Journal of Biomolecular NMR, 4 (1994) 519–542
ESCOM



J-Bio NMR 190

# Calculations of one-, two- and three-bond nuclear spin–spin couplings in a model peptide and correlations with experimental data

Arthur S. Edison[a], John L. Markley[a,b,*] and Frank Weinhold[a,c,*]

[a]*Graduate Biophysics Program,* [b]*Department of Biochemistry and National Magnetic Resonance Facility at Madison and* [c]*Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin at Madison, WI 53706, U.S.A.*

Received 12 July 1993
Accepted 14 January 1994

*Keywords:* Coupling constants; Ab initio calculations; Errors in dihedral angles; Peptides

## SUMMARY

We present ab initio calculations of the Fermi contact term and experimental correlations of six coupling constants, $^{3}J_{H^{N}H^{\alpha}}$, $^{1}J_{C^{\alpha}H^{\alpha}}$, $^{2}J_{C'H^{\alpha}}$, $^{1}J_{C^{\alpha}N}$, $^{2}J_{C^{\alpha}N}$ and $^{1}J_{C'N}$, in a peptide as functions of the backbone dihedral angles, $\phi$ and $\psi$. Given estimates of experimental uncertainties, we find semiquantitative experimental correlations for $^{3}J_{H^{N}H^{\alpha}}$, $^{1}J_{C^{\alpha}N}$ and $^{2}J_{C^{\alpha}N}$, qualitative correlations for $^{1}J_{C^{\alpha}H^{\alpha}}$ and $^{2}J_{C'H^{\alpha}}$, but no experimental correlations of practical utility for $^{1}J_{C'N}$, owing to its complex dependence on at least four dihedral angles. Errors in the estimation of dihedral angles from X-ray crystallographic data for proteins, which result from uncertainties in atom-to-atom distances, place substantial limitations on the quantitative reliability of coupling constant calculations fitted to such data. In the accompanying paper [Edison, A.S. et al., *J. Biomol. NMR*, **4**, 543–551] we apply the results of the coupling constant calculations presented here to the estimation of $\phi$ and $\psi$ angles in staphylococcal nuclease from experimental coupling constants.

## INTRODUCTION

Dramatic advances in biomolecular NMR spectroscopy have been brought about by multidimensional methods (Ernst et al., 1987; Griesinger et al., 1989), methods of protein overproduction and stable isotopic labeling (Markley and Kainosho, 1993), and experimental techniques

*To whom correspondence should be addressed.

*Abbreviations:* AO, atomic orbital; BPTI, basic pancreatic trypsin inhibitor (bovine); CI-2, chymotrypsin inhibitor 2; E.COSY, exclusive correlation spectroscopy (Griesinger et al., 1986); $^{n}J_{AB}$, single bond (n = 1), geminal (n = 2), or vicinal (n = 3) coupling constant between nuclei A and B; LCAO, linear combination of atomic orbitals; NBO, natural bond orbital; n, $\sigma$ and $\sigma^*$, lone pair, bonding and antibonding orbitals, respectively; $\psi$, dihedral angle or molecular orbital wave function; $r^2$, correlation coefficient; RHF, restricted Hartree–Fock; rmsd, root-mean-square deviation; 3-21G and 6-31G*, molecular orbital basis set designations (Hehre et al., 1986).

designed to efficiently correlate atoms along the backbone and side chains of proteins (Bax and Grzesiek, 1993). These advances have led to efficient means for sequentially assigning spectra of proteins up to about 25 kDa (Wagner et al., 1992; Bax and Grzesiek, 1993). However, equally important has been the increased ability to measure parameters such as $T_1$, $T_2$ and homo- and heteronuclear coupling constants. These parameters have been underutilized in studies of proteins, owing in part to difficulties in their accurate measurement.

Experimental determination of coupling constants in proteins is limited by large linewidths, poor spectral resolution and poor sensitivity. Isotopic labeling and multidimensional techniques at least partly solve the problems of sensitivity and spectral resolution. The technique known as E.COSY (an extension of the original COSY (Aue et al., 1976)) was developed to allow accurate measurements of small passive couplings through separation of the peaks by active couplings in a second dimension (Griesinger et al., 1986). This strategy has been applied to the measurement of heteronuclear couplings (Neuhaus et al., 1984; Montelione et al., 1989; Wider et al., 1989; Chary et al., 1991; Delaglio et al., 1991; Edison et al., 1991; Griesinger and Eggenberger, 1992; references contained in Wagner et al., 1992; Madsen et al., 1993). More recently, progress has been made in quantifying couplings from peak intensities (Bax et al., 1992; Blake et al., 1992; Vuister et al., 1993). Although it is now possible to measure many couplings, a satisfactory theoretical basis for their interpretation is lacking.

Advances in computational resources, which have paralleled those in experimental NMR spectroscopy, now permit detailed theoretical calculations of biologically relevant molecules. In this paper we present ab initio molecular orbital calculations of $^3J_{H^NH^\alpha}$, $^1J_{C^\alpha H^\alpha}$, $^2J_{C'H^\alpha}$, $^1J_{C^\alpha N}$, $^2J_{C^\alpha N}$ and $^1J_{C'N}$ for a model peptide (Fig. 1) as functions of the dihedral angles $\phi$ and $\psi$. These coupling constants were chosen because experimental values from proteins with known X-ray structures have been determined recently, allowing us to correlate theoretical and experimental results (Table 1).

This paper is organized as follows. In the Methods section we provide details of the molecular orbital and coupling constant calculations, as well as of the fitting procedures employed in this work. Subsequently, we present relevant geometrical results from the ab initio calculations, coupling constant calculations and estimates of errors, associated with correlations of experimental and theoretical values. For each coupling constant, we show its ($\phi$,$\psi$) surface, the 2D Fourier coefficients of the function, and plots and statistics of experimental and theoretical couplings. In the analysis of errors, we estimate typical uncertainties in dihedral angles as determined from X-ray structures and evaluate our theoretical models in light of uncertainties in the NMR and

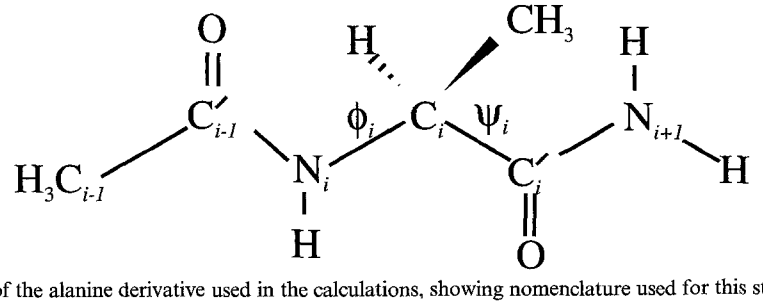


Fig. 1. Structure of the alanine derivative used in the calculations, showing nomenclature used for this study. The angles of rotation are indicated by $\phi_i$ and $\psi_i$. For couplings involving $^{15}N$, the coupling is denoted by the numbering of the $^{15}N$ atom.

TABLE 1
SOURCES OF EXPERIMENTAL DATA

| Protein | References | | | | | |
|---|---|---|---|---|---|---|
| | X-ray structure | Coupling constants | | | | |
| | | $^{3}J_{H^{N}H^{\alpha}}$ | $^{1}J_{C^{\alpha}H^{\alpha}}$ | $^{2}J_{C'H^{\alpha}}$ | $^{1}J_{C^{\alpha}N}$ | $^{2}J_{C^{\alpha}N}$ |
| BPTI | a | e | h | | | |
| Calmodulin | b | | h | i | | |
| CI-2 | c | f | | | | |
| Staphylococcal nuclease: pdTp-$Ca^{2+}$ ternary complex | d | g | h | i | J | j |

[a] Wlodawer et al., 1984.
[b] Babu et al., 1988.
[c] McPhalen and James, 1987.
[d] Loll and Lattman, 1989.
[e] Pardi et al., 1984.
[f] Ludvigsen et al., 1991.
[g] Kay et al., 1989.
[h] Vuister et al., 1993.
[i] Vuister and Bax, 1992.
[j] Delaglio et al., 1991.

X-ray results. We give a general discussion of the reliability of our calculations in light of our use of the same methods to calculate coupling constants in a series of small molecules (Edison et al., 1993). Finally, the results are summarized. The accompanying paper (Edison et al., 1994) demonstrates how these coupling constant calculations can be used with experimental NMR data to estimate $\phi$ and $\psi$ angles for a protein.

## METHODS

### *Ab initio calculations*

Fully geometry-optimized ab initio RHF/3-21G (Hehre et al., 1986) calculations have been obtained for the alanine derivative shown in Fig. 1 for rotations about the dihedral angles $\phi$ and $\psi$. After finding the global minimum at $\phi = -85.3°$, $\psi = 67.6°$ (Head-Gordon et al., 1991), we calculated 169 points on the $(\phi,\psi)$ surface at 30° increments by using the electronic structure package Spartan (Wavefunction, Inc.), installed on an FPS-522 computer. The starting geometries for each point were from an adjacent optimized point, and each calculation could be carried out in memory (requiring 240 Mb) in about 3–5 h. At each optimized geometry, single-point wave functions were calculated by using the GAMESS electronic structure package (Schmidt et al., 1990), in conjunction with the Natural Bond Orbital (NBO) analysis program (Foster and Weinhold, 1980; Reed et al., 1985). The NBO program serves two useful purposes for the coupling constant calculations. First, the NBO archive file provides a convenient, program- and machine-independent method of storing the complete molecular orbital wave function, and we have written a FORTRAN program* that calculates the Fermi contact term from data in the

*A Mathematica (Wolfram, 1991) package containing the calculated coupling surfaces and a FORTRAN program to calculate the Fermi contact contribution to coupling constants from NBO archive files are available upon request from: Operations Assistant, NMRFAM, Biochemistry Department, University of Wisconsin-Madison, 420 Henry Mall, Madison, WI 53706, U.S.A.

archive file. Second, NBO analysis yields an excellent chemical interpretation of the electronic wave function in terms of a set of occupied 'Lewis' and unoccupied 'non-Lewis' localized orbitals. Delocalization effects, important for nuclear spin–spin coupling constants, can be identified from the presence of off-diagonal elements of the Fock matrix in the NBO basis (Reed et al., 1988). In this study, we have used second-order perturbation theory estimates of NBO orbital interactions to qualitatively explain the physical origins of the angular dependence of calculated coupling constants. These orbital interactions are represented in terms of a donor → acceptor relationship, where the donor ('doubly occupied' Lewis-type orbital) is typically a bonding ($\sigma$) orbital or lone pair (n), and the acceptor ('unfilled' non-Lewis orbital) is typically an antibonding ($\sigma^*$) orbital. It must be stressed that, although NBO analysis provides a useful qualitative description of coupling constants, in order to obtain quantitative results for the Fermi contact term, the full molecular orbitals (independent of basis set) must be used.

*Coupling constant calculations*

A complete description of our method can be found elsewhere (Edison et al., 1993), so we will only summarize the calculations here. We used the sum over states (SOS) method to calculate the Fermi contact contribution to the coupling constant (Ramsey, 1953; Pople and Santry, 1964). The equation for the Fermi contact term is

$$J_{AB} = -(8\hbar/9\pi)\mu_0^2\mu_B^2\gamma_A\gamma_B \sum_{i}^{occ}\sum_{j}^{virt} \frac{1}{\varepsilon_j - \varepsilon_i} \langle\psi_i|\delta(r_A)|\psi_j\rangle \langle\psi_j|\delta(r_B)|\psi_i\rangle \quad (1)$$

where $\gamma_A$, $\mu_0$ and $\mu_B$ are the gyromagnetic ratio of nucleus A, the permeability of a vacuum, and the Bohr magneton, respectively, and where $\delta(r_A)$ is the Dirac delta function, which selects the value of occupied (unoccupied) molecular orbital $\psi_{i(j)}$ (with eigenvalue $\varepsilon_{i(j)}$) at nucleus A. Pople and Santry (1964) originally developed this approach for semiempirical wave functions and made the approximation that the valence atomic orbital (AO) centered on nucleus A makes the only contribution to the amplitude of $\psi_i$. We have shown (Edison et al., 1993) that this approximation is invalid, since even with modest split-valence basis sets (i.e., 3-21G), it produces little or no correlation between experimental and theoretical couplings. However, by using a 3-21G basis set and the full linear combination of atomic orbitals (LCAO),

$$\psi_i(A) = \sum_{\mu} c_{i\mu}\phi_\mu(A) \quad (2)$$

the correlation between theory and experiment for a series of small molecules with known geometry was quite high ($r^2 > 0.97$) for $^3J_{HH}$, $^2J_{HH}$ and $^1J_{CC}$, while for $^nJ_{CH}$ (n = 1–3), the correlations were somewhat lower ($r^2 = 0.90$ (n = 1), 0.75 (n = 2) and 0.84 (n = 3)), probably from the neglect of contributions other than the Fermi contact term (Edison et al., 1993). For excellent reviews of the theory of coupling constants, see Barfield and Grant (1965) and Kowalewski (1982).

Although the results of our calculations usually show a high linear correlation with experiment, the absolute magnitudes are always too small, for reasons discussed previously (Edison et al., 1993). We have concluded that for most couplings in 'regular' organic molecules, the RHF/3-21G level adequately represents important chemical trends (such as changes of dihedral angles, electronegativity and orientation of attached substituents, bond lengths and bond angles) at a computational cost low enough to allow investigations of relatively large (biologically significant) mole-

cules. However, as a result of the low magnitudes, we must resort to an empirical linear scaling of the calculated couplings

$$J^{fit} = aJ^{calc} + b \tag{3}$$

where a and b are determined from a linear regression analysis with experimental data.

*Fitting procedures*

Two separate fitting procedures have been used for this work. First, the calculated points (169, at 30° increments) were fitted to a 2D Fourier series

$$g(\phi,\psi) = \sum_{i}^{M}\sum_{j}^{N} \{\cos(i\phi)\cos(j\psi) + \cos(i\phi)\sin(j\psi) + \sin(i\phi)\cos(j\psi) + \sin(i\phi)\sin(j\psi)\} \tag{4}$$

by a least-squares minimization of the calculated points and $g(\phi,\psi)$. We found that a second-order fit ($M = N = 2$) with 25 coefficients was satisfactory.

As mentioned above, the absolute magnitudes of the couplings obtained from Eq. 1 have little direct resemblance to experimental values (though they are linearly highly correlated). Therefore, we used linear regression analysis to fit $J^{calc}$ (in its Fourier representation) to $J^{exp}$ (experimental couplings) to give $J^{fit}$, which in turn was used for the construction of the error surfaces shown in the following paper (Edison et al., 1994).

## RESULTS

*Selected geometries from the ab initio calculations*

Pople and co-workers (Head-Gordon et al., 1991) have published an analysis of the geometries and energetics of the $(\phi,\psi)$ surface of a similar alanine peptide and of a glycine peptide. Our results agree in all respects with theirs. A striking feature found on the potential energy surface (and on many coupling surfaces shown below) is the presence of cusped regions along diagonals from ($\phi = -180°$, $\psi = 0°$) to ($\phi = 0°$, $\psi = -180°$) and from ($\phi = 0°$, $\psi = 180°$) to ($\phi = 180°$, $\psi = 0°$) (Head-Gordon et al., 1991). Pople's group found that this cusp originates from large deviations in the planarity of the peptide bond (up to about 40° in $\omega$), resulting from severe steric clashes from groups along the backbone for some $(\phi,\psi)$ angles. Along the cusp, the peptide group shifts from a large deviation in one direction to a large deviation in the opposite direction. Both amide bonds (i and i + 1) flanking a $(\phi,\psi)$ pair exhibit similar responses. After extensive calculations at higher levels of theory, Pople and co-workers concluded that the distortions in the peptide group are physically real effects.

We find that numerous parameters respond in predictable ways to the peptide distortions. For example, Fig. 2 shows the $(\phi,\psi)$ surfaces of the peptide bond lengths (see Fig. 1 for atom nomenclature). The peptide bonds on both sides of the residue in question lengthen by up to about 0.05 Å in the regions of the largest deviations from peptide planarity, indicating a reduction of double-bond character in these bonds.

As shown below, changes in the single-bond coupling constants across these bonds are proportional to changes in their peptide bond lengths. However, since the lengths of both peptide bonds

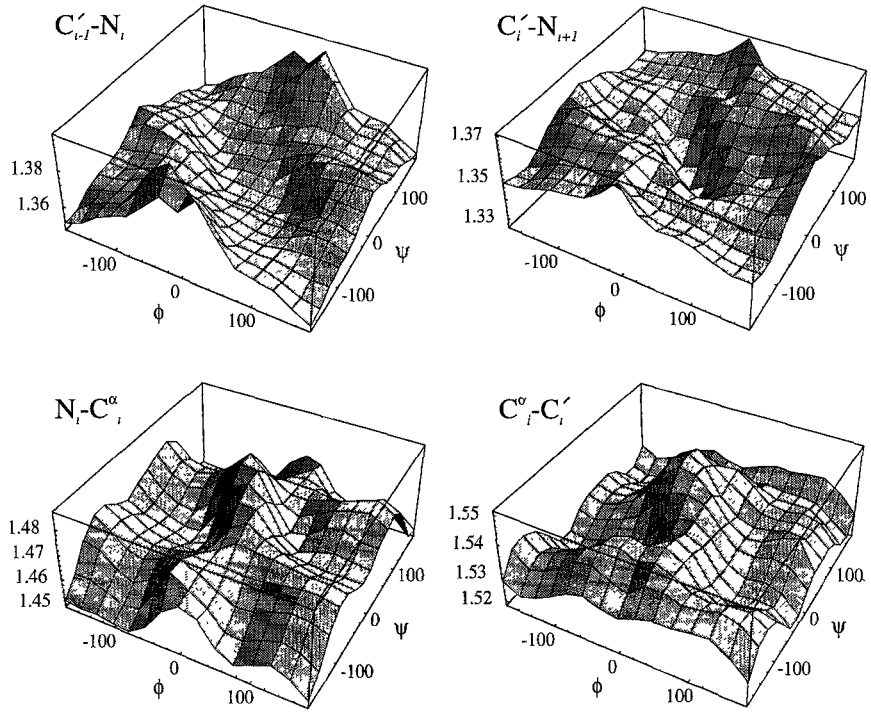


Fig. 2. The dependence of calculated peptide backbone bond lengths (in Å) on the dihedral angles $\phi_i$ and $\psi_i$ (see Fig. 1). Shown are the peptide bonds preceding ($C'_{i-1}$-$N_i$) and following ($C'_i$-$N_{i+1}$), the angles of rotation and the bonds across $\phi_i$ ($N_i$-$C^{\alpha}_i$) and $\psi_i$ ($C^{\alpha}_i$-$C'_i$). Points originated from optimized geometries at RHF/3-21G, calculated at 30° increments. The surfaces were generated by using the 'Interpolation' function in Mathematica (Wolfram, 1991).

respond to rotations about ($\phi$,$\psi$), $^1J_{C'N}$ has a very complicated dependence on these dihedral angles. Figure 2 shows that the bonds along the axes of rotation (N-$C^{\alpha}$ and $C^{\alpha}$-C') lengthen at points where groups become eclipsed. This is completely analogous to the ethane rotational barrier, where the C-C bond lengthens by about 0.014 Å in eclipsed geometries (RHF/6-31G*). For obvious reasons, this bond lengthening is most pronounced at ($\phi$ = 0°, $\psi$ = 0°). Not surprisingly, $^1J_{C^{\alpha}N}$ responds directly to the changes in bond length.

As discussed by Head-Gordon et al. (1991), these geometry changes point to a much more flexible peptide bond than is commonly assumed. We suspect that, if this flexibility were incorporated into the potentials of restrained molecular dynamics or simulated annealing algorithms, used for deriving protein structures from NMR data, the problems associated from local minima might be reduced.

*Calculated couplings and experimental correlations*

In this section, we present calculations of six coupling constants ($^3J_{HNH^{\alpha}}$, $^1J_{C^{\alpha}H^{\alpha}}$, $^2J_{C'H^{\alpha}}$, $^1J_{C^{\alpha}N}$, $^2J_{C^{\alpha}N}$ and $^1J_{C'N}$) for the model alanine peptide (Fig. 1), along with correlations of these couplings to experimental NMR data from proteins with known structures. All references to experimental data are shown in Table 1. In order to compare $J^{exp}$ and $J^{fit}$ ($\phi$,$\psi$), we have used the

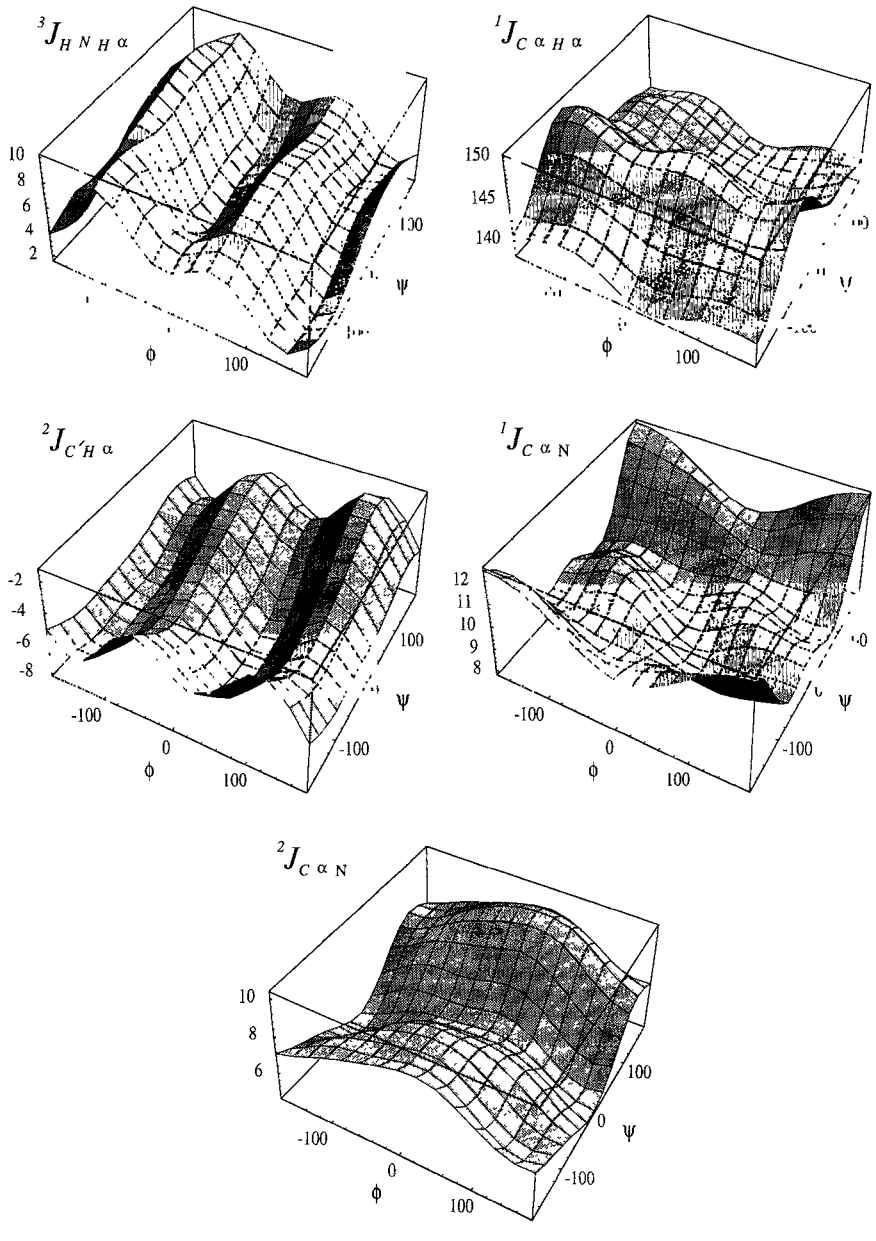


Fig. 3. Calculated coupling constant surfaces as functions of ϕ and ψ. The surfaces have points calculated at 30° increments using Eq. 1. The calculated points were first fitted to a 2D Fourier series to second order, and next linearly fitted to experimental protein data as described in the text. The Fourier coefficients giving rise to these surfaces are given in Table 2.

ϕ and ψ angles of the crystal structure to evaluate $J^{fit}$ (ϕ,ψ). For each coupling we include the calculated (ϕ,ψ) surface of $J^{fit}$ (ϕ,ψ) (Fig. 3), 2D Fourier coefficients (Table 2), experimental vs. theoretical plots (Fig. 4) and linear correlation parameters (Table 3). In addition, qualitative

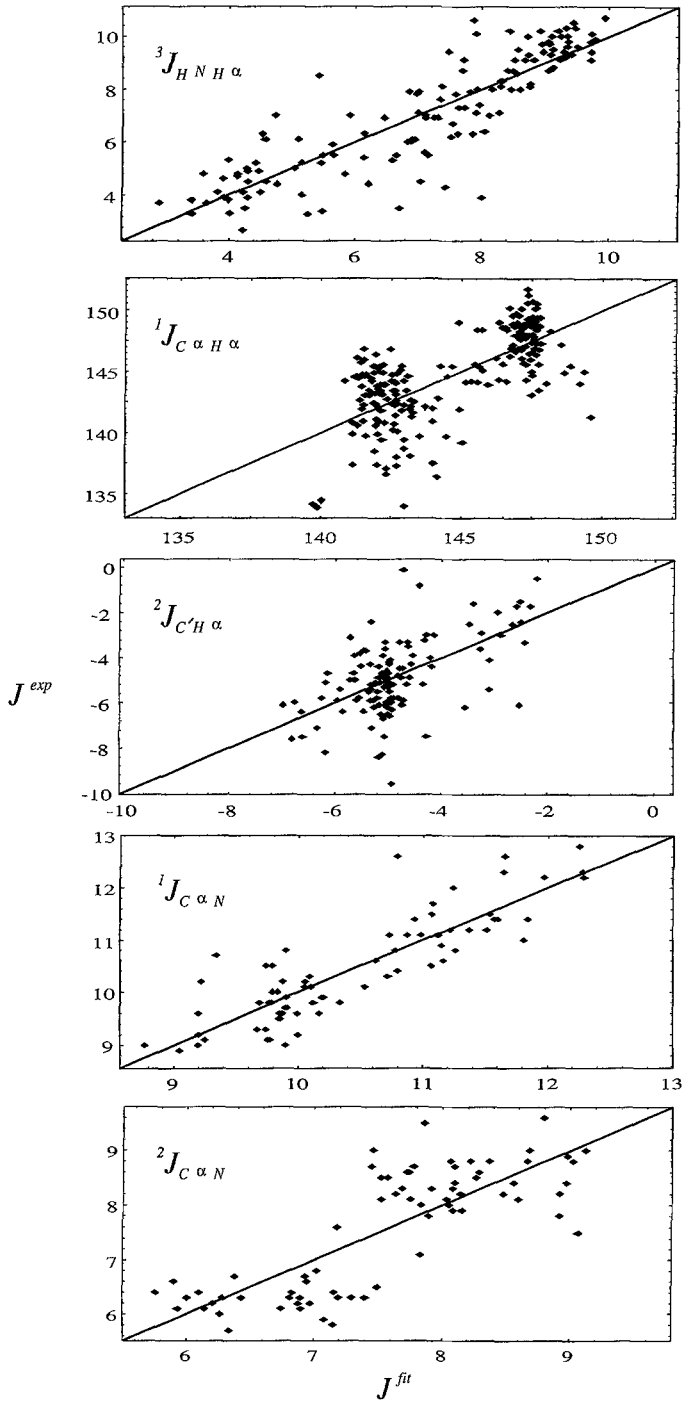


Fig. 4. Experimental ($J^{exp}$) vs. theoretical ($J^{fit}$) correlations of coupling constants. The points for $J^{fit}$ were obtained by evaluating the functions plotted in Fig. 3 at the X-ray crystal structure angles for the residue giving rise to $J^{exp}$. Statistics of the correlations are shown in Table 3.

descriptions of the physical origins of the angular dependence are presented in terms of NBO analysis (Reed et al., 1985).

### $^3J_{H^NH^\alpha}$

The $^3J_{H^NH^\alpha}$ coupling is the most widely measured and interpreted coupling constant in protein NMR. This coupling is interpreted by means of the remarkably simple and successful Karplus equation (Karplus, 1959):

$$J(\theta) = A\cos^2\theta + B\cos\theta + C \quad (5)$$

where $\theta$ is the dihedral angle between the coupled nuclei and A, B and C are empirically derived constants. The Karplus coefficients have been reparameterized several times on the basis of

TABLE 2
COEFFICIENTS DERIVED FROM FITTING THE CALCULATED COUPLING CONSTANTS BY SECOND-ORDER 2D FOURIER ANALYSIS IN TERMS OF $\phi$ AND $\psi$

| Coefficient | Coupling constants (Hz) | | | | |
|---|---|---|---|---|---|
| | $^3J_{H^NH^\alpha}$ | $^1J_{C^\alpha H^\alpha}$ | $^2J_{C'H^\alpha}$[a] | $^1J_{C^\alpha N}$ | $^2J_{C^\alpha N}$ |
| Constant | 5.6689 | 143.1890 | −5.0988 | 9.5893 | 7.4067 |
| cos(φ) | −0.3820 | −0.6600 | −0.2371 | −0.4213 | 1.1583 |
| cos(2φ) | −1.4542 | 0.1916 | −0.5616 | −0.1220 | −0.1630 |
| cos(ψ) | −0.0153 | 2.0431 | −0.7822 | −1.0621 | −1.1768 |
| cos(2ψ) | 0.0838 | −0.7417 | 0.0655 | 0.6450 | −0.4438 |
| sin(φ) | 1.0475 | −1.0455 | 0.3032 | −0.2758 | 0.0519 |
| sin(2φ) | 2.0166 | 0.3998 | −1.4789 | 0.0309 | 0.1325 |
| sin(ψ) | 0.0615 | −3.2753 | 1.1808 | 0.0353 | 0.5599 |
| sin(2ψ) | 0.2553 | −1.2630 | −0.2871 | −0.0999 | −0.3415 |
| cos(φ)cos(ψ) | −0.8415 | 0.4164 | −0.3793 | 0.3139 | −0.0369 |
| cos(2φ)cos(ψ) | 0.0537 | 0.7840 | 0.0473 | −0.2383 | −0.0828 |
| cos(φ)cos(2ψ) | 0.0851 | 0.0724 | 0.1183 | −0.4527 | −0.1618 |
| cos(2φ)cos(2ψ) | 0.1054 | −0.0286 | 0.0373 | −0.1618 | 0.0185 |
| cos(φ)sin(ψ) | −1.005 | −0.4843 | 0.1265 | 0.1086 | −0.0735 |
| cos(2φ)sin(ψ) | 0.5109 | 0.1541 | −0.3874 | −0.0448 | −0.0193 |
| cos(φ)sin(2ψ) | 0.2761 | −0.0304 | −0.0275 | −0.0050 | −0.0837 |
| cos(2φ)sin(2ψ) | −0.0774 | −0.0077 | 0.0066 | −0.0077 | 0.0570 |
| sin(φ)cos(ψ) | 0.7413 | 0.2247 | −0.3426 | −0.0971 | 0.0182 |
| sin(2φ)cos(ψ) | 0.5195 | 0.6147 | −0.1622 | 0.0130 | −0.0365 |
| sin(φ)cos(2ψ) | 0.0590 | 0.2539 | 0.0283 | 0.0074 | −0.0434 |
| sin(2φ)cos(2ψ) | −0.1065 | −0.3016 | −0.1086 | −0.0511 | 0.0267 |
| sin(φ)sin(ψ) | −0.1229 | −0.4461 | 0.1744 | 0.0897 | 0.0491 |
| sin(2φ)sin(ψ) | 0.1131 | −0.1301 | −0.0551 | 0.0885 | 0.0830 |
| sin(φ)sin(2ψ) | −0.1631 | −0.0600 | −0.0523 | 0.3231 | −0.3743 |
| sin(2φ)sin(2ψ) | −0.2914 | −0.2067 | −0.0863 | 0.3103 | 0.2400 |

The numbers shown have been corrected by linear fitting to the experimental data with correlation parameters shown in Table 3.

[a] Coefficients for $^2J_{C'H^\alpha}$ were obtained using all experimental data points.

experimental coupling data for proteins whose structures are known from X-ray crystallography. In an early study, Pardi et al. (1984) used the X-ray structures of the protein BPTI (Deisenhofer and Steigemann, 1975; Walter and Huber, 1983) to evaluate experimental $^3J_{H^N H^\alpha}$ data in terms of Eq. 5. Here the dihedral angle between the coupled protons was defined as

$$\theta = |\phi - 60°| \tag{6}$$

In more recent studies this empirical calibration has been improved slightly by including more measured data (Kay et al., 1989; Ludvigsen et al., 1991), but the values of A, B and C have remained essentially the same. Although there is little doubt that the general features of the Karplus equation are correct, many other factors such as bond angle, substituents, ionization and orientation of neighboring bonds are known to complicate the simple relationship of Eq. 5 (e.g., Karplus, 1963; Pachler, 1972; Marshall et al., 1976; Haasnoot et al., 1980; Barfield and Smith, 1992). We thought that the adjacent dihedral angle $\psi$ would be a likely candidate to influence $^3J_{H^N H^\alpha}$ (Fig. 1).

The ab initio coupling surface for $^3J_{H^N H^\alpha}$ $(\phi,\psi)$ is shown in Fig. 3. Our calculations show an obvious 'Karplus-like' behavior along $\phi$ with a slight dependence along $\psi$. A qualitative description of the angular dependence of $^3J_{H^N H^\alpha}$ is shown in Fig. 5.

As seen in this figure, the interaction of the $C_i^\alpha$-$H_i^\alpha$ bond with the $N_i$-$H_i^N$ antibond ($\sigma_{C\text{-}H} \rightarrow \sigma^*_{N\text{-}H}$) is closely related to the vicinal coupling constant. The cis orientation has a much smaller vicinal interaction than is apparent in the coupling constant. This demonstrates the need to evaluate *all* the orbital density at the coupled nuclei when calculating the Fermi contact term. The same type of interaction has been shown to be largely responsible for the rotational barrier in ethane-type molecules (Reed and Weinhold, 1991). Two related phenomena that give rise to the $\psi$-dependence need further explanation. First, rather large discontinuities occur at the cusp regions (Head-Gordon et al., 1991), as described above. These discontinuities, which arise from higher-order

TABLE 3
PARAMETERS AND STATISTICS FROM LINEAR FITS OF THE CALCULATED COUPLINGS TO THE EXPERIMENTAL COUPLINGS

| Coupling type | Slope (error) | Intercept (error) | $r^2$ | Rmsd (Hz) |
|---|---|---|---|---|
| $^3J_{H^N H^\alpha}$ (ab initio) | 4.57 (0.22) | 2.49 (0.24) | 0.76 | 1.09 |
| $^3J_{H^N H^\alpha}$ (Karplus)[a] | | | 0.78 | 1.02 |
| $^1J_{C^\alpha H^\alpha}$[b] | 2.54 (0.18) | 25.63 (8.43) | 0.48 | 2.67 |
| $^2J_{C'H^\alpha}$ | 5.01 (0.68) | −12.99 (1.10) | 0.31 | 1.38 |
| $^2J_{C'H^\alpha}$[c] | 5.29 (0.10) | −13.35 (0.52) | 0.45 | 1.08 |
| $^1J_{C^\alpha N}$[d] | −2.00 (0.14) | 4.21 (0.44) | 0.75 | 0.50 |
| $^2J_{C^\alpha N}$[d] | −7.04 (0.65) | −9.52 (1.57) | 0.63 | 0.69 |

The calculated couplings were first fitted to a 2D Fourier series to second order.

[a] The Karplus equation was fitted to all the data as described in the text, giving:
$6.4\cos^2(|\phi - 60|) - 1.54\cos(|\phi - 60|) + 1.65$.

[b] Adjustments for random-coil values were made as described in the text.

[c] Seven points deviating by more than 3 Hz were deleted from the data used for the linear regression.

[d] The signs of the calculated couplings were opposite to the reported experimental values for couplings involving $^{15}N$ (see text).

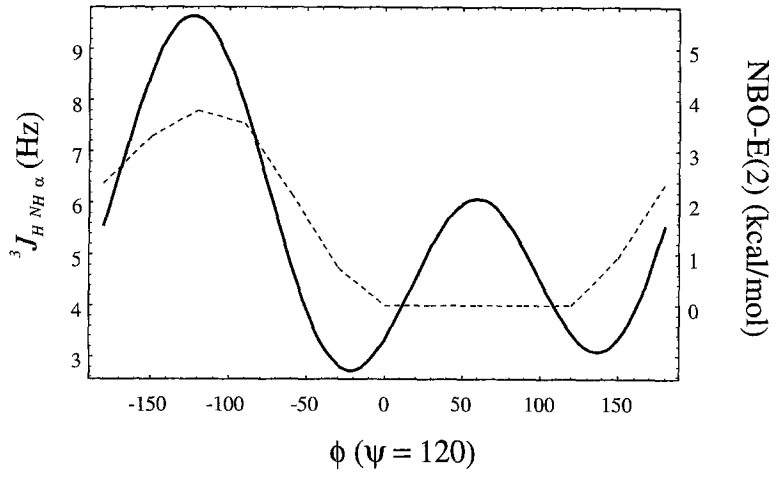


Fig. 5. Qualitative comparison of the NBO second-order perturbation interactions between $\sigma_{C^\alpha\text{-}H^\alpha} \rightarrow \sigma^*_{N\text{-}H^N}$ (dashed line), which are proportional to the ϕ-dependence (at ψ = 120°) of $^3J_{H^N H^\alpha}$ (solid line). The NBO points are spaced 30° apart, and the function for $^3J_{H^N H^\alpha}$ is from a slice of the 2D Fourier series (Table 2). The region from ϕ = 0° to ϕ = 120° is below a printing threshold of 0.5 kcal/mol.

Fourier coefficients, have been attenuated by fitting the data to a 2D Fourier series (Eq. 4). We believe that effects from these regions should be measurable, but that the situation is complicated by the fact that peptide bonds both preceding and following a particular (ϕ,ψ) pair are affected (Fig. 2). Therefore, the deviations from 'Karplus-like' behavior about $\phi_i$ will depend not only on $\psi_i$, but on $\phi_{i-1}$ and $\psi_{i-1}$ as well. The latter effects are neglected here. The second phenomenon

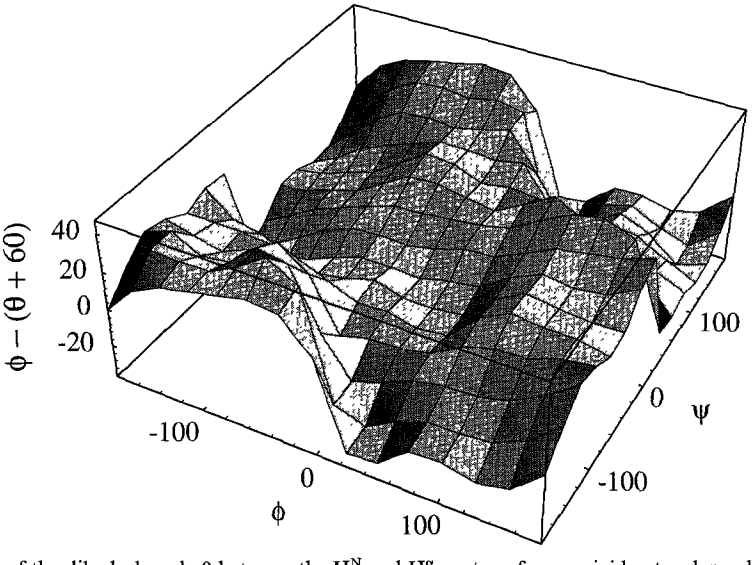


Fig. 6. Deviations of the dihedral angle θ between the $H^N$ and $H^\alpha$ protons from a rigid-rotor dependence on ϕ. The plot shows that the usual relationship θ = |ϕ − 60°| is an oversimplification. However, the large uncertainties from empirical correlations probably prevent this effect from being too noticeable.

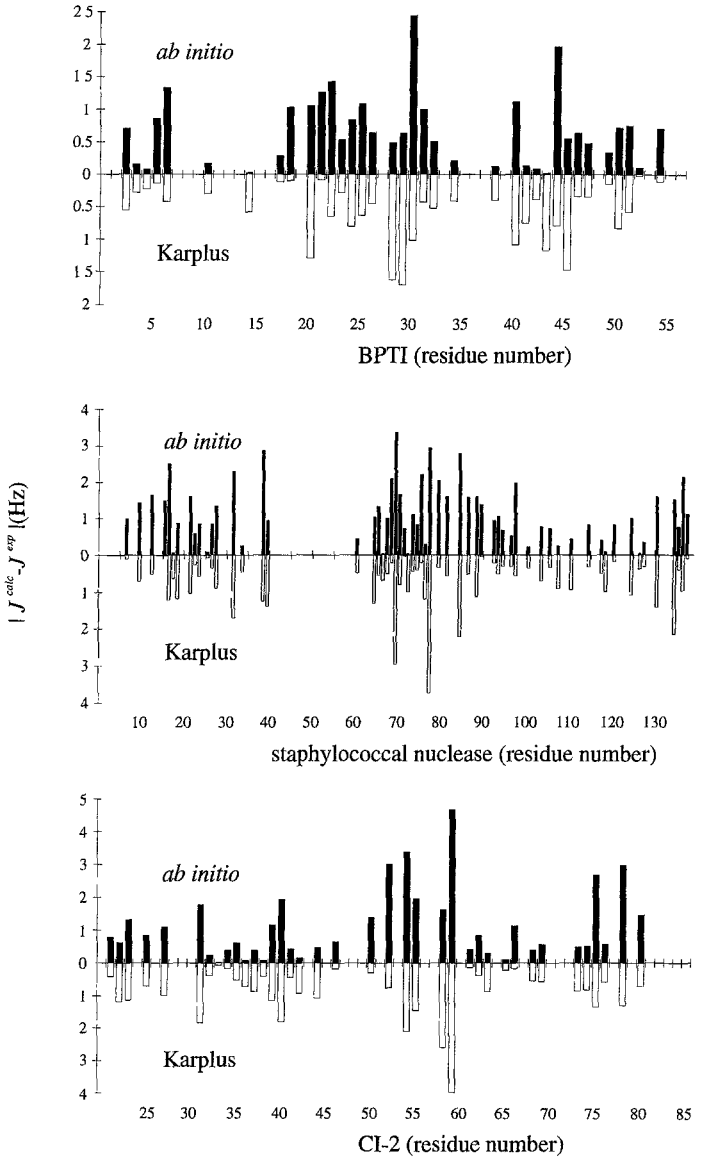


Fig. 7. Absolute values of the sequence-dependent deviations of the ab initio calculations (top) and Karplus (1959) equation (bottom) for BPTI, staphylococcal nuclease and CI-2. The deviations are highly correlated and tend to occur at regions with large crystal structure B-factors or with intermolecular contacts in the crystals. This suggests that many of the deviations between calculated and experimental couplings may be due to internal mobility or structural differences in solution and in the crystals.

giving rise to the $\psi$-dependence (related to the first) is that the actual dihedral angle between $H^N$ and $H^\alpha$ is *not* given by Eq. 6. Rather, this dihedral angle distorts in such a way that maxima occur along the cusps. The calculated dihedral angles and the ones predicted by Eq. 6 deviate by up to 40° (Fig. 6). These deviations give rise to a 'bowing out' of the maxima around $\psi = 0°$.

To our surprise, the ab initio results for $^3J_{H^N H^\alpha}$ were almost identical to those derived with the

Karplus equation (Eq. 5). Table 3 shows that the linear correlation for the Karplus equation and the ab initio results gave almost identical root-mean-squared deviations (rmsd) and regression coefficients ($r^2$). Figure 7 shows the absolute value of the deviations of theory (Karplus equation and our ab initio results) from experiments as a function of protein sequence for BPTI, staphylococcal nuclease and CI-2. It is clear from Fig. 7 that the deviations of the two theoretical approaches for treating the experimental NMR data from the X-ray results are highly correlated.

From an analysis of the estimated experimental errors in the X-ray and NMR data (see below), it appears that the rms deviations and $r^2$ values for both the ab initio- and Karplus-derived coupling constants are at or near the limits imposed by experimental error. Thus, it is not surprising that more rigorous methods fail to give 'improvements' (as measured by $r^2$ correlations with experiment). The errors in the data used to calibrate relationships such as the Karplus equations, chemical shift correlations or our ab initio calculations place fundamental limitations on the accuracy of such theoretical models. This problem has been largely ignored in the literature. Another reason for the similar results is that regions giving rise to large distortions in the peptide bond are not highly populated in a protein, so the largest differences will not be seen experimentally.

### $^1J_{C^\alpha H^\alpha}$

The $^1J_{C\alpha H\alpha}$ coupling has been the subject of considerable theoretical and empirical investigations (Barfield and Johnston Jr., 1972; Bystrov, 1976; Egli and Von Philipsborn, 1981; Hansen, 1981; Gil and Von Philipsborn, 1989; Vuister et al., 1992,1993). A partial list of the phenomena known to influence $^1J_{C\alpha H\alpha}$ includes: solvent and related electric field effects (Barfield and Johnston Jr., 1972), substituent effects (Hansen, 1981), lone electron pairs, and dihedral angle orientation (Egli and Von Philipsborn, 1981; Gil and Von Philipsborn, 1989). Hansen (1981) estimated that solvent effects on single-bond C-H couplings can be as large as 7%; this number is supported by experimental and theoretical results compiled by Barfield and Johnston Jr. (1972). Several empirical studies have demonstrated the dependence of $^1J_{C\alpha H\alpha}$ on the dihedral angles $\phi$ and $\psi$ in peptides (Egli and Von Philipsborn, 1981; Hansen, 1981; Gil and Von Philipsborn, 1989) and proteins (Vuister et al., 1992,1993). The relative contributions of $\phi$ and $\psi$ to $^1J_{C\alpha H\alpha}$ are disputed. The original work of Egli and Von Philipsborn (1981), which was based on a relatively small group of constrained peptides, was interpreted as showing that $^1J_{C\alpha H\alpha}$ depends more heavily on $\phi$ than on $\psi$. Kessler's group extended this work to a larger number of peptides and proteins and concluded again that $^1J_{C\alpha H\alpha}$ depends more on $\phi$ than on $\psi$ (Mierke et al., 1992). More recent empirical work with a large protein data set, however, has shown a much stronger dependence on $\psi$ than on $\phi$ (Vuister et al., 1992,1993). Amino acid substituents have been shown to influence this coupling (Hansen, 1981), and the random-coil values for each amino acid group (except glycine and cysteine) have been measured (Vuister et al., 1993). We are not aware of a previous theoretical study of the full ($\phi$,$\psi$) dependence of $^1J_{C\alpha H\alpha}$.

Our ab initio surface for $^1J_{C\alpha H\alpha}$ is shown in Fig. 3. This surface shows predominantly a $\psi$-dependence with a smaller $\phi$-dependence, in accordance with recent empirical calibrations from proteins (Vuister et al., 1992,1993). The general features of the empirical surface (Vuister et al., 1993) shown in Fig. 8 are extremely similar to the ab initio surface (Fig. 3). Comparison of Figs. 3 and 8 reveals that the major difference is in the region around $\psi = 120°$, where the ab initio calculation shows a smaller maximum.

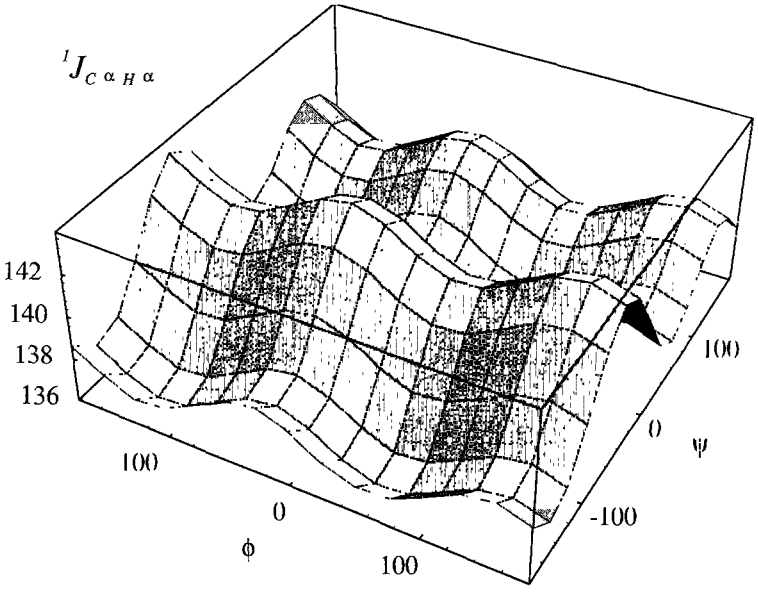


Fig. 8. Plot of the empirical function derived by Vuister et al. (1993): $^1J_{C^\alpha H^\alpha} = A + B \sin(\psi + 138°) + C \cos^2(\psi + 138°) + D \cos^2(\phi + 30°)$, with A = 140.3, B = 1.4, C = −4.1 and D = 2.0. This function is similar to our calculated results for $^1J_{C^\alpha H^\alpha}$ shown in Fig. 3, except for the more pronounced maximum near $\psi = 120°$.

To demonstrate the origins of the dihedral angle effects, 1D slices through the ab initio $^1J_{C^\alpha H^\alpha}$ surface at $\phi = -120°$ and $\psi = -60°$ are shown with selected second-order NBO (Reed et al., 1985) interaction energies (see the Methods section above) in Fig. 9. Qualitatively, the angular dependence along $\phi$ depends primarily on the interaction of the electron pair of $N_i$ with the antibond of $C_i^\alpha$-$H_i^\alpha$ (Gil and Von Philipsborn, 1989). The $\psi$-dependence of $^1J_{C^\alpha H^\alpha}$ is somewhat more complicated. We see large angular dependencies for the interactions of the bonding orbital of $C_i^\alpha$-$H_i^\alpha$ with both of the antibonding orbitals of $C_i'$-$O_i$ and the antibonding orbital of $C_i'$-$N_{i+1}$. Interactions with the $C_i'$-$N_{i+1}$ antibond coincide with an increase in $^1J_{C^\alpha H^\alpha}$, while those with the $C_i'$-$O_i$ coincide with a decrease.

The experimental vs. theoretical correlations of $^1J_{C^\alpha H^\alpha}$ are shown in Fig. 4 and Table 3. It is clear that the correlation is only qualitative and that, in addition to $\phi$ and $\psi$, many parameters might influence $^1J_{C^\alpha H^\alpha}$ (Hansen, 1981). Unlike the model for $^3J_{H^N H^\alpha}$, that for $^1J_{C^\alpha H^\alpha}$ introduces an rmsd of up to 2 Hz beyond that expected from experimental uncertainties (see below). The experimental couplings used to make the comparison were corrected by the random-coil values (Vuister et al., 1993) to account for the fact that our calculations were only for an alanine residue. For a given residue i, we added to the measured value of i the difference between the random-coil value of i and the random-coil value of alanine. It is interesting to note the similarity between the second plot in our Fig. 4 and Fig. 4 from Vuister et al. (1993). Both correlations show dominant ridges along the experimental axis. In analyzing the discrepancies between our results and experiment, we discovered that arginine, aspartic acid and lysine had unusually high percentages of disagreement compared with other residues. Proline was found to have unusually large values of $^1J_{C^\alpha H^\alpha}$ (Vuister et al., 1993), but we found that, after the correction for random-coil values was

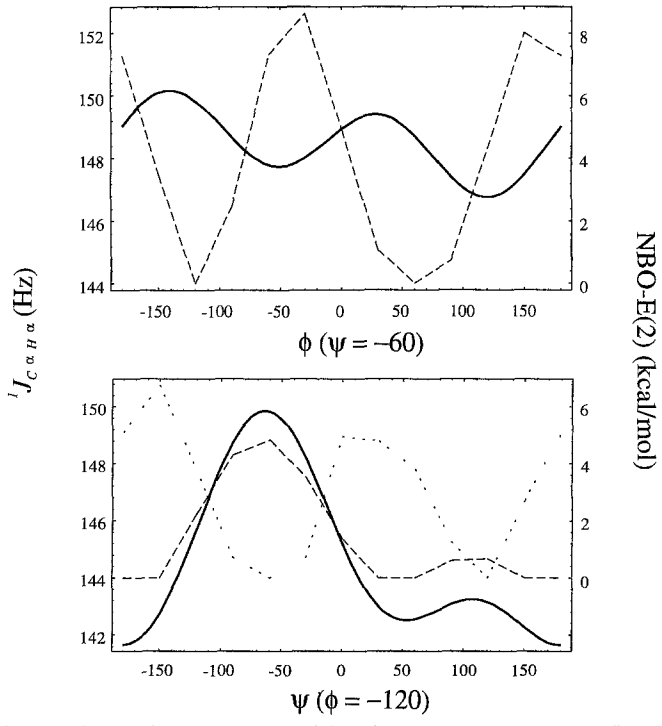


Fig. 9. Qualitative analysis of the physical interactions giving rise to the ϕ- (top) and ψ- (bottom) dependence of $^1J_{C^\alpha H^\alpha}$ (solid lines) from perturbative estimates of NBO interaction energies (dotted and dashed lines). The ϕ-dependence can be justified by the vicinal interaction of the nitrogen lone pair with the $C_i^\alpha$-$H_i^\alpha$ antibond ($n_N \rightarrow \sigma^*_{C^\alpha\text{-}H^\alpha}$). The ψ-dependence of $^1J_{C^\alpha H^\alpha}$ is primarily dependent on two interactions with opposite effects: a positive interaction of the $C_i^\alpha$-$H_i^\alpha$ bond with the vicinal $C'_i$-$N_{i+1}$ antibond ($\sigma_{C^\alpha\text{-}H^\alpha} \rightarrow \sigma^*_{C'\text{-}N}$; bottom panel, dashed line) and a negative interaction of the $C_i^\alpha$-$H_i^\alpha$ bond with the vicinal $C'_i = O_i$ sigma-like antibond ($\sigma_{C^\alpha\text{-}H^\alpha} \rightarrow \sigma^*_{C'\text{-}O}$; bottom panel, dotted line).

made (as described in the text), proline residues deviated only slightly more than the average deviation for all residues. These discrepancies suggest many possibilities, including sequence-dependent electric field or solvent effects (Barfield and Johnston Jr., 1972; Hansen, 1981). Also, it is possible that salt bridge interactions prevented the 'random-coil' values determined from unfolded proteins (Vuister et al., 1993) from being truly random.

### $^2J_{C'H^\alpha}$

Geminal coupling constants are notoriously difficult to treat quantitatively, but trends resulting from electronegative substituents can be understood qualitatively (Pople and Bothner-By, 1965; Jameson and Damasco, 1970). The orientations of adjacent lone pairs and π-orbitals are known to modulate the magnitudes of many geminal couplings (Jameson and Damasco, 1970; Barfield et al., 1976; Bystrov, 1976; Mohanakrishnan and Easwaran, 1979; Gil and Von Philipsborn, 1989). The large dependence on bond angles and bond lengths, which gives rise to large changes in molecular orbital nodal regions in the vicinity of the coupled nuclei, complicates the calculation of geminal couplings (Pople and Bothner-By, 1965; Hansen, 1981). A rather large pH dependence has been observed for $^2J_{C'H^\alpha}$ couplings in peptides: values may vary by as much as 1 Hz

(~ 20%), resulting in lower values of the coupling at pH > 10 and higher values at low pH (Hansen et al., 1975). Since these pH ranges are not within the expected pKa values of either the amide proton or carbonyl oxygen atoms of the polypeptide backbone (Creighton, 1984), the mechanism for the variation of $^2J_{C'H\alpha}$ is uncertain. Our ab initio surface of $^2J_{C'H\alpha}$ shows a regular and strong $\phi$-dependence and a moderate $\psi$-dependence (Fig. 3). Figure 10 shows that certain NBO orbital interactions are proportional to the calculated couplings. We find that much of the $\phi$-dependence can be associated with the interactions of the nitrogen lone pair and the $C_i^\alpha$-$H_i^\alpha$ antibond ($n_N \rightarrow \sigma^*_{C\text{-}H}$). The $\psi$-dependence appears to be complicated by many simultaneous effects, but the interactions of the $C_i^\alpha$-$H_i^\alpha$ bond with the sigma-like vicinal $C'_i$-$O_i$ antibond ($\sigma_{C\text{-}H} \rightarrow \sigma^*_{C\text{-}O}$) is roughly proportional to the calculated couplings (Fig. 10).

The agreement of the ab initio calculations with experiment for $^2J_{C'H\alpha}$ is only qualitative (Fig. 4 and Table 3). Our $^2J_{C'H\alpha}$ calculations probably rely on an oversimplified model. Kowalewski and co-workers have shown that geminal couplings have a large dependence on electron correlation (Kowalewski et al., 1979; Laaksonen and Kowalewski, 1981; Kowalewski, 1982). Others have found that, in addition to the Fermi contact term (Eq. 1), the spin-orbit and spin-dipolar terms are important for geminal couplings (Lee and Schulman, 1979). The reported pH dependence (Hansen et al., 1975) of $^2J_{C'H\alpha}$ might cause some significant deviations across the tertiary

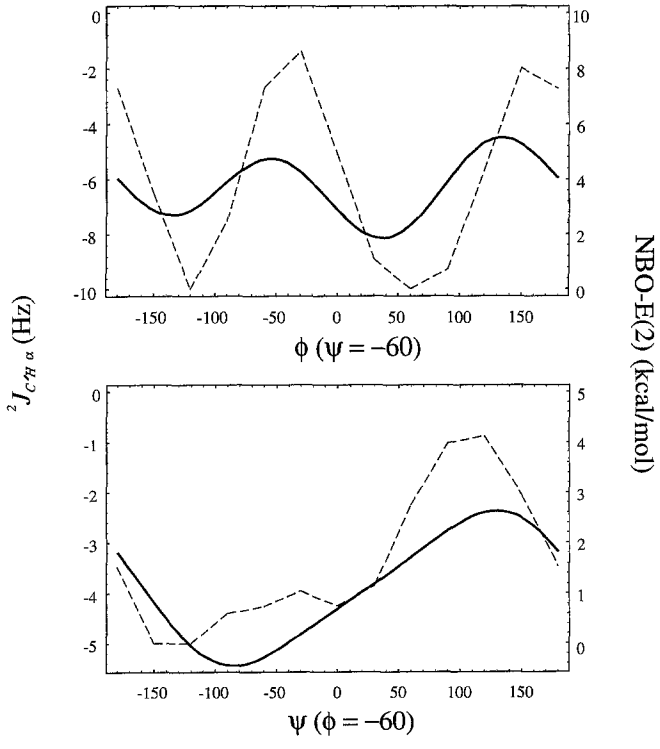


Fig. 10. Qualitative analysis of the physical interactions giving rise to the $\phi$- (top) and $\psi$- (bottom) dependence of $^2J_{C'H\alpha}$ (solid lines) from perturbative estimates of NBO interaction energies (dashed lines). The $^2J_{C'H\alpha}$ $\phi$-dependence is proportional to the interaction of the neighboring nitrogen lone pair with the $C_i^\alpha$-$H_i^\alpha$ antibond ($n_N \rightarrow \sigma^*_{C^\alpha\text{-}H^\alpha}$). The $\psi$-dependence shows a strong interaction of the $C_i^\alpha$-$H_i^\alpha$ bond with the vicinal $C'_i = O_i$ antibond ($\sigma_{C^\alpha\text{-}H^\alpha} \rightarrow \sigma^*_{C\text{-}O}$).

structure of a protein, resulting from changes in dielectric and hydrophobic properties in different regions. Unlike the other couplings examined in this study, several (seven) experimental points for $^2J_{C'H^\alpha}$ deviated significantly from the rest. In Table 3 we show linear regressions, obtained with and without these seven experimental points. The theoretical difficulties and large number of parameters known to contribute to $^2J_{C'H^\alpha}$ conspire against a rigorous, quantitative interpretation of this coupling. However, the calculated angular $\phi$- and $\psi$-dependence (Fig. 3) seems to suggest a qualitative relationship between $^2J_{C'H^\alpha}$ and these conformational angles.

### $^1J_{C^\alpha N}$

The ab initio surface for $^1J_{C^\alpha N}$ is shown in Fig. 3. Not surprisingly, the length of the $C^\alpha$-N bond (Fig. 2) is inversely proportional to $^1J_{C^\alpha N}$. The bond-length dependence on $(\phi,\psi)$ is easily understood, simply by steric considerations: when the peptide backbone is in a fully extended conformation ($\phi \approx \psi \approx 180°$), the $C^\alpha$-N bond is shortest. When either $\phi$ or $\psi$ moves significantly toward an eclipsed conformation, the $C^\alpha$-N bond lengthens by about 0.03 Å. Unexpectedly, the effects of rotation about either $\phi$ or $\psi$ are nearly equal.

The agreement between experiment and theory for $^1J_{C^\alpha N}$ is quite good (Fig. 4 and Table 3). Our calculations of the couplings involving $^{15}N$ predict negative signs, in disagreement with the reported signs of the experimental couplings (Delaglio et al., 1991) but in agreement with previous reports of single-bond C-N couplings (Levy and Lichter, 1979). Our error analysis (shown below) suggests that the theoretical model for $^1J_{C^\alpha N}$ gives values that lie within the expected experimental uncertainties.

### $^2J_{C^\alpha N}$

Figure 3 shows that $^2J_{C^\alpha N}$ is characterized by a strong $\psi_{i-1}$-dependence, which is very similar to the empirical correlation (Delaglio et al., 1991)*. In addition, we find a noticeable $\phi_{i-1}$-dependence with a maximum when $\phi_{i-1}$ approaches 0°. A rigorous interpretation of the physical origins underlying the dihedral angle dependence is difficult for $^2J_{C^\alpha N}$. Obviously, all of the theoretical uncertainties associated with calculations of geminal couplings discussed for $^2J_{C'H^\alpha}$ apply here as well. Figure 11 shows that the calculated $\psi$-dependence for $^2J_{C^\alpha N}$ varies with the NBO vicinal interactions of the $N_i$-$C_i^\alpha$ bond with the $C'_i$-$N_{i+1}$ antibond ($\sigma_{N\text{-}C} \rightarrow \sigma^*_{C'\text{-}N}$). We find the calculated $\phi$-dependence of $^2J_{C^\alpha N}$ to be proportional to interactions of the $C_i^\alpha$-$C'_i$ bond with the vicinal $N_i$-$C'_{i-1}$ antibond ($\sigma_{C\text{-}C'} \rightarrow \sigma^*_{N\text{-}C'}$) (Fig. 11).

Figure 4 and Table 3 show that the agreement between experiment and theory is quite good for $^2J_{C^\alpha N}$. Although calculated values of $^2J_{C^\alpha N}$ showed much better correlation with experiment than did $^2J_{C'H^\alpha}$, the rmsd was greater than expected on the basis of estimated experimental uncertainties (discussed below). The excess error (about 0.3 Hz) probably results from an inadequate theoretical model, which might be improved by taking into account solvation and β-substituent dependence.

### $^1J_{C'N}$

We found no correlation between experiment and theory for $^1J_{C'N}$. Figure 12 shows the angular

*When a coupling involving $^{15}N$ crosses the peptide bond, the numbering convention we use refers to the residue containing the $^{15}N$ nucleus as 'i' (Delaglio et al., 1991).

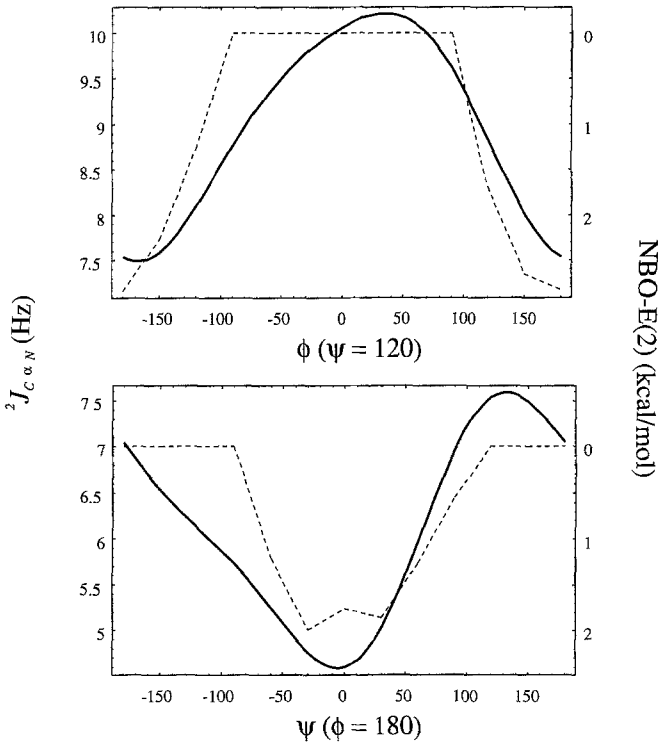


Fig. 11. Qualitative analysis of the physical interactions giving rise to the $\phi$- (top) and $\psi$- (bottom) dependence of $^2J_{C^\alpha N}$ (solid lines) from perturbative estimates of NBO interaction energies (dashed lines). The $\phi$-dependence is proportional to the interaction of the $C_i^\alpha$-$C_i'$ bond with the previous (vicinal) $C_{i-1}'$-$N_i$ antibond ($\sigma_{C^\alpha C'} \rightarrow \sigma^*_{C'N}$). The $\psi$-dependence could be due to the interaction of the $N_i$-$C_i^\alpha$ bond with the $C_i' = O_i$ antibond of the same residue ($\sigma_{N\text{-}C^\alpha} \rightarrow \sigma^*_{C'O}$). Both of the NBO energy estimates fall below a threshold of 0.5 kcal/mol, where they are shown to be zero.

dependence of both $^1J_{C'N(i)}$ and $^1J_{C'N(i+1)}$ as functions of $\phi_i$ and $\psi_i$. The couplings in Fig. 12 are the directly calculated values that have not been fitted to experimental data, so the scales are not in agreement with experimental data. The important feature to recognize is that both couplings ($^1J_{C'N(i)}$ and $^1J_{C'N(i+1)}$) have a strong angular dependence on the same set of dihedral angles, resulting from the peptide bond distortions described above (Head-Gordon et al., 1991).

Delaglio and co-workers (1991) found no correlation with X-ray angles or secondary structure, except that most of the residues with values of $^1J_{C'N}$ greater than 16 Hz were preceded by $\psi$ angles near 0°. Our calculated signs disagree with the reported experimental signs for couplings involving $^{15}N$ (Delaglio et al., 1991) but agree with other reports of couplings involving C and N (Levy and Lichter, 1979). Our surfaces suggest large values of the couplings when both $\phi$ and $\psi$ are near 0°, but we would also expect large values for the couplings in extended sheet conformations with $\phi$ and $\psi$ both near 180°. In short, the behavior of $^1J_{C'N}$ appears to be too complicated for this coupling to be of any practical value in estimating dihedral angles, since it should be modeled as a function of at least $\phi_i$, $\psi_i$, $\phi_{i-1}$ and $\psi_{i-1}$. However, this coupling might provide an experimental approach for evaluating the validity of the calculated conformational flexibility of the peptide bond (Head-Gordon et al., 1991).

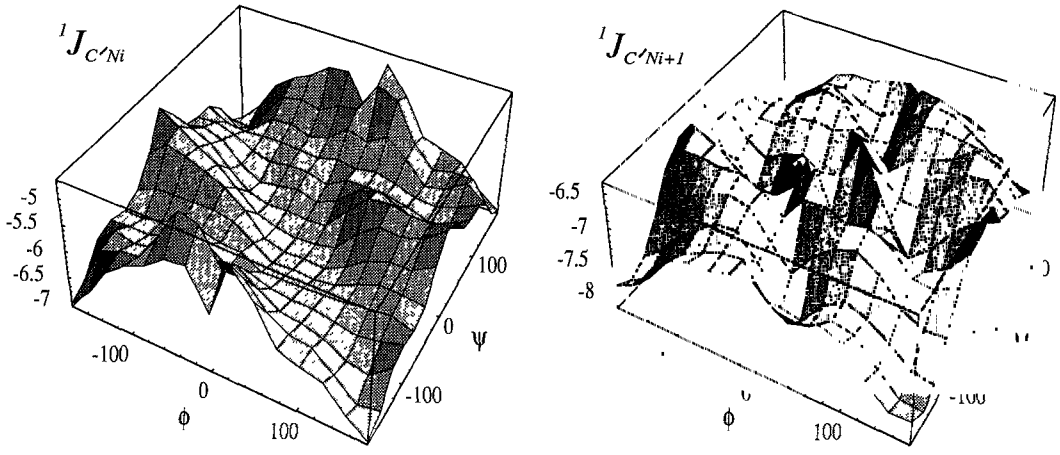


Fig. 12. Dihedral angle dependence of $(^1J_{C'N})_i$ and $(^1J_{C'N})_{i+1}$. The numbers for the coupling constants are from Eq. 1 with no fitting to experimental data, so the absolute numbers have no correspondence to experimental values. Since both couplings depend to almost the same degree on the same set of angles, this indicates that $^1J_{C'N}$ can be modeled only as $^1J_{C'N}(\phi_i,\psi_i,\phi_{i-1},\psi_{i-1})$.

## DISCUSSION

*General reliability of calculated couplings*

The SOS equation for the Fermi contact term of the coupling (Eq. 1) has many theoretical pitfalls (Edison et al., 1993). Among the worst in terms of practical applications is that, as the size of the basis set is increased, the number of terms from the summation of unoccupied orbitals increases without bounds. Thus, the convergence properties of Eq. 1 are uncertain at best. We conducted an extensive study of small molecules with fixed and known geometries in 3-21G and 6-31G* basis sets (Edison et al., 1993). This study showed that, for couplings involving $^{13}C$, the larger basis (6-31G*) performed noticeably worse than the smaller basis (3-21G). The usual assumption, that the molecular orbital density at the nucleus arises entirely from the valence-s orbital centered at that nucleus (Pople and Santry, 1964), was found to be invalid. With this approximation, even the results from the 3-21G basis sets had no useful correlation with experimental values (Edison et al., 1993). However, by evaluating Eq. 1 at the RHF/3-21G level, we found quite high ($r^2 > 0.95$) linear correlations with experimental values for $^3J_{HH}$, $^2J_{HH}$ and $^1J_{CC}$ in small molecules. For $^nJ_{CH}$, we obtained correlation coefficients of 0.89, 0.75 and 0.84 for n = 1, 2 and 3. Although we did not previously study couplings involving $^{15}N$, we expect results similar to those involving $^{13}C$. These previous results for small molecules are generally consistent with the trends that we observe for the couplings calculated in the present work, where we find the best results for $^3J_{HNH\alpha}$ and $^1J_{C\alpha N}$ and more qualitative results for $^1J_{C\alpha H\alpha}$, $^2J_{C'H\alpha}$ and $^2J_{C\alpha N}$. However, the overall experimental correlations found here are considerably worse than those found for small molecules in well-determined geometries (Edison et al., 1993). Clearly, some portion of this additional discrepancy is due to the fact that both NMR and crystallographic data for proteins are subject to considerably larger experimental errors. In the next section, we attempt to characterize the effects of these experimental errors on the correlations with our theoretical models.

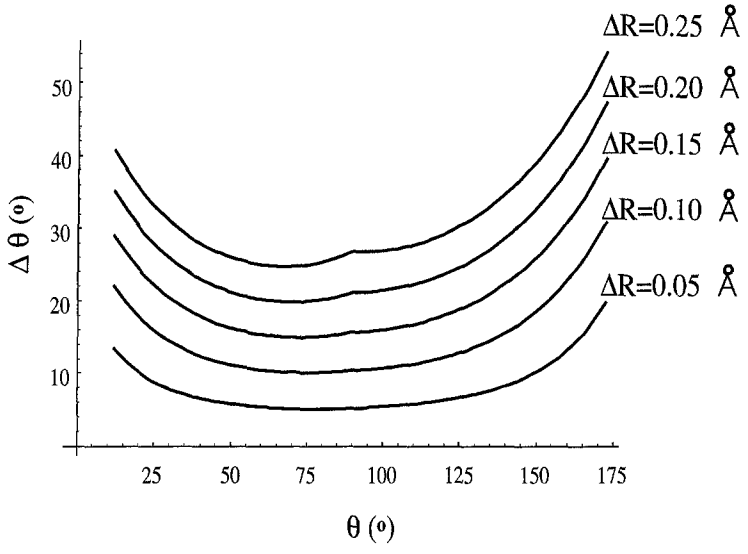


Fig. 13. Propagation-of-errors calculation of the error in the dihedral angle θ between atoms A-B-C-D as a function of errors in distances (in Å) between A and D (shown next to each curve) as described in the text. The relatively large errors constitute one of the problems in using X-ray data to parameterize NMR coupling constants.

*Analysis of experimental errors*

Correlations of experimental couplings from solution NMR data with macromolecular geometries from X-ray crystallographic data have *at least* five serious sources of error:

(1) dihedral angles determined from X-ray data contain uncertainties;

(2) coupling constants measured from NMR data of proteins contain errors;

(3) the crystal and solution structures differ in certain regions;

(4) local motions in proteins in solution may average couplings; and

(5) the theoretical functions that relate couplings to dihedral angles may be erroneous.

Since our aim in this study has been to test the reliability of our calculated couplings, we have attempted to address the first four sources of error and to assume that any further error results from our theoretical treatment.

Errors in distances from biomolecular X-ray crystallographic data are difficult to estimate (Loll and Lattman, 1989 and references contained therein), and the propagation of distance errors into dihedral angle errors is complicated by coupled errors in distances. Of the X-ray structures used in this study, the authors report rms coordinate errors of 0.16, 0.24–0.30 and 0.15 Å for staphylococcal nuclease (Loll and Lattman, 1989), CI-2 (McPhalen and James, 1987), and calmodulin (Babu et al., 1988), respectively. To estimate the error in dihedral angles calculated from X-ray structures, we have used Mathematica (Wolfram, 1991) to calculate the error in the dihedral angle θ between four atoms A-B-C-D, using the standard equation for the dihedral angle

$$\cos(\theta) = \frac{\cos(ABD) - \cos(ABC)\cos(CBD)}{\sin(ABC)\sin(CBD)} \tag{7}$$

The terms of the right-hand side of Eq. 7 are easily calculated by the cosine law from the X-ray distances. Since we are actually calculating the error in cos(θ), the error in θ is nonlinear and

TABLE 4
ERROR ANALYSIS OF CALCULATED COUPLINGS[a]

| | Rmsd (Hz) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma J^{exp}$ (Hz) | $^3J_{H^NH^\alpha}$ | | | | $^1J_{C^\alpha H^\alpha}$ | | | | $^2J_{C'H^\alpha}$ | | | |
| 0.0 | 0 | 0.53 | 0.74 | 1.00 | 0 | 0.35 | 0.47 | 0.66 | 0 | 0.26 | 0.34 | 0.47 |
| 0.5 | 0.29 | 0.58 | 0.82 | 1.07 | 0.29 | 0.43 | 0.60 | 0.73 | 0.29 | 0.37 | 0.47 | 0.54 |
| 1.0 | 0.58 | 0.79 | 0.97 | 1.16 | 0.57 | 0.68 | 0.75 | 0.88 | 0.58 | 0.67 | 0.69 | 0.76 |
| | $^1J_{C^\alpha N}$ | | | | $^2J_{C^\alpha N}$ | | | | | | | |
| 0.0 | 0 | 0.15 | 0.22 | 0.28 | 0 | 0.15 | 0.23 | 0.31 | | | | |
| 0.3 | 0.18 | 0.22 | 0.27 | 0.33 | 0.17 | 0.23 | 0.28 | 0.35 | | | | |
| 0.5 | 0.28 | 0.33 | 0.39 | 0.41 | 0.30 | 0.33 | 0.35 | 0.42 | | | | |

[a] For each type of coupling, we show the rmsd from a linear correlation of $J^{fit}(\phi,\psi)$ with no error vs. $J^{fit}(\phi,\psi)$ with errors as described in the text. The $(\phi,\psi)$ points used for the correlations were the same as those employed in the experimental correlations in Table 3. For each coupling $\sigma(\phi,\psi)$ values from left to right are: 0°, 10°, 15° and 20°. The errors in the dihedral angles (horizontal) and measured couplings (vertical) were derived numerically by using pseudorandom numbers in Mathematica (Wolfram, 1991).

nonsymmetric. In Fig. 13 we show an estimate of the error in θ, where we add the error if cos(θ) is less than 0 and subtract the error if cos(θ) is greater than 0 (in order to prevent cos(θ) from getting larger than 1 or smaller than −1).

To be as conservative as possible about the error in the dihedral angle, we assumed that there were no errors in any bond lengths or bond angles and that the only error in defining cos(θ) was in the distance between the terminal atoms A and D. Figure 13 shows the error in the dihedral angle with 0.05, 0.10, 0.15, 0.20 and 0.25 Å errors in the distance from A to D. It is clear that even a 0.10-Å error in distance (smaller than any of the reported errors from protein structures used in this study) leads to a minimum error in the dihedral angle of about 10°, and as much as 30–40° near the coplanar syn or anti conformations. The errors reported for the proteins used in this study would generate minimum errors of at least 20°. It must be stressed that these errors are not necessarily related to interpreting the electron density map; in tracing electron density, crystallographers have the benefit of using side chains and other heavy atoms in addition to the four atoms defining a dihedral angle. However, for the purpose of calculating dihedral angles from the interatomic distances of four atoms, we feel that the errors shown in Fig. 13 are relevant and representative.

Errors in the measurement of NMR coupling constants are highly dependent on the type of coupling, the type of experiment used to measure the coupling, signal-to-noise ratios, and signal overlap. The errors reported (see Table 1 for experimental references) for the particular couplings used here were: $^3J_{H^NH^\alpha}$: 0.5–1.0 Hz; $^1J_{C^\alpha H^\alpha}$: 0.5–1.0 Hz; $^2J_{C'H^\alpha}$: 1.0 Hz; and $^1J_{C^\alpha N}$ and $^2J_{C^\alpha N}$: 0.3 Hz. Recently, the effects of interference between spin–spin couplings and cross relaxation have been analyzed (Harbison, 1993). In this work, Harbison estimated that for correlation times typical of biological macromolecules, the observed coupling constants might have errors of 25% or more,

independent of methods used for their measurement. These effects have not been analyzed in our work, and, as a result, we may be underestimating the experimental errors.

To analyze the effect of the dihedral angle and coupling constant errors, we made correlations, between our theoretical functions with no errors and with errors, respectively, of the form

$$J' = J(\phi + \sigma_a, \psi + \sigma_a) + \sigma_m \tag{8}$$

where $\sigma_m$ and $\sigma_a$ are pseudorandom errors within prescribed ranges for the measured coupling and dihedral angles, respectively. The results are shown in Table 4.

On comparing Tables 3 and 4, it is clear for $^3J_{H^NH^\alpha}$ and $^1J_{C^\alpha N}$ that the rmsd we obtain for the estimated experimental errors is practically identical to the rmsd of the correlation with experimental values. Therefore, we conclude that the models for $^3J_{H^NH^\alpha}$ and $^1J_{C^\alpha N}$ are performing adequately within the errors of experiment. The small discrepancy of about 0.3 Hz between the rmsd of Tables 3 and 4 for $^2J_{C^\alpha N}$ can be attributed to minor deficiencies of the theoretical model. The much larger differences observed with $^1J_{C^\alpha H^\alpha}$ and $^2J_{C'H^\alpha}$, on the other hand, must be attributed to rather gross errors in the theoretical calculations or to some other systematic error, and these couplings must thus be ascribed less weight in predicting protein conformations. It is interesting to note that these unusually large differences are no larger than the reported solvent, pH and electric field dependencies of $^1J_{C^\alpha H^\alpha}$ and $^2J_{C'H^\alpha}$ (Barfield and Johnston Jr., 1972; Hansen, 1981).

To address possible real differences between the X-ray and NMR structures, we note the results shown in Fig. 7. For $^3J_{H^NH^\alpha}$, both the Karplus equation and the ab initio calculations exhibit very similar deviations as a function of amino acid sequence. Since these two theoretical models are independent* of one another, we suggest that regions that show such deviations represent parts of the molecules whose solution and crystal structures are different. It is worth noting that these regions correspond to those that have large isotropic B-factors, crystal contacts, or consist of extended loops. At present we are unable to determine whether these differences result from different static conformations or different degrees of mobility.

## CONCLUSIONS

We have provided qualitative or semiquantitative functions for the $\phi$- and $\psi$-dependence of $^3J_{H^NH^\alpha}$, $^1J_{C^\alpha N}$ and $^2J_{C^\alpha N}$ and somewhat more approximate functions for $^1J_{C^\alpha H^\alpha}$ and $^2J_{C'H^\alpha}$. We found no useful experimental correlations with calculated values of $^1J_{C'N}$, owing to its complex dependence on $\phi_i$, $\psi_i$, $\phi_{i-1}$ and $\psi_{i-1}$. Our analysis of the errors in dihedral angles from crystal structures indicates that uncertainties of more than 20° may be common. These errors, in association with errors in measurements, and possible systematic errors in the measured couplings (Harbison, 1993), suggest the serious limitations in attempting to empirically adjust any theoreti-

*Both the Karplus equation and the ab initio calculations were fitted to the experimental data. However, the ab initio results have only a linear adjustment, so all the curvature in the function results entirely from the calculations. The Karplus equation, on the other hand, assumes a trigonometric form of the coupling and thus has *all* of its curvature derived from the experimental data.

cal model to experimental X-ray and NMR data. In light of the numerous theoretical and empirical uncertainties, we reiterate with approval the warning of Karplus (1963): 'Certainly with our present knowledge, the person who attempts to estimate dihedral angles to an accuracy of one or two degrees does so at his own peril'.

## ACKNOWLEDGEMENTS

We thank Drs. Milo Westler and Željko Džakula for helpful discussions. Professors Ivan Rayment, Hazel Holden and Wayne Hendrickson provided helpful comments about errors associated with X-ray structures. Drs. Geerten Vuister and Ad Bax kindly supplied us with unpublished results, and Dr. Gerard Harbison provided a preprint of his work. A.S.E. is a trainee of an NIH Molecular Biophysics Training Grant (GM 08293). J.L.M. has support from NIH Grant RR 02301. Ab initio calculations were performed on an FPS model 522 computer with funds from NIH Grant RR 05659. The ab initio software was Spartan, distributed by Wavefunction, Inc., and kindly made available by Prof. W.J. Hehre; and GAMESS (General Atomic and Molecular Electronic Structure System), kindly provided by Dr. Mike Schmidt.

## REFERENCES

Aue, W.P., Bartholdi, E. and Ernst, R.R. (1976) *J. Chem. Phys.*, **64**, 2229–2246.
Babu, Y.S., Bugg, C.E. and Cook, W.J. (1988) *J. Mol. Biol.*, **204**, 191–204.
Barfield, M. and Grant, D.M. (1965) *Adv. Magn. Reson.*, **1**, 149–193.
Barfield, M. and Johnston Jr., M.D. (1972) *Chem. Rev.*, **73**, 53–73.
Barfield, M., Hruby, V.J. and Meraldi, J.-P. (1976) *J. Am. Chem. Soc.*, **98**, 1308–1314.
Barfield, M. and Smith, W.B. (1992) *J. Am. Chem Soc.*, **114**, 1574–1581.
Bax, A., Max, D. and Zax, D. (1992) *J. Am. Chem. Soc.*, **114**, 6923–6925.
Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131–138.
Blake, P.R., Lee, B., Summers, M.F., Adams, M.W.W., Park, J.-B., Zhou, Z.H. and Bax, A. (1992) *J Biomol. NMR*, **2**, 527–533.
Bystrov, V.F. (1976) *Prog. NMR Spectrosc.*, **10**, 44–81.
Chary, K.V., Otting, G. and Wüthrich, K. (1991) *J. Magn. Reson.*, **93**, 218–224.
Creighton, T.E. (1984) *Proteins: Structures and Molecular Properties*, Freeman, New York, NY.
Delaglio, F., Torchia, D.A. and Bax, A. (1991) *J. Biomol. NMR*, **1**, 439–446.
Deisenhofer, J. and Steigemann, W. (1975) *Acta Crystallogr.*, **B31**, 238–250.
Edison, A.S., Westler, W.M. and Markley, J.L. (1991) *J. Magn. Reson*, **92**, 434–438.
Edison, A.S., Markley, J.L. and Weinhold, F. (1993) *J. Phys. Chem.*, **97**, 11657–11665.
Edison, A.S., Weinhold, F., Westler, W.M. and Markley, J.L. (1994) *J. Biomol. NMR*, **4**, 543–551.
Egli, H. and Von Philipsborn, W. (1981) *Helv. Chim. Acta*, **64**, 976–988.
Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, New York, NY.
Foster, J.P. and Weinhold, F. (1980) *J. Am. Chem. Soc.*, **102**, 7211–7218.
Gil, V.M.S. and Von Philipsborn, W. (1989) *Magn. Reson. Chem.*, **27**, 409–430.
Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1986) *J. Chem. Phys.*, **85**, 6837–6843.
Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1989) *J. Magn Reson.*, **84**, 14–63.
Griesinger, C. and Eggenberger, U. (1992) *J. Magn. Reson.*, **97**, 426–434.
Haasnoot, C.A.G., De Leeuw, F.A.A.M. and Altona, C. (1980) *Tetrahedron*, **36**, 2783–2792.
Hansen, P.E., Feeney, J. and Roberts, G.C.K. (1975) *J. Magn. Reson.*, **17**, 249–261.
Hansen, P.E. (1981) *Prog. NMR Spectrosc.*, **14**, 175–296.
Harbison, G.S. (1993) *J. Am. Chem. Soc.*, **115**, 3026–3027.

Head-Gordon, T., Head-Gordon, M., Frisch, M.J., Brooks III, C.L. and Pople, J.A. (1991) *J. Am. Chem. Soc.*, **113**, 5989–5997.
Hehre, W.J., Radom, L., Schleyer, P.v.R. and Pople, J.A. (1986) *Ab Initio Molecular Orbital Theory*, Wiley, New York, NY.
Jameson, C.J. and Damasco, M.C. (1970) *Mol. Phys.*, **18**, 491–504.
Karplus, M. (1959) *J. Chem. Phys.*, **30**, 11–15.
Karplus, M. (1963) *J. Am. Chem. Soc.*, **85**, 2870–2871.
Kay, L.E., Brooks, B., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *J. Am. Chem. Soc.*, **111**, 5488–5490.
Kowalewski, J., Laaksonen, A., Roos, B. and Siegbahn, P.J. (1979) *J. Chem. Phys.*, **71**, 2896–2902.
Kowalewski, J. (1982) *Annu. Rev. NMR Spectrosc.*, **12**, 81–176.
Laaksonen, A. and Kowalewski, J. (1981) *J. Am. Chem. Soc.*, **103**, 5277–5283.
Lee, W.S. and Schulman, J.M. (1979) *J. Am. Chem. Soc.*, **101**, 3182–3184.
Levy, G.C. and Lichter, R.L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance Spectroscopy*, Wiley, New York, NY.
Loll, P.J. and Lattman, E.E. (1989) *Protein Struct. Funct. Genet.*, **5**, 183–201.
Ludvigsen, S., Andersen, K.V. and Poulsen, F.M. (1991) *J. Mol. Biol.*, **217**, 731–736.
Madsen, J.C., Sørensen, O.W., Sørensen, P. and Poulsen, F. (1993) *J. Biomol. NMR*, **3**, 239–244.
Markley, J.L. and Kainosho, M. (1993) In *NMR of Biological Molecules* (Ed., Roberts, G.C.K.) Practical Approach Series, Oxford Press, New York, NY, pp. 101–152.
Marshall, J.L., Walter, S.R., Barfield, M., Marchand, A.P., Marchand, N.W. and Segre, A.L. (1976) *Tetrahedron*, **32**, 537–542.
McPhalen, C.A. and James, M.N.G. (1987) *Biochemistry*, **26**, 261–269.
Mierke, D.F., Grdadolnik, S.G. and Kessler, H. (1992) *J. Am. Chem. Soc.*, **114**, 8283–8284.
Mohanakrishnan, P. and Easwaran, K.R.K. (1979) *Org. Magn. Reson.*, **12**, 196–198.
Montelione, G.T., Winkler, M.E., Rauenbuhler, P. and Wagner, G. (1989) *J. Magn. Reson.*, **82**, 198–204.
Neuhaus, D., Wagner, G., Vasák, M., Kägi, J.H.R. and Wüthrich, K. (1984) *Eur. J. Biochem.*, **143**, 659–667.
Pachler, K.G.R. (1972) *J. Chem. Soc. Perkin Trans. II*, 1936–1940.
Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, **180**, 741–751.
Pople, J.A. and Santry, D.P. (1964) *Mol. Phys.*, **8**, 1–18.
Pople, J.A. and Bothner-By, A.A. (1965) *J. Chem. Phys.*, **42**, 1339–1349.
Ramsey, N.F. (1953) *Phys. Rev.*, **91**, 303–307.
Reed, A.E., Weinstock, R.B. and Weinhold, F. (1985) *J. Chem. Phys.*, **83**, 735–746.
Reed, A.E., Curtiss, L.A. and Weinhold, F. (1988) *Chem. Rev.*, **88**, 899–926.
Reed, A.E. and Weinhold, F. (1991) *Isr. J. Chem.*, **31**, 277–285.
Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Jensen, J.H., Koseki, S., Gordon, M.S., Nguyen, K.A., Windus, T.L. and Elbert, S.T. (1990) *QCPE Bull.*, **10**, 52–54.
Vuister, G.W. and Bax, A. (1992) *J. Biomol. NMR*, **2**, 401–405.
Vuister, G.W., Delaglio, F. and Bax, A. (1992) *J. Am. Chem. Soc.*, **114**, 9674–9675.
Vuister, G.W., Delaglio, F. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 67–80.
Wagner, G., Hyberts, S.G. and Havel, T.F. (1992) *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 167–198.
Walter, J. and Huber, R. (1983) *J. Mol. Biol.*, **167**, 911–917.
Wider, G., Neri, D., Otting, G. and Wüthrich, K. (1989) *J. Magn. Reson.*, **85**, 426–431.
Wlodawer, A., Walter, J., Huber, R. and Sjölin, L. (1984) *J. Mol. Biol.*, **180**, 301–329.
Wolfram, S. (1991) *Mathematica: A System for Doing Mathematics by Computer*, 2nd ed., Addison-Wesley, Redwood City, CA.