1992, 31, 10390.

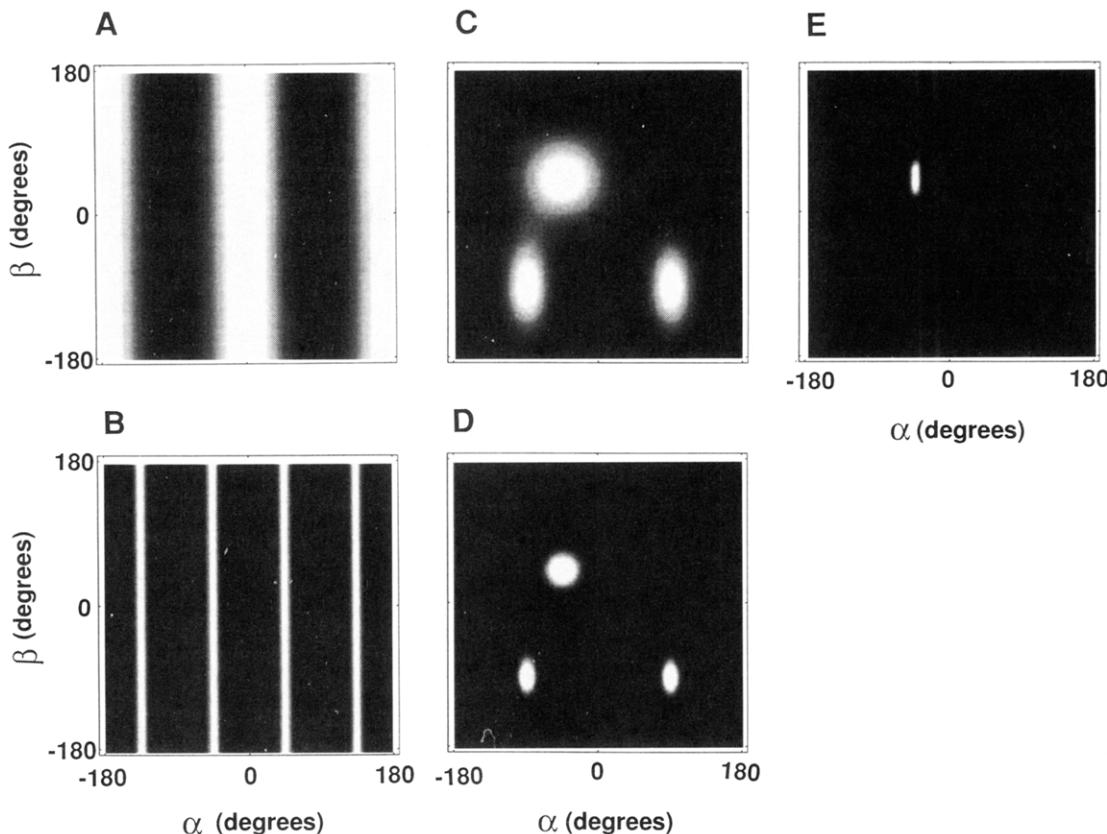
(7) Seavey, B. R.; Farr, E. A.; Westler, W. M.; Markley, J. L. *J. Biomol. NMR* 1991, 1, 217.

(8) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* 1993, 260, 1491.

(9) Ditchfield, R. *Mol. Phys.* 1974, 27, 789.

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

(11) Karplus, M. *J. Am. Chem. Soc.* 1963, 85, 2870. Bystrov, V. F. *Prog. NMR Spectrosc.* 1976, 10, 41.



**Figure 1.** Hypothetical parameter surface and Z-surface plots, where black corresponds to 0 and white to 1. (A) Parameter surface,  $P = f(\alpha) = \cos^2 \alpha$ ; (B)  ${}^1Z$  surface computed from A,  $Z = e^{-(P_{\text{expt}} - f(\alpha))^2/W}$ , where  $P_{\text{expt}} = 0.5$ ; (C) parameter surface,  $P = f(\alpha\beta) = \cos^6[(\alpha+45^\circ)/2] \cos^6[(\beta-45^\circ)/2] + \sin^6 \alpha \sin^6[(\beta-90^\circ)/2]$ ; (D) Z surface from C,  $Z = e^{-(P_{\text{expt}} - f(\alpha,\beta))^2/W}$ , where  $P_{\text{expt}} = 1.0$ ; (E)  ${}^2Z$  surface obtained by multiplying B and D. There are four likely solution regions for  $\alpha$  in B ( $f(\alpha) = 0.5$ ;  $\alpha = -135, -45, 45$ , and  $135^\circ$ ), three for  $f(\alpha\beta) = 1.0$ , but only one for  $f(\alpha) = 0.5$ ,  $f(\alpha\beta) = 1.0$ , Figure 1E, where  $\alpha = -45^\circ$  and  $\beta = 45^\circ$ .

$f(\alpha) = 3 \cos^2(\alpha-1)$ . For chemical shifts, much more complex relations are found,<sup>12</sup> and similar two-dimensional surfaces have

$$P \approx f(\alpha, \beta, \dots) \quad (3)$$

been reported for  ${}^1J$  and  ${}^2J$  couplings also.<sup>13,14</sup> Nevertheless, we can still use eq 2 to generate a single or  ${}^1Z$  surface (Figure 1C,D). Multiplication of the two individual  ${}^1Z$  surfaces (Figure 1B,D) generates what we designate as a  ${}^2Z$  surface, Figure 1E, which provides a set of angles,  $\alpha$  and  $\beta$ , most consistent with both spectroscopic parameters, yielding in this case a single most likely solution for  $\alpha$  and  $\beta$ . Thus, the four sets of solutions of Figure 1B and the three solutions of Figure 1D are reduced to a single solution by use of two, rather than one, spectroscopic parameters. Although both spectroscopic observables were given equal weight in this example, different observables may be given different weights by assigning different  $W$ 's for each  ${}^1Z$  surface, allowing the method to be applied to any situation where the behavior of a set of parameters,  $\{P\}$ , as a function of angles,  $\alpha, \beta, \dots$ , is known. We show below several realizations of the Z-surface approach using as examples the backbone  $\phi, \psi$  torsion angles in proteins. Examples of the use of chemical shift and  $J$  coupling parameter surfaces together, or chemical shift (shielding) surfaces alone, are described in what follows.

The results shown in Figure 1 are hypothetical and are given simply to illustrate the Z-surface idea. There are many existing

spectroscopic observables which have been determined for proteins, including  ${}^{13}\text{C}^\alpha, {}^{13}\text{C}^\beta, {}^1\text{H}^\alpha, {}^{13}\text{C}^\alpha, {}^{15}\text{N}^\text{H}$ , and  ${}^1\text{H}^\text{N}$  chemical shifts, as well as  ${}^1J_{\text{C}^\alpha\text{H}^\alpha}, {}^1J_{\text{C}^\alpha\text{N}}, {}^2J_{\text{C}^\alpha\text{N}}$ , and  ${}^3J_{\text{H}^\alpha\text{H}^\alpha}$  spin–spin couplings, all of which can be expressed as functions of either  $\phi, \psi$  or  $\phi$  or  $\psi$  with varying accuracy. In the protein *Staphylococcal* nuclease (+pdTp,  $\text{Ca}^{2+}$ ) SNase, there are 33 residues for which a large set of chemical shift and  $J$  coupling values have been reported,<sup>13</sup> providing a good initial test case for application of the Z-surface approach in structure prediction.

The first step is to compute shielding and  $J$  coupling parameter surfaces from purely experimental data.<sup>7,13</sup> For chemical shift surfaces, we use a fitting function similar to that described by Spera and Bax;<sup>12</sup> the resulting surface is the sum of convolutions of chemical shift and Gaussian distributions:

$$\sum(\phi, \psi) = \frac{\sum_i P(\phi_i, \psi_i) \exp\left\{-\left[\sin^2\left(\frac{\phi_i-\phi}{2}\right) + \sin^2\left(\frac{\psi_i-\psi}{2}\right)\right]/0.03\right\}}{\sum_i \exp\left\{-\left[\sin^2\left(\frac{\phi_i-\phi}{2}\right) + \sin^2\left(\frac{\psi_i-\psi}{2}\right)\right]/0.03\right\}} \quad (4)$$

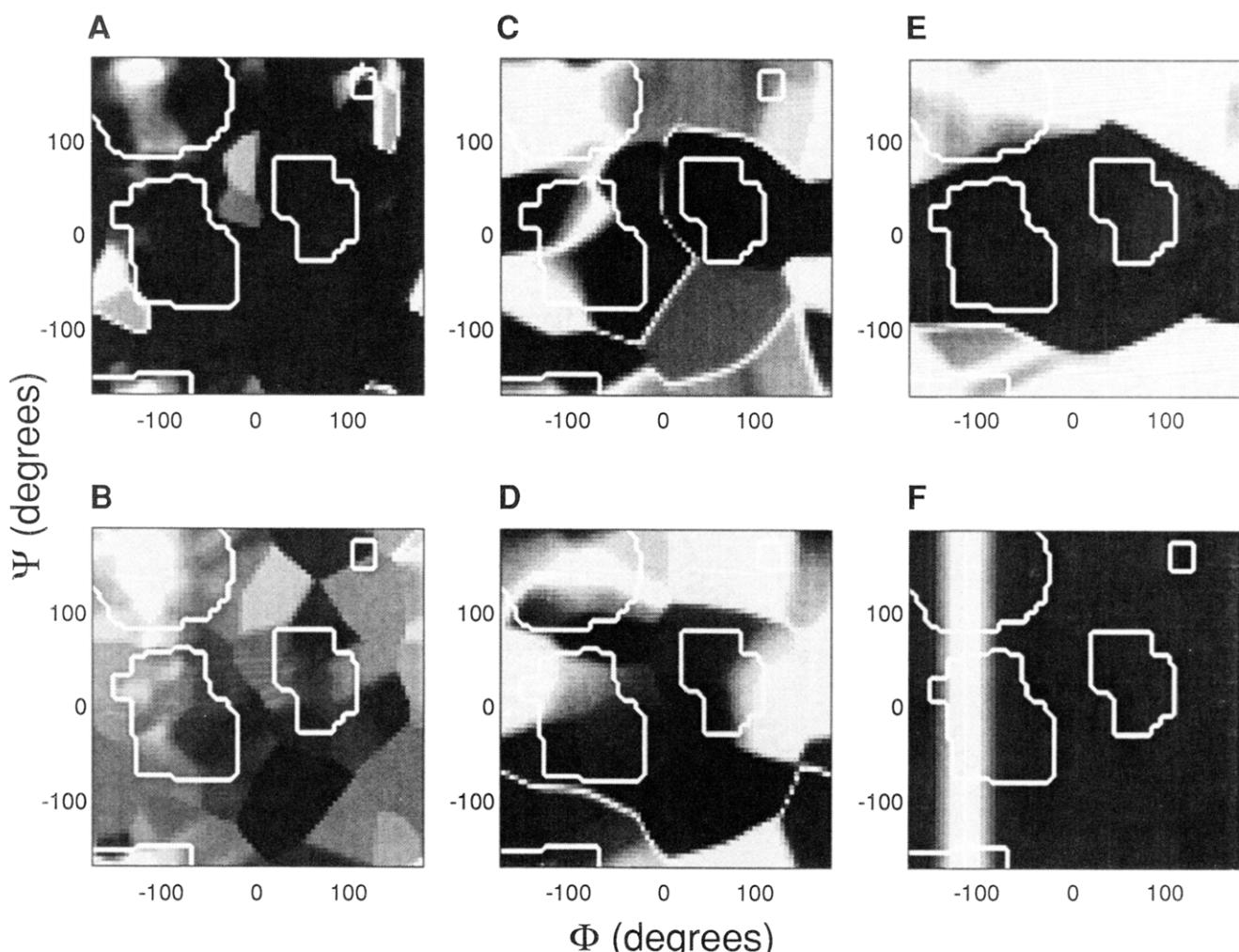
where  $P(\phi_i, \psi_i)$  is the secondary chemical shift<sup>12</sup> for a residue with torsion angles  $(\phi_i, \psi_i)$ , and  $\sum(\phi, \psi)$  is the computed secondary chemical shift parameter surface. A similar approach was then used for the  $J$  couplings. These surfaces were residue nonspecific, due to the lack of sufficient data for the compilation of residue specific empirical parameter surfaces.

The chemical shift and  ${}^1J$  and  ${}^2J$  coupling surfaces produced are analogous to the two-dimensional surface shown in Figure

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

(13) Vuister, G. W.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1992**, *3*, 67.

(14) Vuister, G. W.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **1992**, *114*, 9674. Delaglio, F.; Torchia, D.; Bax, A. *J. Biomol. NMR* **1991**, *1*, 439. Vuister, G. W.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 7772.



**Figure 2.** Representative Z-surface plots for Val<sup>74</sup> in SNase. (A)  ${}^1Z^{(C\alpha)}$ , (B)  ${}^1Z^{(H\alpha)}$ , (C)  ${}^1Z'J_{C\alpha H\alpha}$ , (D)  ${}^1Z'J_{C\alpha N}$ , (E)  ${}^1Z'J_{H^NH\alpha}$ , (F)  ${}^1Z'J_{H^NH\beta}$ . The well-defined surface regions (12) are shown inside the solid line. The regions outside those indicated have a very small number of data points.

1C, while the simpler  ${}^3J$  relation can be used as a one-dimensional filter analogous to that shown in Figure 1A or as a two-dimensional surface having an additional weak  $\psi$ -dependence. We then use the experimental chemical shift and  $J$  coupling constants to compute six individual  ${}^1Z$  surfaces for each residue, and a typical set of  ${}^1Z$  surfaces, for Val<sup>74</sup> in SNase, is shown in Figure 2A–F in which the light regions represent highly probable  $\phi, \psi$  solutions.<sup>12</sup> Multiplication of the individual  ${}^1Z$  surfaces yields a series of increasingly more refined surfaces,  ${}^1Z, {}^2Z, {}^3Z, \dots, {}^6Z$ , as shown in Figure 3A–F, with, in this case, the  ${}^6Z$  surface yielding a unique solution for  $\phi, \psi$  in good accord with that reported via X-ray crystallography,<sup>15</sup> Figure 3F. The procedure was then repeated for each of the other 32 residues in SNase where all six parameters were available, and we find root mean-squared deviation (rmsd) values between the X-ray and NMR data sets of 18 and 15° for  $\phi, \psi$ , respectively.

This level of agreement is quite promising, especially when one considers that it is obtained using purely empirical data to construct global (*i.e.* not residue specific) parameter surfaces. Both the effects of side-chain substitution and uncertainties associated with determining solution-state torsion angles from X-ray crystal structures limit the accuracy attainable using such empirical parameter surfaces. In particular, use of the empirical  ${}^{13}C^\beta$  chemical shift surface actually degrades the accuracy of

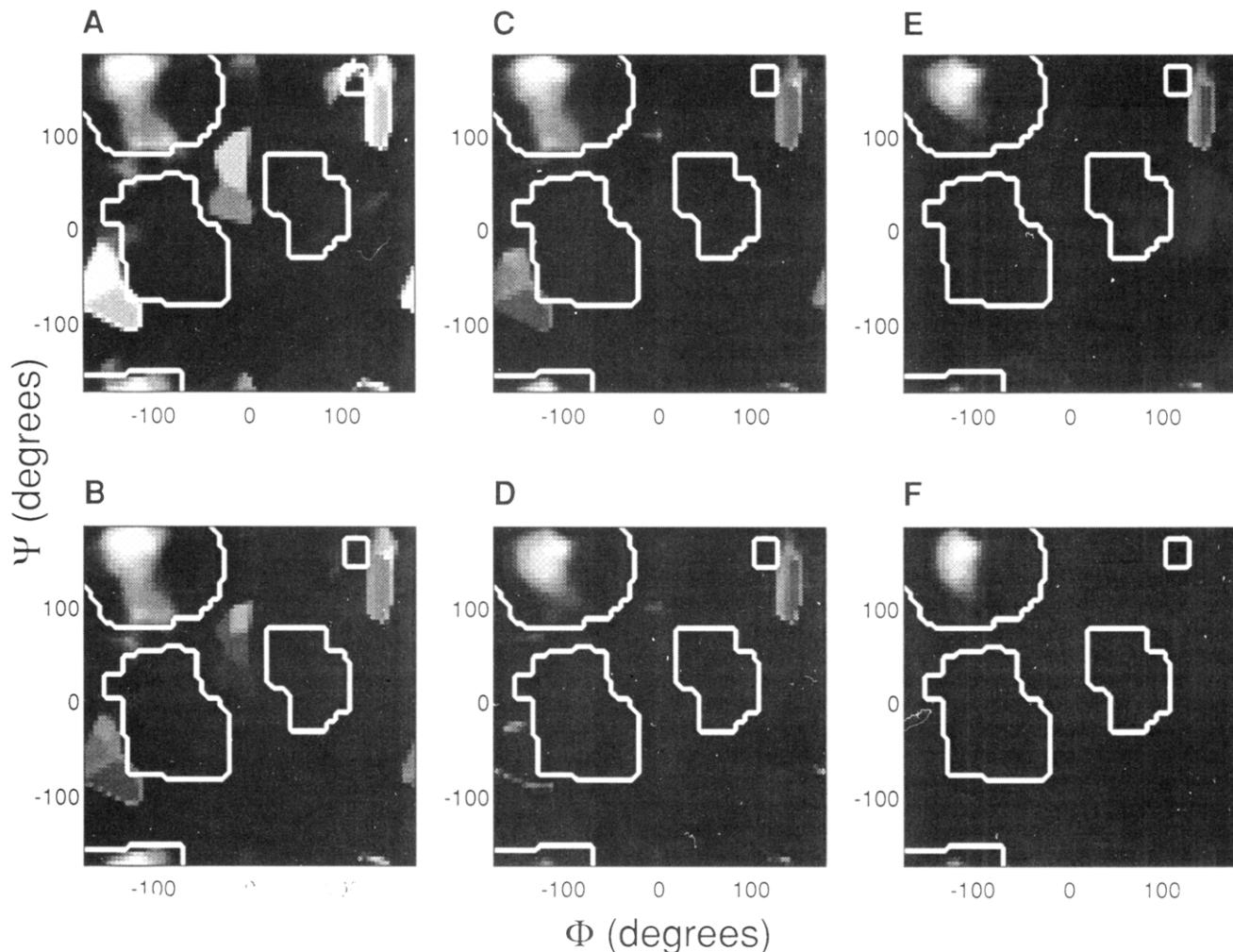
the solutions when compared with the X-ray structure. Improvement of the technique is discussed below and involves quantum chemical computation of individual amino acid shielding surfaces, allowing for the use of residue-specific shielding surfaces. These are free of experimental torsion angle uncertainties, can be evaluated at any  $\phi, \psi$  (and  $\chi$ ), and do not require large data bases, a problem when the rarer amino acids are of interest. We thus next consider the construction of typical  ${}^{13}C$  shielding surfaces using *ab initio* quantum chemical methods and show *inter alia* how they can be used to predict experimental chemical shifts, given  $\phi$  and  $\psi$ .

The  $\phi, \psi$  results shown in Figure 3 are expected to contain inaccuracies, since we now know that different amino acids have very different shielding surfaces.<sup>16,17</sup> That is, the global  ${}^{13}C^\alpha, {}^{13}C^\beta$  shift surfaces, reported by Spera and Bax, represent only general trends in  $\phi, \psi$  effects on shielding and cannot be used for accurate structural predictions. We have thus begun a quantum chemical study of the shielding surfaces of the heavy atoms in each naturally occurring amino acid in order to derive accurate shielding surfaces for each residue type. We use the gauge including atomic orbital (GIAO) quantum chemical method,<sup>9,18</sup> described previously,<sup>8,16</sup> to generate theoretical chemical shieldings as a function of  $\phi, \psi$ , using a locally-dense-

(16) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *J. Am. Chem. Soc.* **1993**, *115*, 9768.

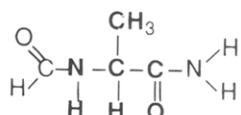
(17) de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*, 5307.

(18) Hinton, J. F.; Guthrie, P.; Pulay, P.; Wolinski, K. *J. Am. Chem. Soc.* **1992**, *114*, 1604.



**Figure 3.** Representative  $^nZ$ -surface plots for Val<sup>74</sup> in SNase. (A)  $^1Z^{\delta(C^\alpha)}$ , (B)  $^2Z^{\delta(C^\alpha)\delta(H^\alpha)}$ , (C)  $^3Z^{\delta(C^\alpha)\delta(H^\alpha)}J_{C^\alpha H^\alpha}$ , (D)  $^4Z^{\delta(C^\alpha)\delta(H^\alpha)}J_{C^\alpha H^\alpha}J_{C^\alpha N}$ , (E)  $^5Z^{\delta(C^\alpha)\delta(H^\alpha)}J_{C^\alpha H^\alpha}J_{C^\alpha N^2}J_{C^\alpha N}$ , (F)  $^6Z^{\delta(C^\alpha)\delta(H^\alpha)}J_{C^\alpha H^\alpha}J_{C^\alpha N^2}J_{C^\alpha N^3}J_{H^N H^\alpha}$ .

basis-set approach on suitable molecular fragments. For example, we use *N*-formyl-L-alanine amide as an alanine model compound



using a 6-311G++(2d,2p)//6-31G basis, with the bold atoms carrying the denser basis. We can thus generate  $^{13}C^\alpha$  shielding surfaces,  $\sigma(\phi,\psi)$ , and typical results for glycine, alanine, and valine ( $\chi^1 = 180^\circ$ ) are shown in Figure 4A-C.

In order to test the accuracy of the computed shielding surfaces for the Z-surface approach, it is essential to test their predictive ability; *i.e.*, they must first satisfactorily account for the experimental chemical shifts. Figure 4D shows that this is indeed the case for  $^{13}C^\alpha$  shielding in these three residues. Here, we have used the shielding surfaces from Figure 4A-C together with experimental (X-ray)  $\phi,\psi$  values to predict the theoretical shielding for 57  $^{13}C^\alpha$  sites of glycine, alanine, and valine residues in *Staphylococcal* nuclease and calmodulin.<sup>19-21</sup> The “goodness of fit” parameter,  $R^2$ , is 0.99, and the rmsd between theory and experiment is 1.6 ppm, giving us confidence in the ability of

quantum chemically determined shielding surfaces to model the effects of side-chain substitution and torsion angle variation upon the  $^{13}C^\alpha$  chemical shifts. The slope of  $\sim 0.9$  may be due in part to ro-vibrational differences between Gly and Ala/Val,<sup>22</sup> although basis deficiencies and small electrostatic field effects are also expected to contribute. In either case, the effects are extremely small, and the results shown in Figure 4D, read off of the corresponding shielding surfaces, compare favorably with the same 57  $^{13}C^\alpha$  shielding calculations reported previously, in which charge-field perturbation was explicitly included.<sup>22</sup>

The results of Figure 4 are also important since they show very clearly that  $^{13}C^\alpha$  sites of Gly, Ala, and Val ( $\chi^1 = 180^\circ$ ) have remarkably differently shaped shielding surfaces, Figure 4A-C. Our glycine results, Figure 4A, were obtained without *ab initio* structure optimization (which would add about 1 order of magnitude in time to the shielding calculations on an identical basis), using standard fragment geometries. In the regions of  $\phi,\psi$  space represented by the glycine residues in SNase and calmodulin, we find only a  $\sim 0.16$  ppm rmsd from shifts predicted with the glycine surface reported by Jiao *et al.*<sup>23</sup> obtained by using structure optimization. Thus, there seems to be no need for optimization. Moreover, since all three shielding surfaces are very different, it is clear that predictions made using empirical approaches will be quite prone to error.

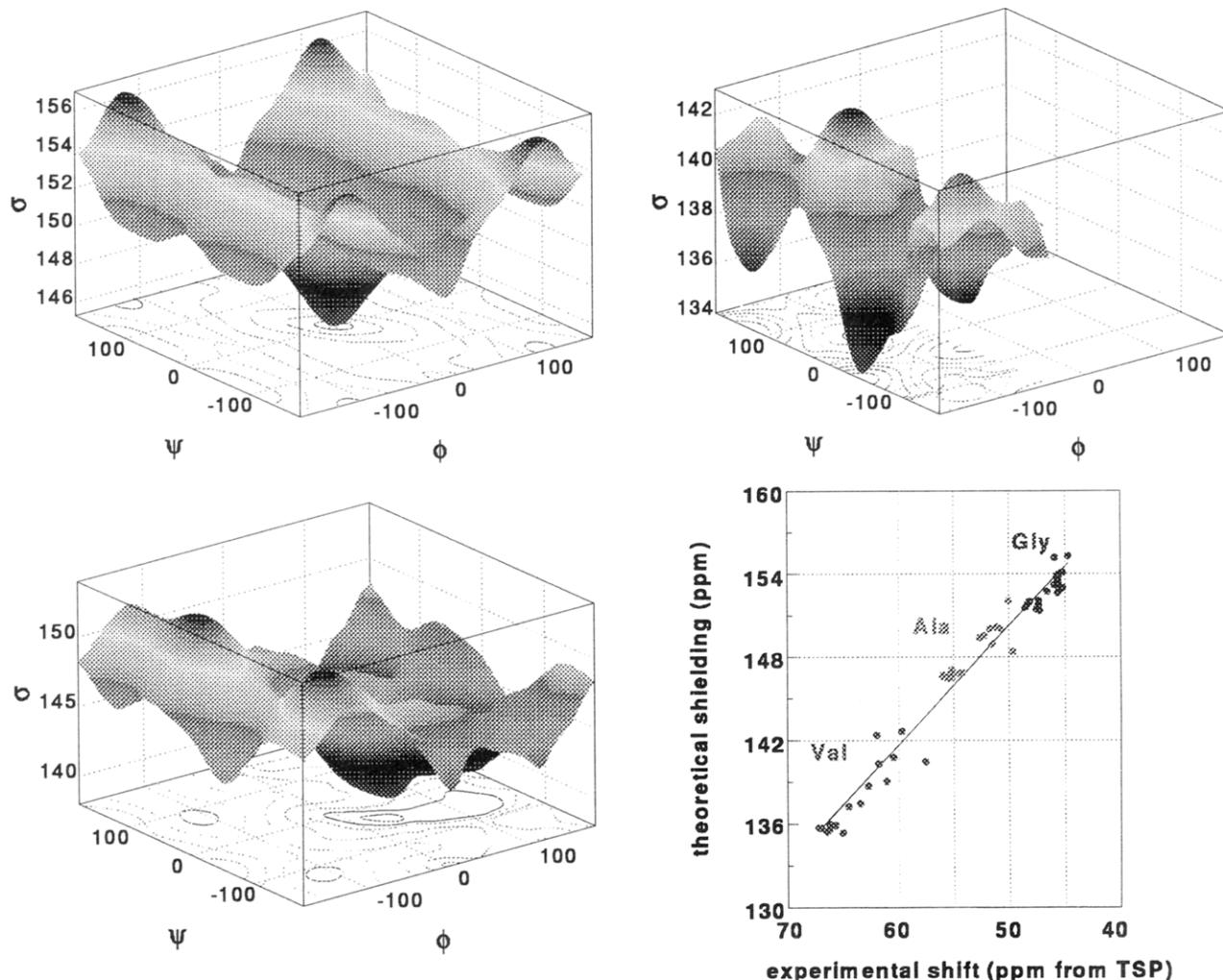
(19) All X-ray structures were obtained from the protein data base (PDB) of Brookhaven National Laboratory.

(20) T. Yamazaki and D. Torchia, private communication.

(21) M. Ikura, private communication.

(22) Laws, D. D.; de Dios, A. C.; Oldfield, E. *J. Biomol. NMR* **1993**, 3, 607.

(23) Jiao, D.; Barfield, M.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, 115, 10883.



**Figure 4.** Computed shielding surfaces for  $^{13}\text{C}^\alpha$  residues in glycine, alanine, and valine fragments and a graph showing experimental  $^{13}\text{C}^\alpha$  chemical shifts as a function of predicted shielding. (upper left) Glycine shielding surface, (upper right) valine  $\chi^1 = 180^\circ$  shielding surface, (lower left) alanine shielding surface, (lower right) graph showing experimental  $^{13}\text{C}^\alpha$  chemical shifts (in ppm from TSP) as a function of predicted  $^{13}\text{C}^\alpha$  shielding, rmsd  $\sim 1.6$  ppm. Chemical shieldings were evaluated by using a cluster of IBM RISC computers equipped with 40 GB of disk space and operating at a peak theoretical speed of 1.0 Gflops.

The final step in the Z-surface method is therefore to introduce quantum mechanically computed shielding surfaces for individual amino acid types into structure prediction. As noted above, different types of amino acids have different shielding surfaces, which if not accounted for will greatly affect the accuracy of the Z-surface predictions. We have calculated theoretical shielding surfaces for  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  of alanine in an attempt to make more accurate  $\phi, \psi$  predictions, and together with an empirical alanine  $\text{H}^\alpha$  shift surface (169 points, ref 24), we generate <sup>3</sup>Z-surface predictions of Ala  $\phi$  and  $\psi$  for the four proteins SNase, vertebrate calmodulin, interleukin-1 $\beta$ , and ribonuclease-H, solely using chemical shift information (which is more readily available than most  $J$  couplings).

Theoretical shielding surfaces were calculated for  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  nuclei in alanine using the *N*-formyl-L-alanine amide model fragment described above. We evaluated  $\sigma(^{13}\text{C}^\alpha)$  and  $\sigma(^{13}\text{C}^\beta)$  for 358  $\phi, \psi$  pairs distributed over the entire Ramachandran space. Then, the data points were fitted to a 67-term polynomial series of sinusoids  $\cos \phi, \cos \psi, \sin \phi, \dots, \sin^3 \psi$ , and all possible cross terms. These shielding surfaces were then calibrated against the experimental chemical shifts of Ala  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$ .

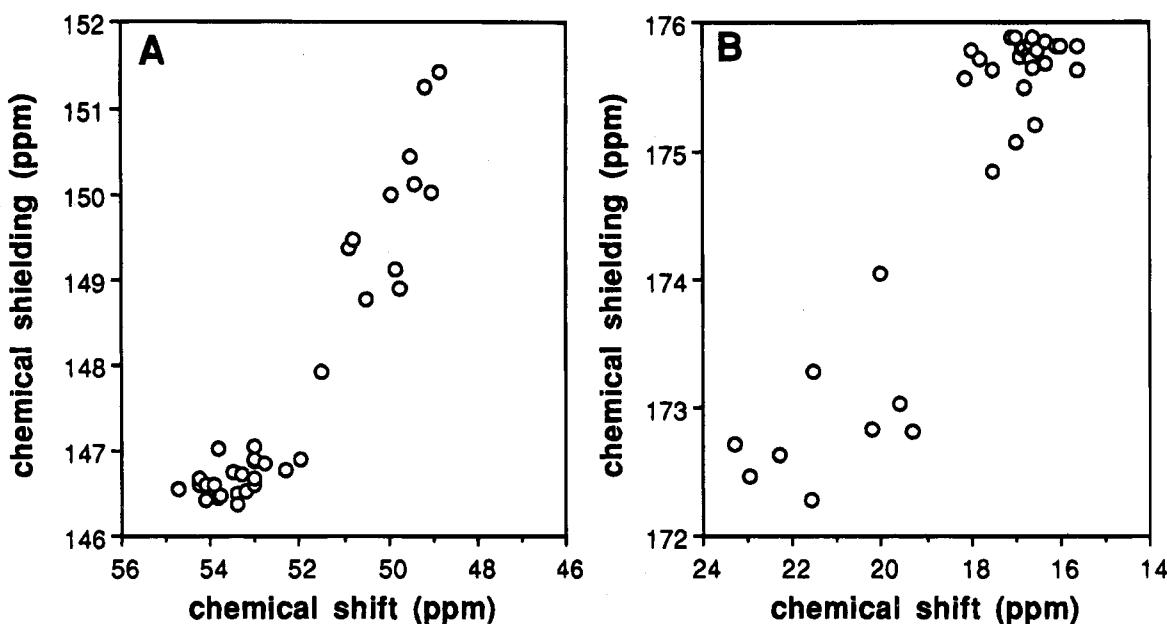
(24) The following PDB files were used for  $^1\text{H}^\alpha$  surface generations: 1SNC, 1CLL, 2RN2, 4IIB, 5PTI, 1F3G, 2GPR, 1BB4, 1LYD, 1LZ1, 2ABX, 2LYM, 2LET, 2TRX, 2WRP, 7LZM, and 9RNT. Unligated SNase was from Fox *et al.* (private communication).

in the four proteins, using X-ray torsion angles as input.<sup>15,19–21,25,26</sup> The results are shown in Figure 5, in which the “back-calculated” chemical shieldings (from X-ray  $\phi, \psi$  values) are reported as a function of the experimental chemical shift. Three out of 39 residues appeared to be “outliers”, falling well-off of a generally good correlation. These points were deleted from the calibration procedure, since we believe they represent either misassignments, large crystal–solution structural differences, or motionally averaged shifts. Crystal–solution structural differences (*e.g.*  $\chi^1$  of valine in calmodulin) are well-known (ref 17 and references cited therein), and the ribonuclease-H data were from a 2D rather than a 3D NMR study, and such lower dimensional assignments are more prone to error.

We then converted the shielding surfaces into chemical shift surfaces by adjusting the shielding surface amplitude with the linear regression coefficients from the data shown in Figure 5. Since we believe the rmsd’s between the calibrated chemical shift surfaces and the experimental chemical shifts are the best measure of the uncertainty in our prediction scheme, they were used to determine the width parameter,  $W$ , used in eq 2. Specifically,  $W$  was set to be equal to twice the mean-square

(25) Clore, G. M.; Bax, A.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. *Biochemistry* 1990, 29, 8172.

(26) Yamazaki, T.; Yoshida, M.; Kanaya, S.; Nakamura, H.; Nagayama, K. *Biochemistry* 1991, 30, 6036.



**Figure 5.** Calibration plots for alanine  $^{13}\text{C}^\alpha, ^{13}\text{C}^\beta$  chemical shifts. (A)  $\sigma^{C^\alpha}(\phi, \psi) \rightarrow \delta^{C^\alpha}(\phi, \psi)$  and (B)  $\sigma^{C^\beta}(\phi, \psi) \rightarrow \delta^{C^\beta}(\phi, \psi)$ . In both cases, Ala chemical shieldings were derived from theoretically calculated shielding surfaces,  $\sigma(\phi, \psi)$ , using X-ray structure torsion angles for Ala's in *Staphylococcal* nuclease, calmodulin, interleukin 1- $\beta$ , and ribonuclease-H. Three of the 39 Ala's were deemed to be outliers (SNase 109 and Rn-H 51 and 125), presumably due to either resonance misassignment and/or crystal-solution structural differences. Similarities in regression coefficients and goodness of fit parameters between the total data set (36 Ala's,  $\sigma^{C^\alpha} = 192.2 \text{ ppm} - 0.86\delta^{C^\alpha}$ ,  $R^2(C^\alpha) = 0.90$ ;  $\sigma^{C^\beta} = 184.4 \text{ ppm} - 0.54\delta^{C^\beta}$ ,  $R^2(C^\beta) = 0.85$ ), the SNase-only data set (11 Ala's,  $\sigma^{C^\alpha} = 189.8 \text{ ppm} - 0.81\delta^{C^\alpha}$ ,  $R^2(C^\alpha) = 0.93$ ;  $\sigma^{C^\beta} = 184.0 \text{ ppm} - 0.51\delta^{C^\beta}$ ,  $R^2(C^\beta) = 0.86$ ), and the Cal + IL-1 $\beta$  + Rn-H data set (25 Ala's,  $\sigma^{C^\alpha} = 193.2 \text{ ppm} - 0.88\delta^{C^\alpha}$ ,  $R^2(C^\alpha) = 0.88$ ;  $\sigma^{C^\beta} = 185.2 \text{ ppm} - 0.58\delta^{C^\beta}$ ,  $R^2(C^\beta) = 0.84$ ) indicate that the calibration coefficients are not dominated by any individual protein. The W values used in the text were obtained from these results. For  $C^\alpha$ ,  $W = 0.52 \text{ ppm}^2$ , and for  $C^\beta$ ,  $W = 0.96 \text{ ppm}^2$ , where  $W$  is twice the mean-squared deviation between experiment and prediction.

variation (Figure 5) between experiment and prediction ( $0.52 \text{ ppm}^2$  for  $C^\alpha$ ,  $0.96 \text{ ppm}^2$  for  $C^\beta$ ). The larger  $W$  corresponds to a decrease in “predictive power” of the given  $Z$  surface. “Self-points” were removed from the calibration plots of  $\sigma(\phi, \psi)$  surfaces for each residue, generating slightly different  $\delta(\phi, \psi)$  surfaces for each residue. For example, the  $\delta(\phi, \psi)$  surfaces used to make torsion angle predictions for Ala<sup>60</sup> in SNase were calibrated against experimental data with Ala<sup>60</sup> data removed.

Here, it is important to consider the implications of these surface scalings. For  $^{13}\text{C}^\alpha$ , slopes of  $\sim -0.86$  are in good accord with experiment, and on the basis of the excellent agreement between experimental chemical shifts and those deduced from shielding surfaces, we conclude that our fragment geometries and bases are adequate. For  $^{13}\text{C}^\beta$ , the observed slopes of  $\sim -0.55$  cannot be due to an inadequate basis, since the calculated values are already basis-set saturated. Since we find a 0.91 correlation coefficient,  $R$ , between predicted and experimental  $^{13}\text{C}^\beta$  shifts, very good agreement even though the slope is only  $-0.55$ , our results suggest that electrostatic, hydrogen bonding, or “other” non- $\phi, \psi$  contributions to shielding are involved and are in fact correlated with  $\phi, \psi$ . The observation that hydrogen bonding and charge field perturbation do permit a more accurate prediction of  $^{13}\text{C}^\beta$  shifts<sup>8,27</sup> indicates these are the major origins of the decreased slope. In addition, the observation that the predictions for  $^{13}\text{C}^\alpha$  (expected to require an even larger basis) are accurate also implies a significant nonlocal contribution to shielding for alanine  $^{13}\text{C}^\beta$ . On the basis of our work with pentapeptides, it appears likely that adjacent peptide carbonyls contribute to  $^{13}\text{C}^\beta$  shielding by at least 1 ppm.<sup>17</sup>

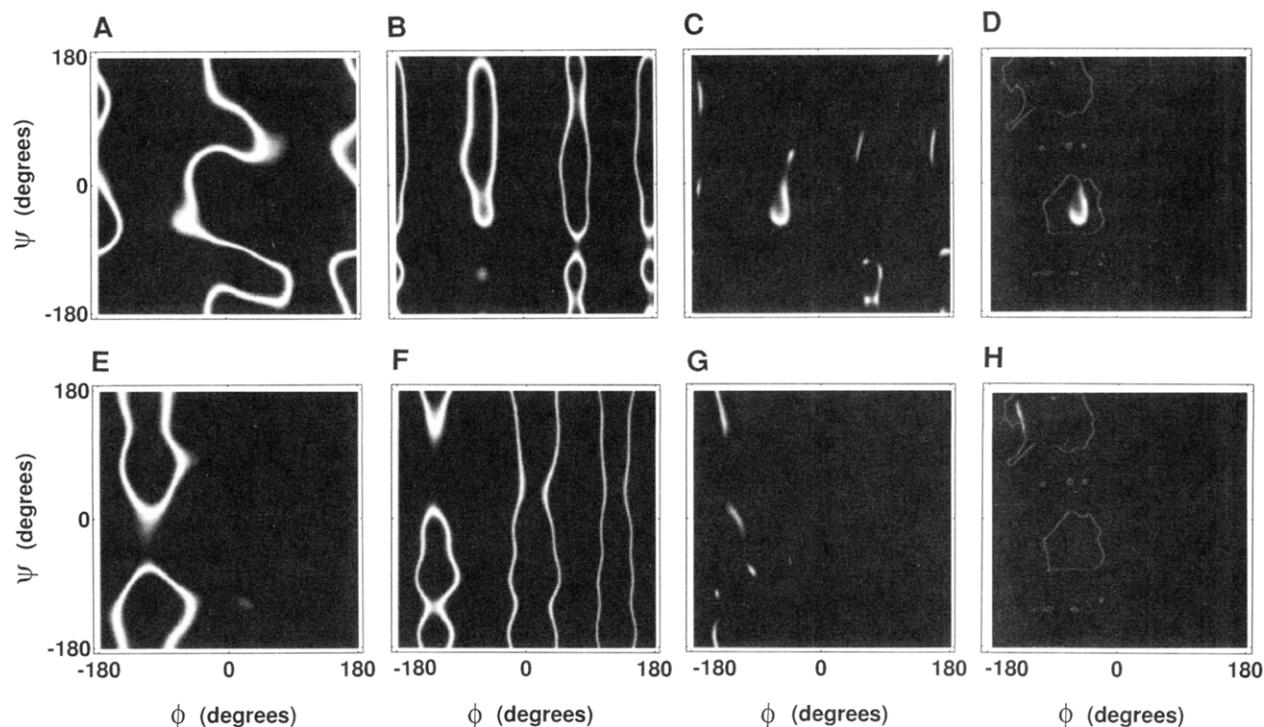
We now consider Z-surface results for some specific residues in *Staphylococcal* nuclease. Figure 6A shows a  $^1Z$  surface for  $^{13}\text{C}^\alpha$  in a typical helical residue, Ala<sup>60</sup> in SNase, while Figure

6B shows the corresponding  $^1Z$  surface for  $^{13}\text{C}^\beta$ , both obtained by using calibrated shielding surfaces. Helical Ala residues show a characteristic “S”-shaped  $^{13}\text{C}^\alpha$   $^1Z$  surface, while  $^{13}\text{C}^\beta$  results are typically less  $\psi$ -dependent. Multiplication of both  $^1Z$  surfaces yields the  $^2Z$  surface shown in Figure 6C, in which a prominent helical  $\phi, \psi$  solution is apparent. This can be further refined by use of a  $\delta^{H^\alpha}$  surface to yield the  $^3Z$  surface shown in Figure 6D, which has  $\phi, \psi = -66.0, -44.2^\circ$ , to be compared with the X-ray values of  $-65.7, -40.6^\circ$ .<sup>19</sup> At present, only empirical  $^1H^\alpha$  chemical shift surfaces are available; thus, we again add a box (Figure 6D) in which the enclosed surface is well-defined by the data. In the future, however, it should be possible to compute  $^1H^\alpha$  shielding surfaces, as well as  $J$  coupling surfaces, allowing for the elimination of this requirement.

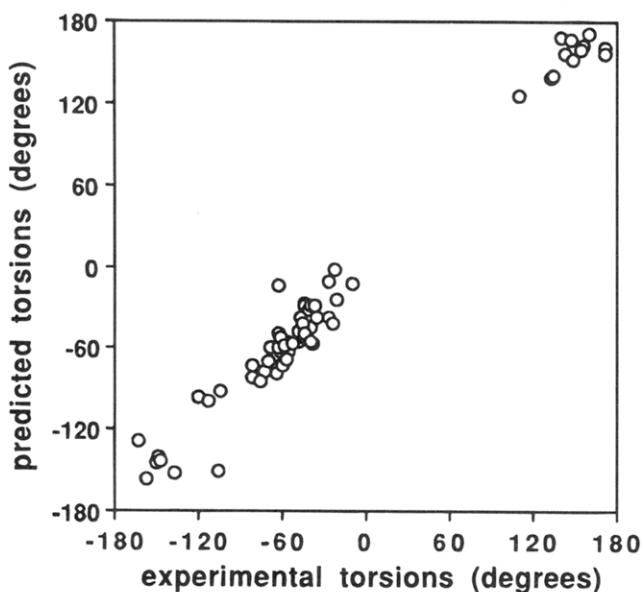
For a typical sheetlike residue, Ala<sup>112</sup> in SNase, we obtain the corresponding Z-surface results shown in Figure 6E–H. Ala residues having sheetlike  $\phi, \psi$  values have characteristic “oval”  $^1Z$   $^{13}\text{C}^\alpha$  surfaces, as shown in Figure 6E, which when multiplied by the  $^{13}\text{C}^\beta$   $^1Z$  surface yield relatively well-defined  $^2Z$  sheet solutions, as shown in Figure 6G. Again, use of an  $H^\alpha$  Z surface gives a further improvement; in this case, we find  $\phi, \psi = -141.0, 156.7^\circ$ , to be compared with the X-ray crystallographic values of  $-148.0, 172.1^\circ$ . A much larger series of  $\phi, \psi$  predictions are compared with crystallographic values in Figure 7.

Here, we again consider  $\phi, \psi$  predictions for a total of 39 Ala residues in four proteins: ligated SNase, vertebrate calmodulin, interleukin-1 $\beta$ , and ribonuclease-H. For 37 of these residues, we predict rmsd's of  $12.6^\circ$  in  $\phi$  and  $14.4^\circ$  in  $\psi$  using only three shift restraints, good agreement between the X-ray results and the  $^3Z$ -surface predictions. For two residues, Ala<sup>51</sup> and Ala<sup>125</sup> in RNase, we find large errors. What are the likely reasons for these apparent errors? First, the method may not adequately account for all interactions in all residues. Second, there could be errors in the NMR assignments or X-ray

(27) J. G. Pearson, H. Le, J. F. Wang, J. L. Markley, and E. Oldfield, unpublished results.



**Figure 6.** Z-surface prediction of peptide backbone  $\phi, \psi$  angles for representative alanine helical and sheet residues in SNase using chemical shift information. (A) Ala<sup>60</sup>  $^1Z^{(\text{C}^\alpha)}$ , (B) Ala<sup>60</sup>  $^1Z^{(\text{C}^\beta)}$ , (C) Ala<sup>60</sup>  $^2Z^{(\text{C}^\alpha)\delta(\text{C}^\beta)}$ , (D) Ala<sup>60</sup>  $^3Z^{(\text{C}^\alpha)\delta(\text{C}^\beta)\delta(\text{H}^\alpha)}$ , (E) Ala<sup>112</sup>  $^1Z^{(\text{C}^\alpha)}$ , (F) Ala<sup>112</sup>  $^1Z^{(\text{C}^\beta)}$ , (G) Ala<sup>112</sup>  $^2Z^{(\text{C}^\alpha)\delta(\text{C}^\beta)}$ , (H) Ala<sup>112</sup>  $^3Z^{(\text{C}^\alpha)\delta(\text{C}^\beta)\delta(\text{H}^\alpha)}$ . White = probable solution. Experimental  $^{13}\text{C}$  chemical shift data sets were referenced against one another by using mean experimental lysine  $^{13}\text{C}^\delta$  and  $^{13}\text{C}^\epsilon$  chemical shifts as internal standards. The experimental Ala H $^\alpha$  shift surface was based on 169 points. For the empirical  $\delta^\text{H}^\alpha$  surface, we used the gradient of the shift surface,  $\nabla\delta(\phi, \psi)$ , to delineate unphysical solutions (due to lack of data points or edge effects),  $\nabla\delta \geq 0.5 \text{ ppm deg}^{-1}$  being chosen as an upper bound (white line in D and H).



**Figure 7.** Comparison between experimental (X-ray) and predicted (Z-surface)  $\phi, \psi$  values for alanine residues in SNase, calmodulin, interleukin-1 $\beta$ , and ribonuclease-H.

structures. Third, there could be crystal–solution structural differences. For Ala<sup>125</sup> in ribonuclease-H, both  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  resonances are 2–3 ppm shifted from their expected positions based on  $\phi, \psi$ , which could indicate a large crystal–solution structural difference or an alternative chemical shift assignment in the 2D NMR spectrum.<sup>26</sup> For Ala<sup>51</sup> in ribonuclease-H, the  $^2Z$  surface shows no likely solutions, *i.e.*  $^2Z(\phi, \psi) < 0.1$  for all  $\phi$  and  $\psi$ . This is not surprising, for both Ala<sup>51</sup> and Ala<sup>125</sup> in ribonuclease-H were among the three residues excluded from the calibration procedure, again indicating the presence of either

large crystal–solution structural differences, errors in the resonance assignments, or dynamical effects.

Finally, we combined the theoretical  $^1Z$  surfaces for  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  chemical shifts together with empirical  $^1\text{H}^\alpha$  and  $J$  coupling results for the 12 Ala residues in SNase to try to estimate the current limitations of the method. Not all parameters were available for each alanine; thus, we obtained from  $^5Z$  to  $^7Z$  surface predictions, as shown in Figure 7. For 22 out of 24  $\phi, \psi$  values, the rmsd between solution NMR prediction and the X-ray result was  $\sim 9^\circ$ , very good agreement given the empirical nature of most surfaces, although we were surprised to find for the Ala residues involved that this in fact did not represent an improvement over the  $^3Z$  surface results described above. For Ala<sup>94</sup> and Ala<sup>109</sup>, the  $\phi$  error was  $\sim 30^\circ$ , regardless of the number of surfaces used beyond  $^3Z$ , possibly indicating small crystal–solution structural differences. This could also explain the lack of improvement on addition of  $J$  couplings to the  $^3Z$  solutions. That is to say  $\sim 10^\circ$  could represent the *real* rmsd between solution and X-ray  $\phi, \psi$  values.

Evidence to support this idea is that we have computed NOE-based solution structures of SNase ( $+Ca^{2+}$ ,  $+pdTp$ ) using X-PLOR<sup>27,28</sup> and the Z-surface  $\phi, \psi$  predictions for alanine,<sup>27</sup> and these structures (including hydrogen-bonding and electrostatics) permit accurate  $^{13}\text{C}^\alpha, ^{13}\text{C}^\beta$  chemical shift predictions. There are no significant increases in NOE violations in NOE + shift structures,<sup>27</sup> while use of the X-ray structures alone gives significant shift errors for predicted Ala  $^{13}\text{C}^\alpha, ^{13}\text{C}^\beta$  chemical shifts. Of course, there are situations in which simple application of the Z-surface method will be inaccurate, *e.g.* in the presence of internal motion. However, it is worth noting that the use of shielding surfaces,  $\sigma(\phi, \psi)$ , permits an evaluation of  $\sigma$  (or  $\delta$ ) in the presence of motion *via* use of the shielding trajectory

approach,<sup>8</sup> allowing future structural models to be tested against observed chemical shifts even in the presence of molecular motion.

We should finally note that interpretation of our joint probability surfaces strictly in terms of a single most probable value, rather than a range, does not make full use of the information content of the Z surfaces. In work to be described elsewhere,<sup>27</sup> we will describe the incorporation of full chemical shift "Z surfaces into an NOE/J coupling/hydrogen bonding structure determination using X-PLOR,<sup>28</sup> including an extrapolation of this study to analysis of  $\phi,\psi$  and  $\chi$  results for valine.<sup>27</sup> The NOE structure refinement obtained using chemical shift information again has an  $\sim 10\text{--}15^\circ$  rmsd *versus* the X-ray results, with the mean  $\phi,\psi$  values being in close agreement with those found in the present work.

### Conclusions

The results we have shown above are of interest for a number of reasons. First, we have introduced an alternative method for predicting elements of protein structure, the Z-surface approach, which is in principle equally applicable in the solid or liquid state and permits incorporation of isotropic chemical shift, spin–spin ( $J$  or dipolar) couplings, quadrupolar, and anisotropic chemical shift data, depending on the nucleus and the state of the system under investigation. Second, we have shown how a variety of experimental chemical shift and  $J$  coupling results can be combined together in order to predict backbone,  $\phi,\psi$  torsion angles in a protein, *Staphylococcal*

nuclease, with an  $\sim 17^\circ$  rmsd between theory and experiment (X-ray). Third, we have shown how quantum mechanically computed shielding surfaces can be used to predict  $^{13}\text{C}$  chemical shifts in proteins, with an rmsd of  $\sim 1\text{--}2$  ppm for  $^{13}\text{C}^\alpha$  of glycine, alanine, and valine residues. Fourth, we have shown how such quantum chemically predicted shielding surfaces can be used to help predict protein  $\phi,\psi$  values. For 22 out of 24 Ala  $\phi,\psi$  torsion angles in SNase, the rmsd is  $\sim 10^\circ$ . Thus, especially when combined with additional restraints, from NOE data,  $J$  couplings, quadrupolar splittings, anisotropic chemical shifts, etc., the combined use of isotropic chemical shift surfaces from quantum chemical calculation and the Z-surface approach should provide a useful way of facilitating protein structure prediction, refinement, and validation. In solution NMR, the method can be expected to be of most use in situations where there are too few NOE restraints to define structure, while in the solid state, the method can be used to complement both distance- and orientation-based methods and should be of particular use for investigation of random powder samples, where often only limited amounts of distance information may be available.

**Acknowledgment.** We thank P. Pulay, J. F. Hinton, and K. Wolinski for providing their GIAO program; G. Vuister for providing  $J$  coupling constant data; and A. Bax, C. E. Dykstra, A. Gronenborn, C. Jameson, J. Markley, D. Torchia, and J. Wand for helpful comments.

JA9413022