A New Analysis of Proton Chemical Shifts in Proteins

Klára Ösapay and David A. Case*

Contribution from the Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037. *Received April 12, 1991*

Abstract: We present an empirical analysis of proton chemical shifts from 17 proteins whose X-ray crystal structures have been determined. The crystal structures are used to estimate the conformation-dependent part of the shift, that is, the difference between the observed shift and that of a "random-coil" linear peptide. The results indicate that a significant improvement over ring-current theories can be made by including the effects of the magnetic anisotropy of the peptide group and estimates of backbone electrostatic contributions. For 5678 protons bonded to carbon, we find a linear correlation coefficient of 0.88 between calculated and observed structural shifts, with a root-mean-square error of 0.23 ppm; contributions from the peptide group are especially noticeable for protons at the Cα position. If we consider only side-chain protons in non-heme proteins, the rms error is 0.18 ppm, and methyl protons show an rms error of 0.13 ppm. New estimates of intensity factors for various ring-current contributions are given (including those arising from the heme group) which suggest more nearly equal contributions from various rings than found in earlier studies. Predictions for protons bonded to nitrogen are much poorer than for protons bonded to carbon, but significant qualitative insights can be obtained. Prospects for using calculated chemical shifts in the final refinement of protein solution structures are discussed.

Introduction

Advances in NMR instrumentation and methodology have now made it possible to determine site-specific proton chemical shift assignments for a large number of proteins.[1] It has been known for some time that the "structural" chemical shifts (the differences between the resonance positions in a protein and in a "random-coil" polypeptide[2]) carry useful information about the structure. For example, Cα protons are predominantly shifted downfield in β-sheets and upfield in α-helices,[3-7] and similarities among homologous proteins have often been used to help assign the spectra or to interpret the shifts.[8,9] Here we report progress toward a more quantitative theory, based on an analysis of proton chemical shifts in 17 proteins whose crystal structures are known; these results are a significant extension on earlier studies, carried out nearly a decade ago, that concentrated on the pancreatic trypsin inhibitor and lysozyme.[10-12]

It is useful to decompose contributions to proton chemical shifts into local and nonlocal contributions:[13]

$$\delta_{tot} = \delta_d(\text{local}) + \delta_p(\text{local}) + \delta_{rc} + \delta_m + \delta_{el} + \delta_s \qquad (1)$$

The first two terms represent local "diamagnetic" and "paramagnetic" contributions; here we will assume that these terms can be approximated by the observed shifts of short peptides that appear to be in random-coil conformations.[2] The final four terms in eq 1 represent contributions from more distant parts of the molecule deriving from aromatic ring currents, other magnetic anisotropies (in our case, from the peptide groups), and electrostatic and solvent effects. Although this decomposition is neither complete nor unique, it represents well a considerable amount of empirical knowledge about chemical shift propensities.[13]

The conformation-dependent contribution to chemical shifts that is best understood is the ring-current contribution associated with conjugated groups. Various functional forms have been tested and calibrated against observed data.[14] For protons bound to side-chain carbons, these theories give a reasonable explanation for many of the large structural shifts, with linear correlations between calculated and observed shifts in the range 0.7–0.8, and root-mean-square (rms) errors of 0.2–0.4 ppm.[11,15,16] Ring-current calculations can occasionally be useful in helping to assign spectra,[8,15] but there are a significant number of examples where large discrepancies occur between the predicted and observed resonance positions. Further, many side-chain protons exhibit structural shifts of up to 1 ppm where no rings are nearby and no ring-current contribution is expected (or calculated). Correlations between observed structural shifts and ring-current results for backbone protons (at the NH and CαH positions) are generally poorer than those for side chains. These results (which have generally been known for some time,[11,12,16,17]) suggest that additional mechanisms are contributing to the observed structural shifts in proteins.

Two prominent candidates for additional contributions to the shielding tensors are the magnetic anisotropy of the peptide group and electrostatic or hydrogen-bonding effects. The former has been considered by a number of authors,[3,18-20] mainly based on analogies to carbonyl shifts in ketones. Until now, the lack of a large number of protein assignments has limited its application in proteins. Nonlocal electrostatic effects can be considered to arise from their ability to polarize the C–H bond, so that greater or less electron density is in the vicinity of the proton itself.[21] Although empirical estimates can be made of the expected magnitude of this contribution, applications to proteins are plagued by well-known problems of understanding electrostatic interactions at a microscopic level.[22,23] An additional complication is that some portion of the solvent interaction (especially for water)

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(2) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285-297.
(3) Clayden, N. J.; Williams, R. J. P. *J. Magn. Reson.* **1982**, *49*, 383-396.
(4) Dalgarno, D. C.; Levine, B. A.; Williams, R. J. P. *Biosci. Repl.* **1983**, *3*, 443-452.
(5) Gross, K.-H.; Kalbitzer, H. R. *J. Magn. Reson.* **1988**, *76*, 87-99.
(6) Szilágyi, L.; Jardetzky, O. *J. Magn. Reson.* **1989**, *83*, 441-449.
(7) Pastore, A.; Saudek, V. *J. Magn. Reson.* **1990**, *90*, 165-176.
(8) Reid, D. G.; Saunders, M. R. *J. Biol. Chem.* **1989**, *264*, 2003-2012.
(9) Weiss, M. A.; Hoch, J. C. *J. Magn. Reson.* **1987**, *72*, 324-333.
(10) Perkins, S. J.; Wüthrich, K. *Biochim. Biophys. Acta* **1979**, *576*, 409-423.
(11) Perkins, S. J.; Dwek, R. A. *Biochemistry* **1980**, *19*, 245-258.
(12) Pardi, A.; Wagner, G.; Wüthrich, K. *Eur. J. Biochem.* **1983**, *137*, 445-454.
(13) Harris, R. K. *Nuclear Magnetic Resonance Spectroscopy, A Physicochemical View*; Longman, Scientific & Technical: Essex, England, 1986.
(14) Haigh, C. W.; Mallion, R. B. *Prog. NMR Spectrosc.* **1980**, *13*, 303-344.
(15) Redfield, C.; Hoch, J. C.; Dobson, C. M. *FEBS Lett.* **1983**, *159*, 132-136.
(16) Gippert, G. P.; Yip, P. F.; Wright, P. E.; Case, D. A. *Biochem. Pharm.* **1990**, *40*, 15-22.
(17) Wagner, G.; Pardi, A.; Wüthrich, K. *J. Am. Chem. Soc.* **1983**, *105*, 5948-5949.
(18) Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 2293-2294.
(19) Zürcher, R. F. *Prog. NMR Spectrosc.* **1967**, *2*, 205-257.
(20) ApSimon, J. W.; Beierbeck, H. *Can. J. Chem.* **1971**, *49*, 1328-1334.
(21) Buckingham, A. D. *Can. J. Chem.* **1960**, *38*, 300-307.
(22) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301-332.
(23) Warshel, A.; Russell, S. T. *Quart. Rev. Biophys.* **1984**, *17*, 283-422.

0002-7863/91/1513-9436$02.50/0

Table I. Data Used in the Calculations[a]

		crystallographic data			NMR assignments				
								no. of	
no.	protein	PDB code	resolution	*R*-factor (%)	pH	temp (°C)	Cα shifts	side-chain shifts	rings
1	hen egg white lysozyme	2LZT	1.97	12.4	3.8	35	110	253	1H 3F 3Y 6W
2	reduced cytochrome *c*	5CYT	1.5	15.9	5.75	40	72	225	2H 3F 5Y 2W hem
3	dihydrofolate reductase	3DFR	1.7	15.2	6.5	35		79	7H 8F 5Y 4W[b]
4	ribonuclease T-1	2RNT	1.8	14.9	5.5	40	80	181	3H 4F 9Y 1W
5	bovine calbindin D-9k	3ICB	2.3	17.8	6.0	27	67	257	5F 1Y
6	oxidized thioredoxin		1.7	16.5	5.7	35	97	351	1H 4F 2Y 2W
7	bovine pancreatic trypsin inh	5PTI	1.0	20	4.6	36	50	164	4F 4Y
8		6PTI	1.7	16					
9	ribonuclease A	1RN3	1.45	24	3.2	30	116	204	4H 3F
10					4.0	35	119	235	6Y
11	barley serine proteinase inh	2CI2	2.0	19.8	4.2	42	59	180	1F 1Y 1W
12	tendamistat	1HOE	2.0	19.9	3.2	50	65	190	2H 6Y 1W
13	reduced cytochrome b_5	2B5C	2.0		7.2	40	74	198	5H 3F 4Y 1W hem
14	bacteriophage T4 lysozyme	3LZM	1.7	15.7	5.6	20	146	196	1H 5F 6Y 3W
15	turkey ovomucoid third domain	3SGB	1.8	12.5	4.0	25	45	131	1H 2F 3Y
16	human ubiquitin	1UBQ	1.8	17.6	3.6	27	68	215	1H 2F
17					5.8	30	69	252	1Y
18	human lysozyme	1LZ1	1.5	17.7	3.8	35	117	283	1H 2F 6Y 5W
19	myoglobin CO	1MBC	1.5	17.1	5.6	36	77	190	12H 6F 3Y 2W hem
20	reduced cytochrome c_{551}	451C	1.6	18.7	6.8	25	72	177	1H 2F 1Y 2W hem

[a] References to the crystal structures and NMR assignments, respectively, for each protein: (1) 45, 46; (2) 47, 48; (3) 49, 50, 51; (4) 52, 53; (5) 54, 55; (6) 56, 57; (7) 30, 58; (8) 31; (9) 59, 60; (10) 61; (11) 62, 63; (12) 64, 65; (13) 66, 67; (14) 68, 69; (15) 70, 71; (16) 72, 28; (17) 29; (18) 73, 74; (19) 75, 76; (20) 77, 78. [b] Ring currents for the ligands were modeled as additional benzene and indole rings.

undoubtedly arises from electrostatic effects as well, so that a careful analysis of many factors is needed. We show below that inclusion of electrostatic effects from peptide groups provides a modest improvement in the statistical reliability of chemical shift predictions, but that more work in this area seems warranted.

Methods

A. Database of Observed Shifts. We chose 17 proteins for which extensive site-specific resonances are available and which also have crystal structures available at fairly high resolution. The basic data and references are given in Table I. There will certainly be differences between solution and crystal structures, but, for a large enough database, one may hope that errors in the computed shifts arising from these differences will contain both upfield and downfield contributions, so that the mean correlations seen will be valid ones. One particular concern in this regard involves the charge state of side chains, since many protein NMR studies have been carried out under conditions acidic enough to protonate many histidines and potentially also aspartate and glutamate side chains. Protons attached to the C2 and C4 positions of histidine, to the Cβ carbon of aspartate, and the Cγ carbon of glutamate are known to be quite sensitive to the state of protonation of the residue.[2] We have therefore eliminated these protons from our statistical sample, since the protonation states of side chains in proteins are often not known. However, we have found that the overall results would not be significantly changed by including them, with the assumption that all of the residues are charged.

It is becoming increasing common to have stereospecific assignments available for prochiral protons, particularly at the Cβ positions. Nevertheless, such information is still relatively rare for the particular set of proteins chosen for this study, and we have elected to consider only the average shift for prochiral protons, for both observed and calculated values. The glycine protons attached to Cα are an additional special case, since here there are significant differences in shifts even in short linear peptides, probably arising from contributions from neighboring amide groups. Since we do not know in detail the conformational preferences of "random" peptides, we have omitted the glycine protons from our statistical sample. A further study of these shifts is planned.

Finally, for most phenylalanine and tyrosine residues, only a single, averaged shift is observed for protons attached to the Cδ and Cε positions. We have averaged the calculated shifts for the two protons involved and discarded residues where the ring-flip exchange was not fast on the NMR time scale.

Our final "database" on which we carried out calculations consisted of 5678 protons in 17 proteins, as outlined in Table I. For each observed shift we have subtracted the reference shift seen in the GGXA peptides studied by Bundi and Wüthrich.[2] The statistical correlations presented below are all in terms of this difference, i.e., to the conformation-dependent shift. In cases where chemical shifts were given relative to TSP (sodium 2,2,3,3-tetradeutero-3-(trimethylsilyl)propionate) the chemical shifts were corrected for the small pH dependence of the reference TSP by[24]

$$\delta[\text{ppm}] = \delta_{\text{obs}} - 0.019(1 + 10^{(5.0-\text{pH})})^{-1} \quad (2)$$

In many cases, the correlation between the X-ray and solution structures should be straightforward, and there is often good evidence that the solution structure must be fairly close to that in the crystalline environment. Several of our examples deserve special comment.

(1) With cytochrome *c*, we have chosen to compare shifts measured for the horse heart protein with a structure from tuna.

(24) DeMarco, A. *J. Magn. Reson.* **1977**, *26*, 527–528.

The sequence homology between these two proteins is quite high, with five amino acid changes. The only aromatic substitution (which would affect ring-current contributions) involves Trp 33 in the tuna protein, which becomes histidine in the horse protein. We modeled the histidine side chain with the same χ_2 dihedral angle as found for tryptophan in the crystal structure. Since this study was carried out, a new, more highly refined structure for the oxidized form of the horse protein has been reported;[25] we will consider the implications of this structure, plus other new heme protein data (such as for a yeast cytochrome c[26,27]), in a subsequent paper.

(2) Two chemical shift data sets were used for ubiquitin and for ribonuclease A. These data were collected at different pH values. Although, we did not use δ and ϵ proton chemical shifts for least-squares fitting, the assignments for His 68 given by DiStefano and Wand[28] may be mistaken, based on comparisons to the assignments of Weber et al.[29] and to the calculated values.

(3) In view of the historical importance of the pancreatic trypsin inhibitor to NMR studies, we have carried out calculations on each of two crystal forms.[30,31] Alternate crystal structures are also available for lysozyme and myloglobin, and we report comparisons with these structures below.

(4) For cytochrome b_5, we used the oxidized crystal structure and shifts from the reduced protein; the two conformers are expected to be nearly identical.

(5) For human lysozyme, the crystal structure has a valine at position 130, whereas the NMR study was carried out on a variant with leucine at that position.

B. Ring Current Calculations. The basic ideas of ring-current calculations may be found in textbooks[13] and reviews.[14] The general form of ring-current contributions is

$$\delta_{rc} = iBG(\mathbf{r}) \tag{3}$$

where r is the vector from the observed proton to the aromatic ring, $G(\mathbf{r})$ is a geometric factor, and i and B are constants. It is conventional to incorporate into B those constants that would yield the expected contribution from a benzene ring and to use i (the "ring-current intensity" factor) to represent the ratio of the intensity expected for the ring in question relative to that of a benzene ring. For example, in the Haigh–Mallion theory, which is one of the simplest to implement, the geometric factor is

$$G(\mathbf{r}) = \sum_{ij} S_{ij} \left\{ \frac{1}{r_i^3} + \frac{1}{r_j^3} \right\} \tag{4}$$

Here r_i and r_j are the distances from ring atoms i and j to the proton, and s_{ij} is the area of the triangle formed by atoms i and j and the proton projected onto the plane of the aromatic ring. The sum is over the bonds in the ring. In this case, $B = 5.455 \times 10^{-6}$ Å. Other estimates of the ring-current contribution can be formulated in a similar fashion.

We have used a modified version of codes originally written by Keith Cross to compute both the Johnson–Bovey and Haigh–Mallion estimates of these contributions; as has long been recognized, there is very little difference between these theories,[11] although the approach used by Haigh and Mallion is simpler to work with and to differentiate. Of particular concern to us are the "intensity factors" that relate the magnitudes of ring-current contributions from amino acid side chains to that of benzene.

Table II. Fitted parameters[a]

parameter	original[b]	lsq[c]	nonlinear[d]	uncert[e]
His	0.53	0.90	0.88	0.09
Phe	1.00	1.00	1.01	0.06
Tyr	0.94	0.84	0.86	0.05
Trp(6)	1.04	1.02	1.00	0.08
Trp(5)	0.56	1.04	1.05	0.14
heme pyrrole	0.55	1.08	1.21	0.25
heme macro	1.99	1.68	1.59	0.03
peptide $\Delta\chi$	−5.1	−7.9	−7.3	0.4
A(elec.)	−2.0	−1.2	−1.1	0.1
constant, Cα		−0.75	−0.72	0.03
constant, side chain		−0.04	−0.04	0.01
constant, amides		−0.55		

[a] Ring-current intensity parameters are dimensionless; $\Delta\chi$ is in units of 10^{-6} cm^3/mol; A(elec.) is in units of 10^{-12} esu^{-1}; and the three constants are in ppm. [b] Ring-current intensity parameters taken from refs 32 and 42. The value quoted for $\Delta\chi$ is that for formamide in the gas phase, from ref 36. The "original" value for the constant A is that suggested in ref 21. [c] The linear least-squares estimate. [d] A nonlinear estimate of the parameter, using eq 9. [e] The "jackknife" estimate of the uncertainty in the parameter estimate; see the text.

Although it is possible to generate theoretical arguments for these parameters,[32] they are probably best viewed as empirical constants to be determined by observed data. Table II gives values that have been used in the past, determined largely from semiempirical quantum calculations and from analyses of BPTI and lysozyme. Below, we discuss the recalibration of these parameters based on the current, much larger database.

C. Peptide Group Contributions. It has been recognized for some time that the magnetic anisotropy of the peptide group is likely to contribute significantly to chemical shifts in proteins, but it has been difficult to develop convincing models for this phenomenon. When the observed proton and the "source" of the magnetic anisotropy are far apart, McConnell[33] has shown that the contribution to the local shift tensor depends upon the magnetic anisotropy of the distant group:

$$\delta_m = (3L_0R^3)^{-1} \sum_{i=x,y,z} \chi_{ii} (3\cos^2\theta_i - 1) \tag{5}$$

Here L_0 is the Avogadro constant, R, is the distance from the proton to the distant group, χ_{ii} is a component of the magnetic susceptibility tensor, and θ_i is the angle between the i-axis and the radius vector $\mathbf{R}$. Since there is no direct method to measure the magnetic anisotropy of a peptide group within a protein, all estimates of these effects are to some extent empirical. Flygare's group has measured susceptibilities for a large number of molecules in the gas phase[34] and have developed procedures to analyze these effects in terms of localized contributions.[35] Data for formamide[36] suggest that the peptide group is nearly axially symmetric about the out-of-plane axis, with an anisotropy $\Delta\chi$ of $-5.1 \pm 0.6 \times 10^{-6}$ erg/(G^2-mol). This value is in good agreement with an empirical theoretical approach developed by Pauling, which yields a value of -5.4×10^{-6} erg/(G^2-mol).[18] In the case of axial symmetry, eq 5 becomes

$$\delta_m = (3L_0R^3)^{-1}\Delta\chi(3\cos^2\theta - 1) \tag{6}$$

where θ is the angle between the vector connecting the amide group to the proton in question and the vector normal to the amide plane.

Even assuming, however, that the measured value for formamide can be used to represent peptide groups inside proteins, two essentially empirical questions remain: what is the appropriate origin for the application of eq 6, and how great is the contribution

(25) Bushnell, G. W.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 585–595.
(26) Gao, Y.; Boyd, J.; Williams, R. J. P.; Peklak, G. J. *Biochemistry* **1990**, *29*, 6994–7003.
(27) Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 527–555.
(28) DiStefano, D. L.; Wand, A. J. *Biochemistry* **1987**, *26*, 7272.
(29) Weber, P. L.; Brown, S. C.; Mueller, L. *Biochemistry* **1987**, *26*, 7282–7290.
(30) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J. Mol. Biol.* **1984**, *180*, 301.
(31) Wlodawer, A.; Nachman, J.; Gilliland, G. L.; Gallagher, W.; Woodward, C. *J. Mol. Biol.* **1987**, *198*, 469.
(32) Giessner-Prettre, C.; Pullman, B. *C. R. Hebd. Seances Acad. Sci. Ser. D* **1969**, *268*, 1115–1117.
(33) McConnell, H. M. *J. Chem. Phys.* **1957**, *27*, 226–229.
(34) Flygare, W. H. *Chem. Rev.* **1974**, *74*, 653–687.
(35) Schmalz, T. G.; Norris, C. L.; Flygare, W. H. *J. Am. Chem. Soc.* **1973**, *95*, 7961–7967.
(36) Tigelaar, H. L.; Flygare, W. H. *J. Am. Chem. Soc.* **1972**, *94*, 343–346.

Table III. Statistics of the Fits[a]

data set		parameters used							
		ring-current (original)		ring-current (new)		ring-current + anis		all	
	no. of protons	r	rms	r	rms	r	rms	r	rms
all	5678	0.773	0.314	0.778	0.310	0.870	0.242	0.880	0.233
Hα	1553	0.547	0.468	0.557	0.462	0.832	0.315	0.849	0.298
side chain	4125	0.871	0.230	0.877	0.225	0.894	0.208	0.899	0.203
methyl groups	849	0.940	0.176					0.961	0.136
heme proteins	1085	0.883	0.343	0.886	0.336	0.915	0.288	0.918	0.282
non-heme proteins	4593	0.679	0.306	0.685	0.303	0.833	0.230	0.849	0.220
amides[b]	1627	0.329	0.704					0.575	0.621

[a] The correlation coefficient r between the observed and calculated structural shifts is defined by

$$r = \frac{n\Sigma(xy) - \Sigma(x)\Sigma(y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2] \times [n\Sigma y^2 - (\Sigma y)^2]}}$$

rms gives the root-mean-square error. [b] For 1065 amide protons involved in internal hydrogen bonds, the final parameters yield $r = 0.721$, rms error = 0.506.

of these terms to the "random-coil" values that are serving as our reference shifts. We investigated several possible locations for the origin, placing it at various points along the bisector of the NCO angle at distances between 0 and 1.5 Å from the carbon atom. We found only minor variations in the quality of the resulting fits; our final results correspond to an origin that is 0.7 Å from the carbon, which is roughly in the center of the amide group.

Unlike the ring-current contribution, which cannot contribute to shifts in peptides that have no aromatic side chains other than the one being examined, the peptide anisotropy contribution will contribute to the "random-coil" reference shift as well as to the protein shift. At the present time, we see no good way of calculating this amide contribution, since we do not know in detail what distribution of conformers is actually populated in the short linear, "random" peptides. Test calculations on short peptides showed that the amide contribution was largest for Cα protons and much smaller for all side-chain protons. If it is correct to assume that "random" peptides populate similar distributions of backbone dihedral angles (independent of sequence), then the contribution to the reference shift would be the same for all Cα protons. We have made this assumption and treated the constant as an additional fitting parameter. A similar argument can be made for side-chain protons as a group or for amide protons as a group.

Our final model for amide contributions, then, has three adjustable parameters: the anisotropy $\Delta\chi$, the average reference shift for Cα protons, and the average reference shift for side-chain protons. If the fits are to make good physical sense, the value of $\Delta\chi$ should be fairly close to that found for formamide, and the side-chain constant correction should be small. We indeed find this to be the case, as shown below.

Some earlier analyses of peptide group contributions have started from analyses of carbonyl contributions estimated by analyzing NMR shifts in ketones.[19,20] These models are similar to that described above, except that axial symmetry in the plane of the peptide group is not imposed. Since the magnetic anisotropy parameters for ketones are not expected to match those for amides, these theories essentially have two adjustable anisotropy parameters. In the Zürcher or ApSimon approaches, the anisotropy contribution of C=O groups is

$$\delta_m = (3L_0R^3)^{-1}[\Delta\chi_1(3\cos^2\phi_1 - 1) + \Delta\chi_2(3\cos^2\phi_2 - 1)] \quad (7)$$

In the most successful parameterization, **P**, the origin of the coordinate system, was put at 0.6 Å apart from C along the C=O bond direction; ϕ_1 is the angle between PO and PH, ϕ_2 is the angle between PH and the axis that lies in the plane of the peptide group, and $\Delta\chi_1$ and $\Delta\chi_2$ are magnetic susceptibility differences. We carried out fits with both eqs 6 and 7. Almost no improvement is found for the latter theory, indicating that the present data are not able to distinguish between axial and nonaxial anisotropies for the amide group in proteins. Accordingly, we have based most of our analyses on eq 6, which has one less adjustable parameter.

D. Electrostatic Contributions. A significant contribution to chemical shifts can also arise from distant polar groups, which can polarize the C–H bond and thereby increase or decrease the local shielding by electrons. The most significant term is expected to be proportional to the projection of the local electric field onto the C–H bond vector:

$$\delta_{el} = A\mathrm{E(C\text{–}H)} \quad (8)$$

Buckingham[21] suggested that an appropriate value for A would be -2×10^{-12} esu^{-1}, but this magnitude clearly depends upon the way in which local electric fields are estimated. One simple procedure would involve the use of Coulomb's law along with partial charges derived from a molecular mechanics force field. We investigated this approach and found, as have others,[37] that there seems to be little or no correlation between the fields estimated in this way and observed shifts. A principal problem is likely to be that charged side chains are largely exposed to solvent, and their Coulomb interactions are greatly reduced by the solvent dielectric. A more satisfying theory would include the dielectric effects of water molecules surrounding the protein. We are considering models along these lines, but for the moment have limited our consideration of electrostatic effects to backbone charges. We used the charge model from the CHARMM version 19 parameters, which has −0.35, 0.25, 0.10, 0.55, and −0.55 for N, H, Cα, C, and O atoms, respectively, and −0.20 for the charge of a proline nitrogen. We omitted electrostatic contributions from the residue of the proton in question, assuming that these effects would be incorporated into the reference peptide shifts. As shown below, this approach yields an electrostatic contribution that provides a modest but systematic improvement in the fits to experimental data.

E. Parameter Fitting. Our final model has 11 linear parameters: seven ring-current intensity factors, one overall parameter each for the peptide group anisotropy and for the backbone electrostatic contributions, and two constants, one for Cα protons and one for side-chain protons. We used three statistical methods to estimate the unknown parameters and their uncertainties. The first was a standard linear least-squares algorithm[38] which provides the maximum likelihood estimate, assuming that the errors are normally distributed. Second, as a test of the sensitivity of these results to individual large errors, we also used a nonlinear optimization program to minimize

$$\rho(z) = \sum_{\text{shifts}} \log(1 + 1/2z^2) \quad (9)$$

where $z = (\sigma_{obs} - \sigma_{calc})/c$ with $c = 0.23$ ppm. This has the effect

(37) Hoch, J., personal communication.
(38) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986, Chapter 14.

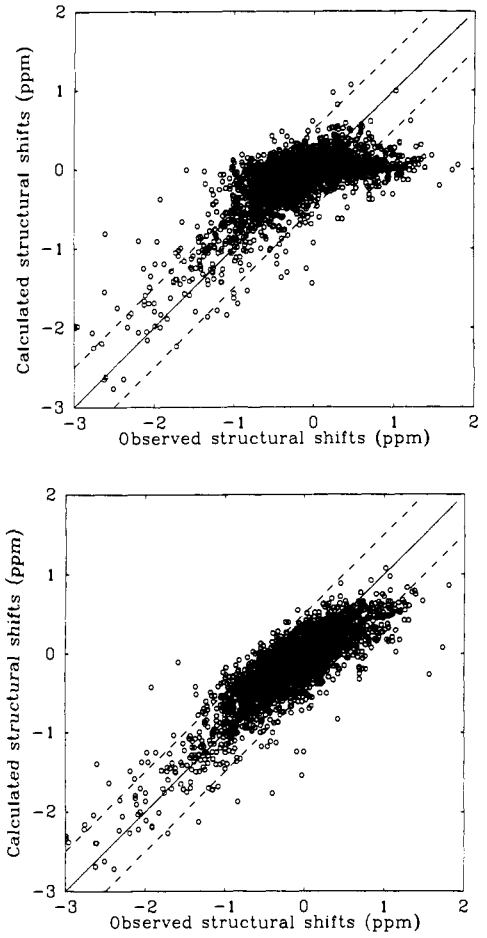

Figure 1. Correlation between observed and calculated structural shifts for all (5678) protons bonded to carbon: (a) calculation using ring currents alone with the "original" parameters (see Table III), $r = 0.773$, rms = 0.314 and (b) calculation with the new fitted parameter set, $r = 0.880$, rms = 0.233 ppm.

of reducing the importance of shifts whose errors are much larger than the rms error over the data set; such outliers might arise, for example, in cases where the solution structure and the crystal structure differ significantly.

Finally, to gain another view on the predictive ability of the correlations and on the uncertainty in the parameters, we performed least-squares optimizations in which each protein in turn was deleted from the input set. The parameters determined in this way can then be used to predict the shifts in the deleted protein, giving an estimate of how well we should expect a new protein to be fit. These results are collected in Table IV. This "jackknife" procedure can also be used to develop an estimate of the uncertainties in the final parameters.[39] By eliminating each protein in turn, we obtain 17 least-squares estimates of the parameters based on data for 16 proteins. One then defines *pseudovalues* for each parameter as

$$y^*_j = k y_{all} - (k-1) y_j, \qquad j = 1,2,...k \tag{10}$$

where y_{all} is the parameter estimate when all of the data is considered, and y_j the least-squares estimate when the jth protein is omitted; for the example here, $k = 17$. Then the jackknife estimate for the parameter is the mean of the y^*_j values, and the estimate of its uncertainty is determined by standard formulas for the uncertainty of a mean.[39] These uncertainties are listed in the final column of Table II. A listing of all of the calculated and observed shifts is included as supplementary material; copies of the computer programs used to calculate the shifts are available from the authors.

(39) Mosteller, F.; Tukey, J. W. *Data Analysis and Regression. A Second Course in Statistics*; Addison-Wesley: Reading, MA, 1977.

Table IV. Predictions from the "Jackknife" Procedure

		predicted	
no.	protein	r	rms
1	hen egg white lysozyme	0.898	0.219
2	reduced cytochrome c	0.940	0.293
3	dihydrofolate reductase	0.919	0.203
4	ribonuclease T-1	0.913	0.230
5	bovine calbindin D-9k	0.800	0.205
6	oxidized thioredoxin	0.776	0.208
7	bovine pancreatic trypsin inh	0.884	0.213
8		0.882	0.212
9	ribonuclease A	0.803	0.269
10		0.812	0.255
11	barley serine proteinase inh	0.800	0.211
12	tendamistat	0.840	0.232
13	reduced cytochrome B_5	0.827	0.348
14	bacteriophage T4 lysozyme	0.826	0.225
15	turkey ovomucoid third domain	0.812	0.223
16	human ubiquitin	0.807	0.194
17		0.773	0.205
18	human lysozyme	0.869	0.224
19	myoglobin CO	0.872	0.264
20	reduced cytochrome c_{551}	0.940	0.301

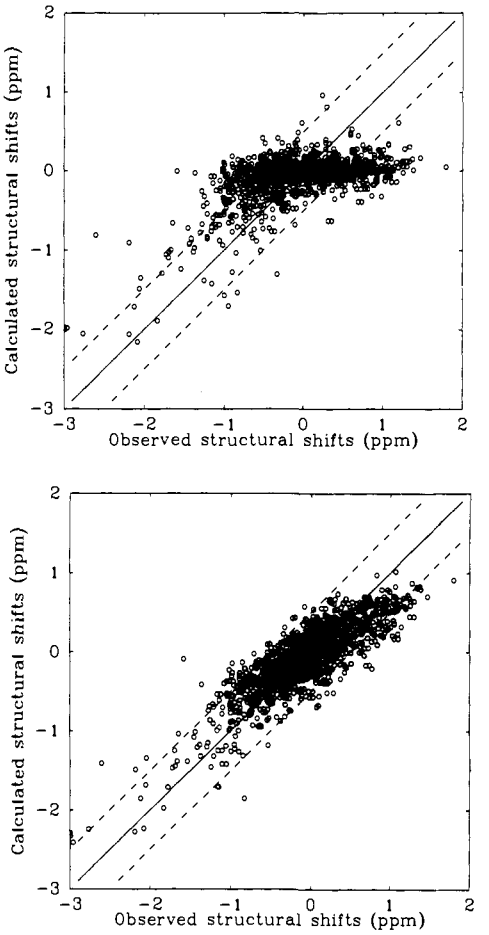

Figure 2. Correlation between observed and calculated structural shifts for 1553 Cα protons: (a) calculation using the original ring currents alone, $r = 0.547$, rms = 0.468 ppm and (b) calculation with the new fitted parameter set, $r = 0.849$, rms = 0.298 ppm.

Results

The basic results of our investigation are collected in Table II, which shows the parameters determined from each of the statistical procedures derived above, along with values for ring-current intensities that have been used in earlier studies. Table III gives statistics on our final fits, and Figures 1 and 2 show predicted vs observed shifts for a ring-current only theory and for one that includes the effects of peptide group magnetic anisotropy and electrostatic interactions. As we pointed out above, when ring

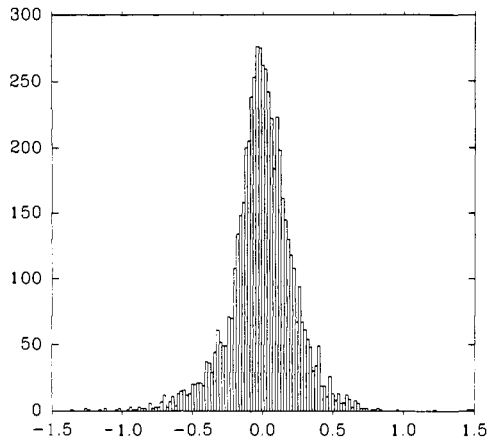

Figure 3. Distribution of errors between calculated and observed structural shifts for all (5678) protons bonded to carbon.

currents alone are used to predict structural chemical shifts, there are a significant number of protons whose predicted shifts are near zero, yet whose observed shifts are significantly different from zero. This is evident in the horizontal ellipse of overlapping points seen in Figures 1a and 2a and is especially noticeable for the protons at the Cα position. When an estimate of the peptide group contribution is added to the ring currents (Figures 1b and 2b), this ellipse is significantly tilted toward a line with a slope of unity. The statistical measures collected in Table III reinforce the visual conclusion from the figures that the peptide group contributions offer a substantial improvement in the quality of fits to chemical shift variations in proteins. In particular, it should be noted that the new ring-current intensities, by themselves, would lead to only a marginal improvement in the fits between observed and calculated values; most of the final improvement indeed arises from the inclusion of peptide group effects.

The distribution of errors in our final results is shown in Figure 3. The results are fairly sharply peaked about a mean of 0.0004 with a standard deviation of $\sigma = 0.23$ ppm. The percentages of errors exceeding σ, 2σ, 3σ, and 4σ are 24.1, 5.8, 1.1, and 0.4, respectively. For a normal distribution, these values would be 31.7, 4.6, 0.3, and 0.006. The distribution of errors is thus approximately Gaussian near the center, but there are significantly more large errors than would be expected from a normal distribution. We have looked at the nature of the protons with the largest deviations. Many are from protons at the Cα position (as is evident from the statistics in Table III), a number of others are in heme proteins (where the ring-current shifts can be much larger than in non-heme proteins), and others cluster in lysine and arginine side chains, where one might commonly expect the solution structure to differ from that in the crystal. If we eliminate these protons from our statistics, then the fraction of errors greater than σ, 2σ, 3σ, and 4σ becomes 16.3, 2.0, 0.25, and 0.03, respectively; the rms error in this subset (2796 protons) is 0.18 ppm. Heme proteins form a special subset where ring-current effects can be very large, and a number of attempts have been made to interpret them;[40-42] we will consider these shifts in detail elsewhere. The likely origins of other residual errors are considered in the Discussion section, below. It is also worth noting that the rms error for the methyl protons, 0.14 ppm (0.13 ppm for 663 methyls in non-heme proteins), is significantly lower than for all side-chain protons, although the origins of this behavior are not clear.

Ring-current contributions to shifts in both small molecules and in proteins have been considered for many years.[14] Although several functional forms have been put forward for the geometric dependence of this effect, the differences between them are not great, except at very short distances.[11] We have chosen to work primarily with the Haigh–Mallion formula, since it is the easiest to differentiate. Applications to proteins received a boost about ten years ago, when comparisons became possible with a significant number of shifts in lysozyme and BPTI.[10,11] Improvements in our understanding of ring-current effects is now possible not only because more proteins have been assigned but also because the quality of many crystal structures has also improved in the last decade, both from higher resolution of data and from advances in crystallographic refinement techniques. Table II compares ring-current parameters from the current fits with those used in most earlier studies. The most common parameters were originally derived from semiempirical π-electron calculations;[32] modified parameters for tryptophan were later suggested by Perkins[11] based on analyses of observed shifts in hen egg white lysozyme.

The values given in Table II are based on a much larger data base of protons than was available to the earlier studies and also on more highly refined crystal structures. The principal change from the earlier values is that five-membered aromatic rings (in histidine, tryptophan and the heme group) appear to have intensity factors more nearly equal to those of six-membered rings, rather than the smaller values suggested from the semiempirical calculations. As indicated in Table III, however, the overall statistics are a relatively flat function of the intensity parameters. Our best-fit parameters yield results only slightly better than those used in the earlier studies: for a ring-current only theory, the rms error in the prediction changes only from 0.314 to 0.310 ppm, and the correlation coefficient increases from 0.773 to 0.778.

One method to estimate the uncertainties in the fitted parameters is through a "jackknife" procedure, in which each protein in turn is eliminated from the input data set and the parameters refit. In every case, the parameter estimate from this procedure is identical to that of the full least-squares procedure and to the number of significant figures quoted in the table, and the predicted standard deviation gives a very rough idea of the level of precision we expect. We emphasize that the observed distribution of errors is not normal and that the concept of a "standard error" in the parameter is correspondingly not well-defined. For this reason, we consider the uncertainty parameters listed in Table II as qualitative guides only; in particular, the very low uncertainty estimated for the heme macrocycle uncertainty may be an artefact arising from the fact that there are only four heme proteins in our sample.

This procedure also allows an estimate of how accurate a prediction of a new protein is expected to be. For each of the 17 proteins in our set, we computed the predicted shifts based on parameters obtained from the other 16 proteins; these results are given in Table IV. As might be expected from size of our data set, the results are not largely different from those in which all of the proteins were included in the input data: the mean correlation coefficient in Table IV is 0.850 and the mean rms error is 0.24 ppm.

Table III also shows results of fits in which fewer adjustable parameters were used than in our final fits. As was mentioned above, if just ring-current formulas are used (with no peptide group contribution), only small improvements in the overall fits can be obtained. Adding in the anisotropy contribution from the peptide group leads to a significant improvement, especially for protons at the Cα position, where the linear correlation between calculated and observed structural shifts increases from 0.557 to 0.832. Adding in the computed electrostatic contributions from the backbone contributes an additional small improvement.

The results discussed above all dealt with protons bonded to carbon. Figure 4 shows results for amide protons, using the parameters of Table II. Figure 4a shows what has been known for some time, that calculated ring-current contributions explain very little of the observed variability in amide shifts. Addition of peptide group effects (excluding the peptide group to which the amide proton belongs) appears to systematically improve the correlations (Figure 4b), but the scatter is still much worse than we found for protons bonded to carbon. It is likely that amide shifts are influenced in an important way by hydrogen-bonding contributions, both to protein carbonyl groups and to solvent. Empirical attempts to model these effects have had some suc-

(40) Abraham, R. J. *Mol. Phys.* **1961**, *4*, 145.
(41) Abraham, R. J.; Fell, S. C. M.; Smith, K. M. *Org. Magn. Reson.* **1977**, *9*, 367–373.
(42) Cross, K. J.; Wright, P. E. *J. Magn. Reson.* **1985**, *64*, 220–231.

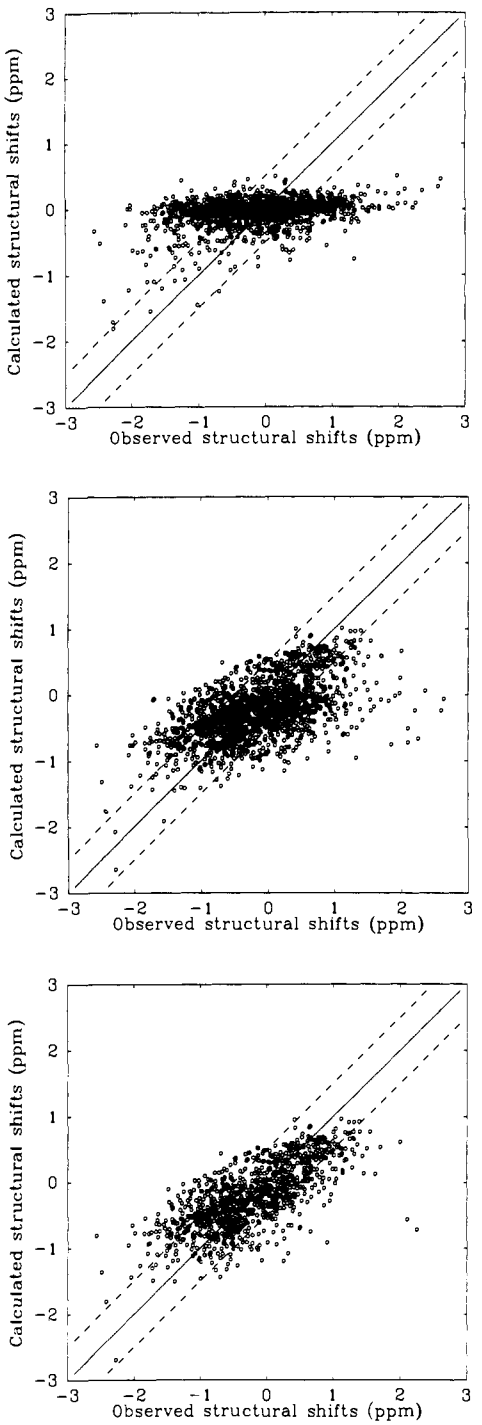

Figure 4. Comparison of observed and calculated structural shifts for amide protons (a) using original ring currents alone, $r = 0.329$, rms = 0.704 ppm, $n = 1627$; (b) and (c) using the parameters fitted for the protons bonded to carbons and a constant of −0.550; (b) all amide protons, $r = 0.575$, rms = 0.621 ppm, $n = 1627$; (c) amide protons involved in internal hydrogen bonds, $r = 0.721$, rms = 0.506 ppm, $n = 1065$.

cess,[12,17,43] but no theory that works as well as the one presented here for protons bonded to carbon has been found. There does appear to be some statistical difference between amides participating in internal hydrogen bonds and those exposed to solvent: Figure 4c shows an improved fit when only amides involved in internal hydrogen bonds are considered, but the scatter is still large. We are in the process of considering more realistic electrostatic (and solvation) models than the simple Coulomb theory used here. For the moment it would seem that shifts of protons bonded to nitrogen are not nearly so well understood as those bonded to carbon.

(43) Kuntz, I. D.; Kosen, P. A.; Craig, E. C. *J. Am. Chem. Soc.* **1991**, *113*, 1406–1408.

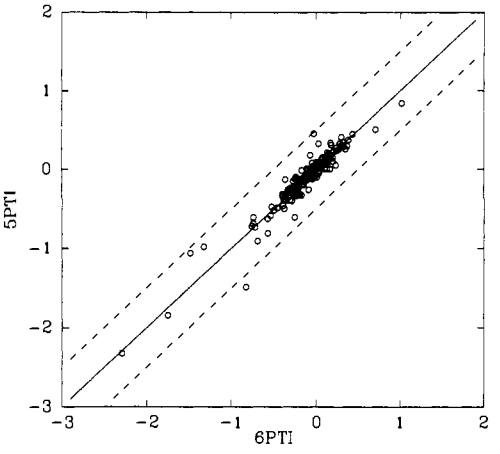

Figure 5. Comparison of calculated structural shifts for pancreatic trypsin inhibitor structures 6PTI and 5PTI: $r = 0.947$, rms = 0.12 ppm.

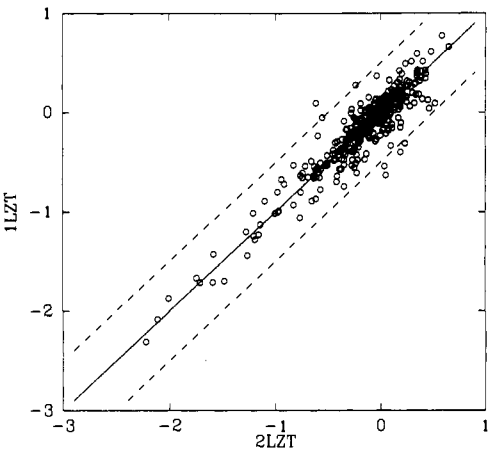

Figure 6. Comparison of calculated structural shifts for lysozyme structures 2LZT and 1LZT: $r = 0.942$, rms = 0.14 ppm.

Discussion

The residual errors in our fits are likely to arise from three sources. First, the "reference" or random-coil peptide value is itself not uniquely defined but probably has some sequence dependence; we are currently examining shifts of a large number of peptides to study this point. Further, as is evident from Table I, the protein measurements were made under a variety of pH and temperatures, and some variation in the structural shifts will arise from these effects. Second, the empirical relations we use to relate structure to chemical shift are certainly only approximate ones and are certainly not capturing electrostatic and solvent influences in a realistic fashion. Finally, there are inevitable differences between the mean solution structure and that seen in the crystal. One measure of this last potential variation can be gained from comparisons of results from different crystal structures. Figures 5–7 illustrate examples of this sort of comparison for BPTI, lysozyme, and myoglobin. In each case, one can compare two relatively high-resolution crystal structures. These scatter plots indicate that the average deviation in prediction shifts between structures is 0.12–0.16 ppm, although individual resonances can have much larger deviations. Since it is likely that the difference between the crystalline and solution environment is at least as large as that between two crystalline environments, these numbers support the notion that a significant portion of our residual errors arise from our use of crystal structures as input to the calculation.

There is always a danger in empirical fits that the mathematical description does not correspond well to the actual physical processes being modeled. For this reason, we have avoided using

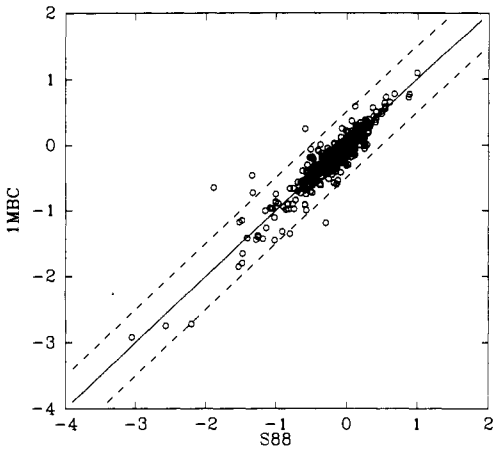

Figure 7. Comparison of calculated structural shifts for myoglobin structures 1MBC and a neutron diffraction refinement ("S88", Schoenborn, B., personal communication; see, also Cheng, X.; Schoenborn, B. P. *J. Mol. Biol.* **1991**, *220*, 381-399): r = 0.927, rms = 0.16 ppm.

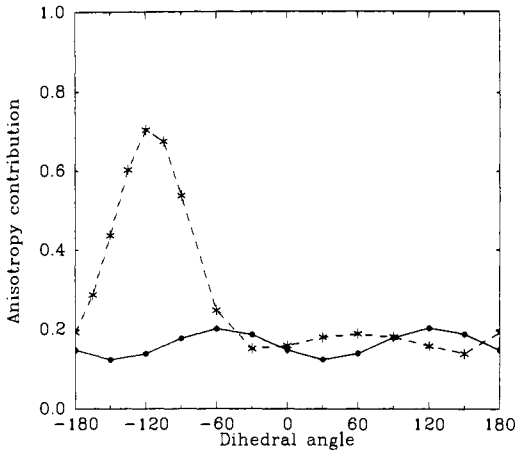

Figure 8. Calculated anisotropy contributions to the Cα proton structural shift when varying the angle ϕ (dashed line) or the angle ψ (solid line).

simple trigonometric functions of torsion angles as basis functions, since it would be difficult to associate the resulting parameters with physical interactions. It is encouraging that the final adjusted parameters (which could have taken on any values whatsoever) in fact converge to physically plausible results. The basic intensities of ring-current effects are in rough accord with those predicted from semicmpirical quantum calculations, and the effective magnetic anisotropy of the peptide group is within 50% of the value measured in the gas phase for formamide. While it is possible, and indeed likely, that other types of interactions are affecting chemical shifts, the contributions from aromatic rings and the peptide group are likely to end up being roughly similar to those predicted here.

It is clear from Table III that the addition of the effects of peptide magnetic anisotropy has a big effect on the predictions of shifts at the Cα position. Since there is a close connection between this shift and the local secondary structure,[3-7] and since the geometric relation of the neighboring peptides depends just on the backbone ϕ and ψ angles (for fixed bond lengths and angles), we computed the peptide anisotropy contributions for a dipeptide model as a function of the backbone angles. These are shown in Figure 8. The ψ dependence is very weak, but there is a strong variation predicted for ϕ in the region −180 to −60°. Most ϕ angles in proteins are in this range, and one should thus expect a strong dependence of the Cα proton shift on this backbone angle. Since helices will have mean ϕ values near −60°, and sheets mean values closer to −120°, a portion of the difference between helix and sheet probably arises from the neighboring peptide group; a further study of Cα proton shifts in regular secondary structures is in progress.

The current fits also have implications for determination of solution structures of proteins by NMR methods. First, since the rms error for a large number of crystal structures is less than 0.25 ppm, it may be hoped that NMR-determined solution structures might achieve similar accuracy. We will report elsewhere results for plastocyanin which indicate that even the most accurate structures obtained using "conventional" distance and angle restraints are not as good as the crystal structures considered here in predicting proton chemical shifts. Second, a reasonable model (such as the one proposed here) may be useful in refining solution structures, even in the presence of residual uncertainties. In this respect, the strong dependence of shift on structure is an advantage: relatively large errors in the formulas or in the computed shifts may translate into only small errors in the predicted atomic positions. In this respect, chemical shifts are somewhat like NOESY intensities in that useful structural information may be available even in the absence of a truly quantitative understanding of the spectral phenomenon.

We have discussed earlier the use of ring-current formulas in making NMR-based penalty functions for structure refinement.[16] The crucial ingredient is the derivative of the shift with respect to atomic coordinates. We have already incorporated such effects for ring currents into the AMBER package (version 4.0)[44] and are

(44) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 4.0; University of California: San Francisco, 1991.
(45) Kurachi, K.; Sieker, L. C.; Jensen, L. H. *J. Mol. Biol.* **1976**, *101*, 11-24.
(46) Redfield, C.; Dobson, C. M. *Biochemistry* **1988**, *27*, 122-136.
(47) Takano, T.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *153*, 79-115.
(48) Wand, A. J.; DiStefano, D. L.; Feng, Y.; Roder, H.; Englander, S. W. *Biochemistry* **1989**, *28*, 186-194.
(49) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650-13662.
(50) Hammond, S. J.; Birdsall, B.; Searle, M. S.; Roberts, G. C. K.; Feeney, J. *J. Mol. Biol.* **1986**, *188*, 81.
(51) Birdsall, B.; Arnold, J. R. P.; Jimenez-Barbero, J.; Frenkiel, T. A.; Bauer, C. J.; Tendler, S. B.; Carr, M. D.; Thomas, J. A.; Roberts, G. C. K.; Feeney, J. *Eur. J. Biochem.* **1990**, *191*, 659-668.
(52) Arni, R.; Heinemann, U.; Maslowska, M.; Tokuoka, R.; Saenger, W. *Acta Crystallogr., Sect. B* **1987**, *43*, 549.
(53) Hoffmann, E.; Rüterjans, H. *Eur. J. Biochem.* **1988**, *177*, 539-560.
(54) Szebenyi, D. M. E.; Moffat, K. *J. Biol. Chem.* **1986**, *261*, 8761.
(55) Kördel, J.; Forsén, S.; Chazin, W. J. *Biochemistry* **1989**, *28*, 7065.
(56) Katti, S. K.; LeMaster, D. M.; Eklund, H. *J. Mol. Biol.* **1990**, *212*, 167-184.
(57) Dyson, H. J.; Holmgren, A.; Wright, P. E. *Biochemistry* **1989**, *28*, 7074-7087.
(58) Wagner, G.; Braun, W.; Havel, T. F.; Schaumann, T.; Go, N.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 611-639.
(59) Borkakoti, N.; Moss, D. S.; Palmer, R. A. *Acta Crystallogr., Sect. B* **1982**, *38*, 2210.
(60) Robertson, A. D.; Purisima, E. O.; Eastman, M. A.; Scheraga, H. A. *Biochemistry* **1989**, *28*, 5930-5938.
(61) Rico, M.; Bruix, M.; Santoro, J.; Gonzalez, C.; Neira, J. L.; Nieto, J. L.; Herranz, J. *Eur. J. Biochemistry* **1989**, *183*, 623-638.
(62) McPhalen, C. A.; James, M. N. G. *Biochemistry* **1987**, *26*, 261.
(63) Kjaer, M.; Ludvigsen, S.; Sorensen, O. W.; Denys, L. A.; Kindtler, J.; Poulsen, F. M. *Carlsberg Res. Commun.* **1987**, *52*, 327-354.
(64) Pflugrath, J. W.; Wiegand, G.; Huber, R.; Vertesy, L. *J. Mol. Biol.* **1986**, *189*, 383.
(65) Kline, A. D.; Wüthrich, K. *J. Mol. Biol.* **1986**, *192*, 869-890.
(66) Mathews, F. S.; Argos, P.; Levine, M. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *36*, 387.
(67) Guiles, R. D.; Altman, J.; Kuntz, I. D.; Waskell, L. *Biochemistry* **1990**, *29*, 1276-1289.
(68) Matsumura, M.; Wozniak, J. A.; Dao-Pin, S.; Matthews, B. W. To be published.
(69) McIntosh, L. P.; Wand, A. J.; Lowry, D. F.; Redfield, A. G.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 6341-6362.
(70) Read, R. J.; Fujinaga, M.; Sielecki, A. R.; James, M. N. G. *Biochemistry* **1983**, *22*, 4420.
(71) Robertson, A. D.; Westler, W. M.; Markley, J. L. *Biochemistry* **1988**, *27*, 2519-2529.
(72) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
(73) Artymiuk, P. J.; Blake, C. C. F. *J. Mol. Biol.* **1981**, *152*, 737.
(74) Redfield, C.; Dobson, C. M. *Biochemistry* **1990**, *29*, 7201-7214.

currently incorporating the peptide group contributions in the same way. Our limited experience with plastocyanin suggests that relatively small adjustments in a fairly "high-resolution" NMR structure can yield chemical shift estimates whose errors are no larger than those for the crystal structures considered here.

Acknowledgment. This research was supported by NIH Grants GM38794 and HL40453. We thank Walter Chazin, Art Palmer, Jane Dyson, and Peter Wright for helpful discussions. Many of the calculations used a modified version of a program originally written by Keith Cross. We thank Benno Schoenborn for providing us coordinates for the neutron structure of MbCO in advance of publication.

Registry No. Lysozyme, 9001-63-2; dihydrofolate reductase, 9002-03-3; ribonuclease T-1, 9026-12-4; trypsin inhibitor, 9035-81-8; ribonuclease A, 9001-99-4; serine protease inhibitor, 37205-61-1; tendamistat, 86596-25-0.

Supplementary Material Available: Table of observed shifts, along with those calculated in the final parameter set, for all of the observations used in the calculations for the protons attached to a carbon atom (33 pages). Ordering information is given on any current masthead page.

(75) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. *J. Mol. Biol.* **1986**, *192*, 133-154.
(76) Dalvit, C.; Wright, P. E. *J. Mol. Biol.* **1987**, *194*, 313-327.
(77) Matsuura, Y.; Takano, T.; Dickerson, R. E. *J. Mol. Biol.* **1982**, *156*, 389.
(78) Detlefsen, D. J.; Thanabal, V.; Pecoraro, V. L.; Wagner, G. *Biochemistry* **1990**, *29*, 9377-9386.

Double-Quantum Filtering in Magic-Angle-Spinning NMR Spectroscopy: An Approach to Spectral Simplification and Molecular Structure Determination

Robert Tycko* and Gary Dabbagh

Contribution from AT&T Bell Laboratories, Murray Hill, New Jersey 07974.
Received May 1, 1991

Abstract: We show how simple radio-frequency pulse sequences can be used to select resonances from pairs of magnetic dipole-coupled nuclear spins and to suppress resonances from isolated spins in magic-angle-spinning (MAS) NMR experiments, thereby simplifying the spectra and providing information about internuclear distances. Non-zero average dipole–dipole couplings are generated by means of DRAMA sequences (Tycko, R.; Dabbagh, G. *Chem. Phys. Lett.* **1990**, *173*, 461–465). Double-quantum filtering techniques are then used to select the resonances of coupled spin pairs. Two experimental demonstrations of double-quantum filtering in ^{13}C MAS spectra of mixtures of organic compounds are presented, one in which the NMR signal from labeled carbon sites in a doubly ^{13}C-labeled compound ($(CH_3)_2C(OH)SO_3Na$) is selected while natural abundance signals from an unlabeled compound (*N*-acetyl-L-valine) are suppressed, and one in which natural abundance signals from a singly ^{13}C-labeled compound (methionine·HCl) are selected while signals from an unlabeled compound (*N*-acetyl-L-valine) are suppressed. The implications of these experiments, including potential applications to the simplification of MAS spectra of complex molecules, such as biopolymers, and to the determination of the structure of selected regions of complex molecules, are analyzed in detail.

Introduction

Magic angle spinning (MAS) is very widely used to obtain high resolution in nuclear magnetic resonance (NMR) spectra of polycrystalline and noncrystalline solids. High resolution is achieved because MAS has the effect of averaging out the (second-rank) orientation-dependent parts of nuclear spin interactions, principally anisotropic chemical shifts (CSA) and nuclear magnetic dipole–dipole couplings. This paper addresses two important problems with the MAS technique. First, the fact that dipole–dipole couplings are averaged out in a conventional MAS experiment also means that the structural information contained in these couplings is lost. Second, even the line narrowing that results from MAS is very often insufficient to produce a spectrum in which resonances from inequivalent nuclei are resolved, for example in ^{13}C MAS spectra of large molecules such as peptides and small proteins where there are many inequivalent nuclei or in spectra of noncrystalline materials such as amorphous synthetic polymers where the MAS lines are inhomogeneously broadened as a consequence of disorder. In such situations, it is desirable to simplify the spectrum by selecting certain resonances of interest and suppressing others.

Several groups have dealt with the first problem by devising somewhat more complicated MAS techniques that allow the retention of both the high resolution of MAS and the information contained in homonuclear[1-14] and heteronuclear[15-20] dipole–dipole couplings. In particular, we have shown[1] that homonuclear dipole–dipole couplings can be prevented from averaging out during

(1) Tycko, R.; Dabbagh, G. *Chem. Phys. Lett.* **1990**, *173*, 461-465.
(2) Meier, B. H.; Earl, W. L. *J. Am. Chem. Soc.* **1987**, *109*, 7937-7941.
(3) Meier, B. H.; Earl, W. L. *J. Chem. Phys.* **1986**, *85*, 4905-4911.
(4) Andrew, E. R.; Bradbury, A.; Eades, R. G.; Wynn, V. T. *Phys. Lett.* **1963**, *4*, 99-100.
(5) Andrew, E. R.; Clough, S.; Farnell, L. F.; Gledhill, T. D.; Roberts, I. *Phys. Lett.* **1966**, *21*, 505-506.
(6) Raleigh, D. P.; Harbison, G. S.; Neiss, T. G.; Roberts, J. E.; Griffin, R. G. *Chem. Phys. Lett.* **1987**, *138*, 285-290.
(7) Raleigh, D. P.; Levitt, M. H.; Griffin, R. G. *Chem. Phys. Lett.* **1988**, *146*, 71-76.
(8) Levitt, M. H.; Raleigh, D. P.; Creuzet, F.; Griffin, R. G. *J. Chem. Phys.* **1990**, *92*, 6347-6365.
(9) Colombo, M. G.; Meier, B. H.; Ernst, R. R. *Chem. Phys. Lett.* **1988**, *146*, 189-196.
(10) Maas, W. E. J. R.; Veeman, W. S. *Chem. Phys. Lett.* **1988**, *149*, 170-174.
(11) Kubo, A.; McDowell, C. A. *J. Chem. Soc., Faraday Trans. 1* **1988**, *84*, 3713-3730.
(12) Gan, Z.-H.; Grant, D. M. *Mol. Phys.* **1990**, *67*, 1419-1430.
(13) Raleigh, D. P.; Creuzet, F.; Das Gupta, S. K.; Levitt, M. H.; Griffin, R. G. *J. Am. Chem. Soc.* **1989**, *111*, 4502-4503.
(14) Creuzet, F.; McDermott, A.; Gebhard, R.; van der Hoef, K.; Spijker-Assink, M. B.; Herzfeld, J.; Lugtenberg, J.; Levitt, M. H.; Griffin, R. G. *Science* **1991**, *251*, 783-786.
(15) Gullion, T.; Poliks, M. D.; Schaefer, J. *J. Magn. Reson.* **1988**, *80*, 553-558.
(16) Gullion, T.; Schaefer, J. *J. Magn. Reson.* **1989**, *81*, 196-200.
(17) Marshall, G. R.; Beusen, D. D.; Kociolek, K.; Redlinski, A. S.; Leplawy, M. T.; Pan, Y.; Schaefer, J. *J. Am. Chem. Soc.* **1990**, *112*, 963-966.

0002-7863/91/1513-9444$02.50/0