Journal of Biomolecular NMR, 4 (1994) 341–348
ESCOM

J-Bio NMR 182

Correlation between ^{15}N NMR chemical shifts in proteins and secondary structure

Hongbiao Le and Eric Oldfield*

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

Received 7 June 1993
Accepted 10 November 1993

Keywords: Secondary structure; Shielding surface; ^{15}N shielding; Torsion angle

SUMMARY

An empirical correlation between the peptide ^{15}N chemical shift, $\delta^{15}N_i$, and the backbone torsion angles ϕ_i, ψ_{i-1} is reported. By using two-dimensional shielding surfaces $\Delta(\phi_i,\psi_{i-1})$, it is possible in many cases to make reasonably accurate predictions of ^{15}N chemical shifts for a given structure. On average, the rms error between experiment and prediction is about 3.5 ppm. Results for threonine, valine and isoleucine are worse (~4.8 ppm), due presumably to χ_1-distribution/γ-gauche effects. The rms errors for the other amino acids are ~3 ppm, for a typical maximal chemical shift range of ~15–20 ppm. Thus, there is a significant correlation between ^{15}N chemical shift and secondary structure.

INTRODUCTION

It is well known that folding a protein into its native conformation causes a large chemical shift dispersion, and these shielding nonequivalences permit application of multidimensional NMR spectroscopy for determining the 3D structures of proteins in solution (Wüthrich, 1986; Clore and Gronenborn, 1991). However, the origins of these folding-induced nonequivalencies have, for most elements, been unclear until recently (de Dios et al., 1993). The reason for this is that there are many parameters which might influence the chemical shift — such as the values of the ϕ,ψ,χ torsion angles, electrostatics (including hydrogen bonding), solvation, dynamics, and so forth. Thus, the chemical shift problem has received little attention, being generally regarded as intractable.

However, more recent observations of strong secondary structure — chemical shift correlations for H^α, C^α and C^β (Case and Ösapay, 1991; Spera and Bax, 1991; Williamson et al., 1992),

*To whom correspondence should be addressed.

combined with improvements in hardware and software architecture (Wolinski et al., 1990), have reopened interest in analyzing the burgeoning protein chemical shift database — which currently consists of ~130 000 reported values (Seavey et al., 1991). These developments have led us to apply Hartree–Fock theory to predicting chemical shifts (of ^{1}H, ^{13}C, ^{15}N and ^{19}F) in proteins (de Dios et al., 1993; Pearson et al., 1993), and we have shown that very good agreement with experimental results can now be obtained. The next step is to begin to use chemical shifts to refine and predict structure. Good progress in this direction has been made recently by using ^{1}H NMR (Wishart et al., 1992; Ösapay and Case, 1993), where clear shift–structure correlations can be seen — much as with ^{13}C NMR (Spera and Bax, 1991). However, for ^{15}N chemical shifts, much less progress has been reported. The reasons for this are clear, based upon our ab initio work (de Dios et al., 1993): there are complex ϕ,ψ,χ contributions to the shielding, and solvation and electrostatic field effects can all be significant.

Nevertheless, in recent work (de Dios et al., 1993) it was apparent that the 'local fragment' or secondary structure appeared to play the single most significant role in determining shielding, and in earlier work Glushka et al. (1989) noticed an interesting empirical correlation between $\delta^{15}N_i$ and ψ_{i-1}, but only in β-sheet regions. Based upon our ab initio results and the observations of Glushka et al. (1989) and Spera and Bax (1991), we felt that it would be more likely that $\delta^{15}N_i$ is a function of at least two independent variables — specifically ϕ_i, ψ_{i-1}. We explore this hypothesis in this paper and show that ^{15}N chemical shifts can be predicted from known structure with an rms error of about 3.5 ppm by use of 2D shielding surfaces. In future work with theoretical shielding surfaces, it can be anticipated that this type of (ϕ,ψ) information will be of use in structure refinement and determination.

COMPUTATIONAL ASPECTS

The database we used consists of ^{15}N amide chemical shifts for the following 14 proteins: bovine pancreatic trypsin inhibitor (Wlodawer et al., 1987; Glushka et al., 1989); calmodulin (Babu et al., 1988; Ikura et al., 1990); *Escherichia coli* IIIGlc (Pelton et al., 1991; Worthylake et al., 1991); *Bacillus subtilis* IIGlc domain A (Liao et al., 1991; Fairbrother et al., 1992); *E. coli* ribonuclease H (Yamazaki et al., 1991; Katayanagi et al., 1992); Staphylococcal nuclease (Loll and Lattman, 1989; Wang et al., 1992); interleukin-1β (Finzel et al., 1989; Driscoll et al., 1990); *Pseudomonas* cytochrome c-551 (Matsuura et al., 1982; Timkovich, 1990); calbindin D9k (Skelton et al., 1992; Svensson et al., 1992); FK506-binding protein (Van Duyne et al., 1991; Xu et al., 1993); dihydrofolate reductase (Bolin et al., 1982; Carr et al., 1991); bacteriophage T4 lysozyme (McIntosh et al., 1990; Matsumura et al., 1993); *Pseudomonas aeruginosa* azurin (Nar et al., 1991; Van de Kamp et al., 1992) and *Aspergillus oryzae* ribonuclease T1 pH_{iso} 3.9 (Martinez-Oyanedel et al., 1991; Schmidt et al., 1991). We used X-ray coordinates reported in the Brookhaven Protein Data Bank (Bernstein et al., 1977; Abola et al., 1987) for all proteins (files 5PTI, 3CLN, 1GPR, 1F3G, 2RN2, 1SNC, 1I1B, 1OFV, 451C, 4ICB, 1FKF, 3DFR, 3LZM, 2AZU, 9RNT). The respective reported resolutions for these proteins are: 2.0, 2.2, 2.1, 1.9, 1.48, 1.65, 2.0, 1.7, 1.7, 1.9, 1.7, 1.7, 1.6, 1.6, 1.6 and 1.5 Å. The 15 structures comprised a total of 1654 residues, and no attempt was made to modify the structures. The function used to fit the $\delta(\phi_i,\psi_{i-1})$ surface is similar to that used by Spera and Bax (1991); the resulting surface is a convolution of chemical shifts and a Gaussian:

$$\Delta(\phi,\psi-) = \frac{\sum_i \delta(\phi_i\psi_i^-)\cdot\exp\left\{-\left[\sin^2\left(\frac{\phi_i-\phi}{2}\right)+\sin^2\left(\frac{\psi_i^- - \psi^-}{2}\right)\right]\Big/0.03\right\}}{\sum_i \exp\left\{-\left[\sin^2\left(\frac{\phi_i-\phi}{2}\right)+\sin^2\left(\frac{\psi_i^- - \psi^-}{2}\right)\right]\Big/0.03\right\}} \quad (1)$$

where $\phi,\psi-$ represents $\phi(i)$, $\psi(i-1)$; $\delta(\phi,\psi-)$ is the secondary shift ($\Delta\delta$) and $\Delta(\phi,\psi-)$ represents the computed surface. We computed such shielding surfaces for all amino acids together (a 'global surface'), as well as individual amino acid chemical shift surfaces for most residues. For prediction of a given chemical shift δ_i, or a set of chemical shifts $\{\delta_i\}$, data for residue i were not used in surface computation. Also, predictions were not made for data points in sparse regions (Spera and Bax, 1991).

RESULTS AND DISCUSSION

Figure 1 shows secondary chemical shift contour plots, $\Delta\delta^{15}N_i$, on a ϕ_i, ψ_{i-1} surface. Incorporated in Fig. 1 is the entire set of 1477 ^{15}N chemical shift data points, corrected for random coil shifts by using either the values reported by Wishart et al. (1991), Fig. 1A, or those by Glushka et al. (1990), Fig. 1B. The qualitative features of the two surfaces are very similar. For example, in the

TABLE 1
ROOT MEAN SQUARE DEVIATIONS BETWEEN PREDICTED ^{15}N CHEMICAL SHIFTS AND EXPERIMENT USING EMPIRICAL SHIELDING SURFACES

		Rmsd (ppm)		
Residue type	Number of residues	Global surface[a]	Global surface[b]	Individual surface
Ala	115	3.04	3.19	2.97
Arg	65	2.87	4.59	2.88
Asn	74	3.97	4.75	3.90
Asp	105	2.88	2.99	2.73
Gln	68	3.32	3.76	3.43
Glu	129	2.81	2.74	2.66
Gly	139	5.20	6.18	3.07
Ile	86	4.86	4.43	4.31
Leu	124	2.94	2.82	2.95
Lys	125	3.11	4.30	2.70
Phe	76	3.30	3.37	3.52
Ser	92	3.32	3.44	3.56
Thr	102	5.23	7.15	5.61
Tyr	48	4.04	4.09	3.43[c]
Val	129	4.57	4.92	4.34

[a] Using random coil values of Wishart et al. (1991).
[b] Using random coil values of Glushka et al. (1990).
[c] Combined surface of phenylalanine, tyrosine residue data.

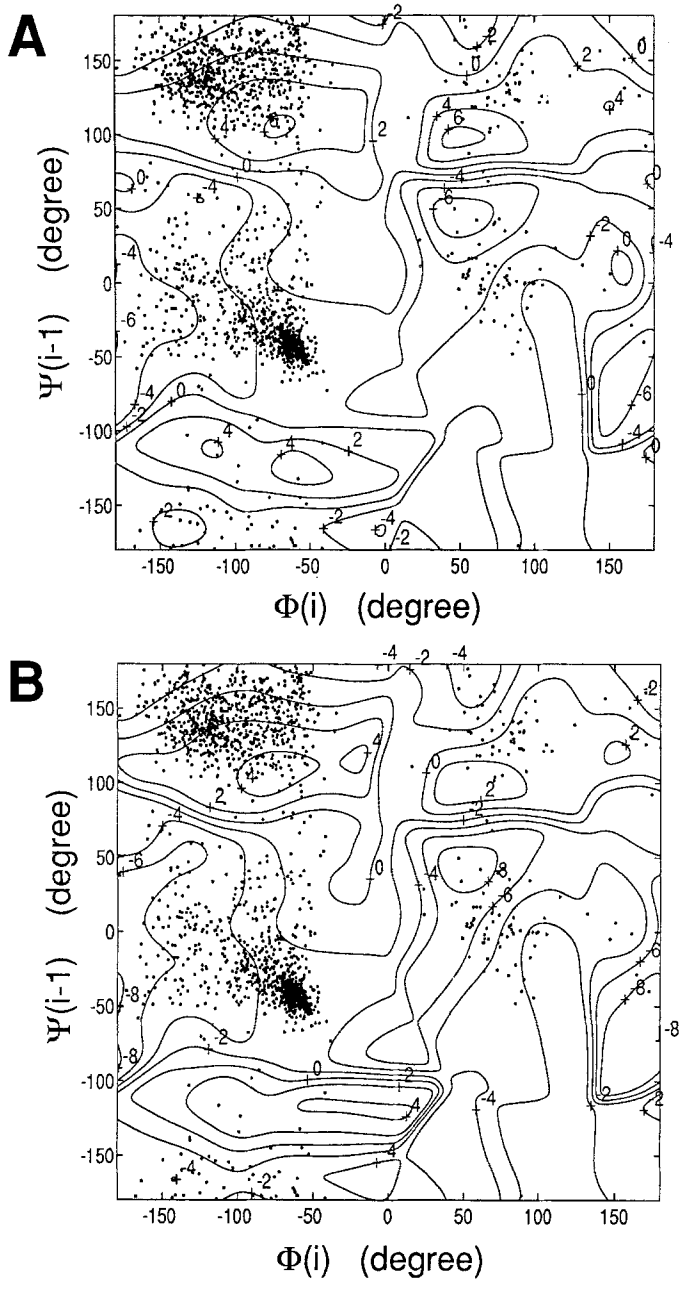


Fig. 1. Secondary ^{15}N chemical shift surfaces for all 1477 residues. (A) Random coil shifts of Wishart et al. (1991) used for calculating secondary shifts; rmsd = 3.5 ppm. (B) Random coil shifts of Glushka et al. (1990) used for calculating secondary shifts; rmsd = 3.9 ppm.

area which corresponds to the β-sheet region in a Ramachandran plot, we find large downfield shifts ($\Delta\delta \sim 4$ ppm). This secondary shift diminishes on approaching the upper left corner, or on

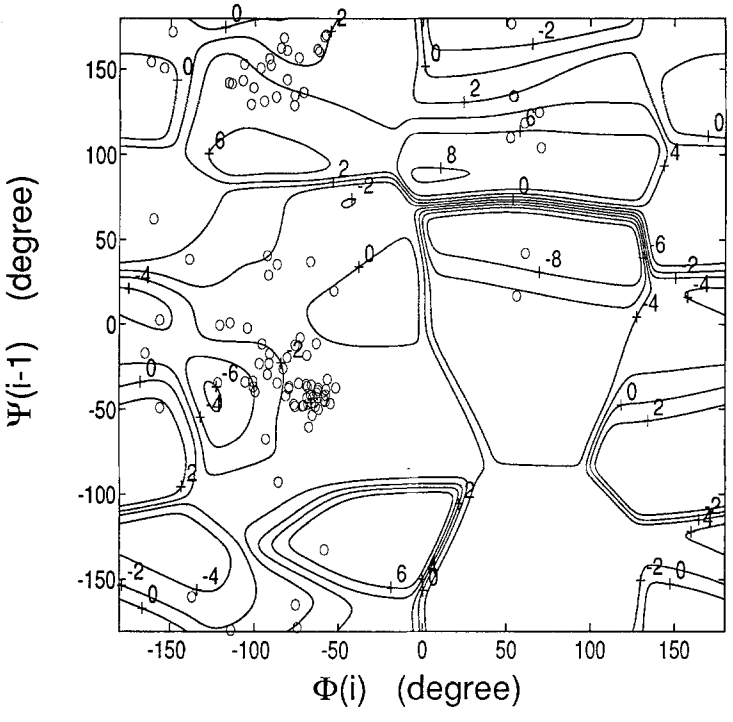


Fig. 2. Secondary ^{15}N chemical shift surface for aspartic acid residues. Random coil values of Wishart et al. (1991) were used for calculating secondary shifts; rmsd = 2.7 ppm.

going down the plot. In the 'helix region', there is a large upfield shift ($\Delta\delta \approx -5$ ppm), and this shift tends to be even more upfield in the 'turn area'.

Table 1 shows rms deviations between experimental ^{15}N chemical shifts and those predicted from the X-ray structures for a number of amino acid residue types. We first computed ^{15}N $\Delta\delta$ values for each type of amino acid, using modified global surfaces on which residues of the type of interest were deleted, to avoid problems with bias of the data set. We used random coil shifts of both Wishart et al. (1991) and Glushka et al. (1990), and found the rmsd values shown in Table 1. In most cases, there are only small differences in rms error between the two surfaces, although for threonine and arginine the difference is almost 2 ppm.

Better chemical shift predictions are obtained when individual amino acid shift surfaces are computed. Figure 2 shows such a surface for aspartic acid, while Fig. 3 shows experimental versus predicted secondary shifts from individual surfaces for aspartic acid (rms error = 2.7 ppm), lysine (rms error = 2.7 ppm), and valine (rms error = 4.3 ppm), and rmsd results for each amino acid are given in Table 1. The correlations in many cases are quite good. The average rmsd when using individual surfaces is 3.5 ppm, although, as shown in Table 1 and Fig. 3C, the results for three residues, i.e., threonine (rmsd = 5.6 ppm), valine (rmsd = 4.3 ppm) and isoleucine (rmsd = 4.3 ppm) stand out as being much worse than the others. Eliminating these three residues, the average rmsd is about 3 ppm, over a $\approx$20 ppm overall range of secondary shifts.

We believe that the cause of the significantly worse rmsd for threonine, valine and isoleucine arises from the presence of a strong (and variable) χ^1 contribution to shielding. All three of these

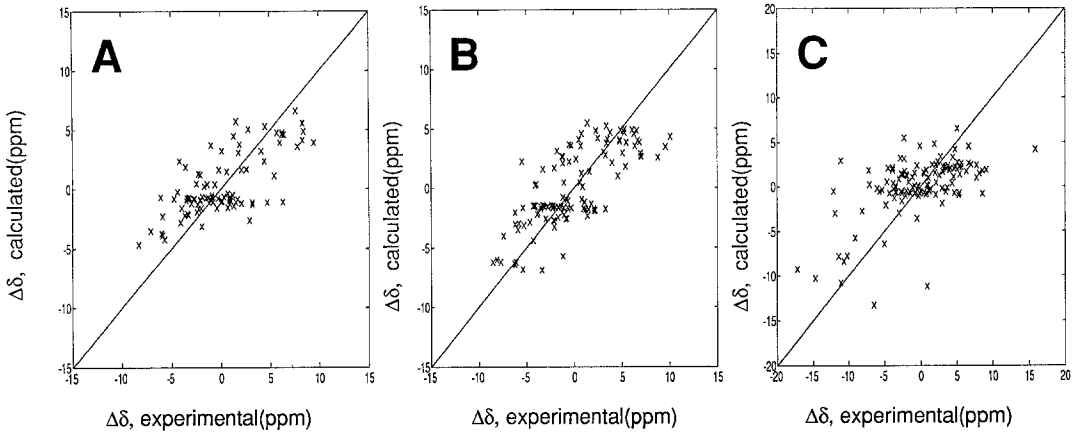
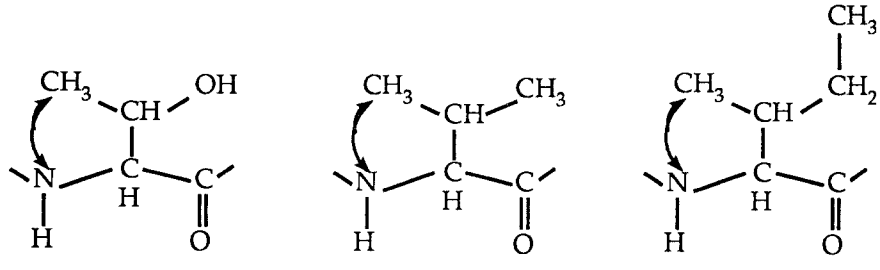


Fig. 3. Correlation between experimental secondary ^{15}N chemical shifts and those computed from individual secondary chemical shift surfaces. (A) aspartic acid; (B) lysine; and (C) valine. The rms errors are 2.7, 2.7 and 4.3 ppm, respectively (Table 1).

amino acids have methyl group substituents on the β-carbon, and χ^1 torsional or 'γ-gauche' effects (Tonelli, 1980) are presumably contributing to $\delta^{15}N$, as shown in the scheme below.



Thr Val Ile

In previous work on ^{15}N shielding in a valine fragment (de Dios et al., 1993), a 5-ppm difference in shielding was computed between the two most common χ^1 conformers. Thus, if (as is certainly the case) there are distributions of χ^1 conformers for these three residues, then the increased rmsd can be readily appreciated, and in future work with a larger database it should be possible to make separate χ^1 maps for each substate, or to compute such surfaces by using ab initio methods (de Dios et al., 1993), thereby decreasing the rms deviations for the three residues.

CONCLUSIONS

In conclusion, the results we have presented above show a correlation between backbone amide ^{15}N chemical shifts and local structure, which appears to make the major contribution to ^{15}N chemical shift nonequivalencies observed in native proteins. In particular, our results show a strong secondary chemical shift correlation with (ϕ_i, ψ_{i-1}), which could be of use as an additional indicator of secondary structure. On average, the rmsd between experiment and prediction is 3.5

ppm. For amino acids that do not have a γ-methyl interaction (i.e. a large χ^1 conformer contribution to shielding), the rmsd is typically ~3 ppm, for an overall experimental shift range of ~15–20 ppm. These results can be expected to be improved upon by incorporation of additional data points, by separating data into χ^1 and possibly χ^1, χ^2 subsets, as well as by editing the data sets for sequence (e.g. i−1 glycine) anomalies, and by considering, for example, the effects of hydrogen bonding.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (grant HL-19481). H.L. is a graduate fellow of the University of Illinois.

REFERENCES

Abola, E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J. (1987) In *Crystallographic Databases: Information Content, Software Systems, Scientific Applications* (Eds, Allen, F.H., Bergerhoff, G. and Sievers, R.) Data Commission of the International Union of Crystallography, Cambridge, pp. 107–132.

Babu, Y.S., Bugg, C.E. and Cook, W.J. (1988) *J. Mol. Biol.*, **204**, 191–204.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Bolin, J.T., Filman, D.J., Mathews, D.A., Hamlin, R.C. and Kraut, J. (1982) *J. Biol. Chem.*, **257**, 13650–13662.

Carr, M.D., Birdsall, B., Frenkiel, T.A., Baver, C.J., Jimenez-Barbero, J., Polshakov, V.I., McCormick, J.E., Roberts, G.C.K. and Feeney, J. (1991) *Biochemistry*, **30**, 6330–6341.

Case, D.A. and Ösapay, K. (1991) *J. Am. Chem. Soc.*, **113**, 9436–9444.

Clore, G.M. and Gronenborn, A.M. (1991) *Science*, **252**, 1390–1399.

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

Driscoll, P.C., Clore, G.M., Marion, D., Wingfield, P.T. and Gronenborn, A.M. (1990) *Biochemistry*, **29**, 3542–3556.

Fairbrother, W.J., Palmer III, A.G., Rance, M., Reizer, J., Saier Jr., M.H. and Wright, P.E. (1992) *Biochemistry*, **31**, 4413–4425.

Finzel, B.C., Clancy, L.L., Holland, D.R., Muchmore, S.W., Watenpaugh, K.D. and Einspahr, H.M. (1989) *J. Mol. Biol.*, **209**, 779–791.

Glushka, J., Lee, M., Coffin, S. and Cowburn, D. (1989) *J. Am. Chem. Soc.*, **111**, 7716–7722.

Glushka, J., Lee, M., Coffin, S. and Cowburn, D. (1990) *J. Am. Chem. Soc.*, **112**, 2843.

Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T. and Morikawa, K. (1992) *J. Mol. Biol.*, **223**, 1029–1052.

Liao, D.-I., Kapadia, G., Reddy, P., Saier Jr., M.H., Reizer, J. and Herzberg, O. (1991) *Biochemistry*, **30**, 9583–9594.

Loll, P.J. and Lattman, E.E. (1989) *Proteins*, **5**, 183–201.

Martinez-Oyanedel, J., Choe, H.-W., Heinemann, W. and Saenger, W. (1991) *J. Mol. Biol.*, **222**, 335–352.

Matsumura, M., Wozniak, J.A., Dao-Pin, S. and Mathews, B.W. (1993) Brookhaven Protein Data Bank, file 3LZM.

Matsuura, Y., Takano, T. and Dickerson, R.E. (1982) *J. Mol. Biol.*, **156**, 389–409.

McIntosh, L., Wand, A.J., Lowry, D.F., Redfield, A.G. and Dahlquist, F.W. (1990) *Biochemistry*, **29**, 6341–6362.

Nar, H., Messerschmidt, A., Huber, R., Van de Kamp, M. and Canters, G.W. (1991) *J. Mol. Biol.*, **257**, 13650–13662.

Ösapay, K. and Case, D.A. (1993) In *Calculation of NMR Shielding Constants and Their Use in the Determination of the Geometric and Electronic Structures of Molecules and Solids* (Ed, Tossel, J.) Kluwer, Dordrecht, p. 572.

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

Pelton, J.G., Torchia, D.A., Meadow, N.D., Wong, C.-Y. and Roseman, S. (1991) *Biochemistry*, **30**, 10043–10057.

Schmidt, J.M., Thüring, H., Werner, A., Rüterjans, H., Quaas, R. and Hahn, U. (1991) *Eur. J. Biochem.*, **197**, 643–653.

Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J.L. (1991) *J. Biomol. NMR*, **1**, 217–236.

Skelton, N.J., Akke, M., Kördel, J., Thulin, E., Forsen, S. and Chazin, W.J. (1992) *FEBS Lett.*, **303**, 136–140.

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.
Svensson, L.A., Thulin, E. and Forsen, S. (1992) *J. Mol. Biol.*, **223**, 601–606.
Timkovich, R. (1990) *Biochemistry*, **29**, 7773–7780.
Tonelli, A.E. (1980) *J. Am. Chem. Soc.*, **102**, 7635–7637.
Van de Kamp, M., Canters, G.W., Wijmenga, S.S., Lommen, A., Hilbers, C., Nar, H., Messerschmidt, A. and Huber, R. (1992) *Biochemistry*, **31**, 10194–10207.
Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clady, J. (1991) *Science*, **252**, 839–842.
Wang, J., Hinck, A.P., Loh, S.N., LeMaster, D.M. and Markley, J.L. (1992) *Biochemistry*, **31**, 921–936.
Williamson, M.P., Asakura, T., Nakamura, E. and Demura, M. (1992) *J. Biomol. NMR*, **2**, 83–98.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.
Wlodawer, A., Deisenhofer, J. and Huber, R. (1987) *J. Mol. Biol.*, **193**, 145–156.
Wolinski, K., Hinton, J.F. and Pulay, P. (1990) *J. Am. Chem. Soc.*, **112**, 8251–8260.
Worthylake, D., Meadow, N.D., Roseman, S., Liao, D.-I., Herzberg, O. and Remington, S.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10382–10386.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.
Xu, R.X., Nettesheim, D., Olejniczak, E.T., Meadows, R., Gemmecker, G. and Fesik, S.W. (1993) *Biopolymers*, **33**, 535–550.
Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H. and Nagayama, K. (1991) *Biochemistry*, **30**, 6036–6047.