Journal of Biomolecular NMR, 3 (1993) 607–612
ESCOM

J-Bio NMR 149

NMR chemical shifts and structure refinement in proteins

David D. Laws, Angel C. de Dios and Eric Oldfield*

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

Received 2 July 1993
Accepted 18 August 1993

Keywords: Chemical shift; Ab initio; Calmodulin; *Staphylococcal* nuclease

SUMMARY

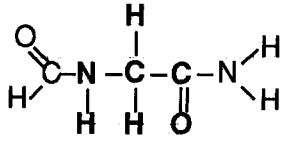
Computation of the $^{13}C^{\alpha}$ chemical shifts (or shieldings) of glycine, alanine and valine residues in bovine and *Drosophila* calmodulins and *Staphylococcal* nuclease, and comparison with experimental values, is reported using a gauge-including atomic orbital quantum-chemical approach. The full ~ 24 ppm shielding range is reproduced (overall r.m.s.d. = 1.4 ppm) using 'optimized' protein structures, corrected for bond-length/bond-angle errors, and rovibrational effects.

Multidimensional nuclear magnetic resonance (Oschkinat et al., 1988) provides a powerful route to analyzing the three-dimensional (3D) structures of proteins in solution (Bax, 1989), similar to that provided by X-ray diffraction studies of crystalline solids (Billeter et al., 1992). However, the origins of the chemical-shift nonequivalencies observed in proteins due to folding — without which NMR structural studies would not be possible — have been poorly understood, especially for the heavier elements. We show in this communication that the full ~ 24 ppm range of $^{13}C^{\alpha}$ chemical shifts in glycine, alanine and valine residues in two proteins, *Drosophila* calmodulin and *Staphylococcal* nuclease, can now be reproduced by using quantum-chemical methods (de Dios et al., 1993). However, good agreement (r.m.s.d. ~ 1.4 ppm) between theory and experiment is achieved only when highly relaxed structures are used, due to the extreme sensitivity of ^{13}C chemical shifts to bond-length errors. These findings should open up new avenues to structure refinement and determination, as well as providing a route for spectral assignment verification.

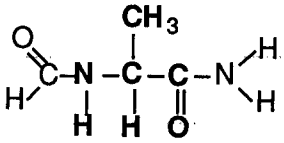
The C^{α} sites of glycine, alanine and valine residues in proteins are known to display a very large chemical-shift range, with glycine being most shielded, followed by alanine, while valine is most deshielded (Wishart et al., 1991). Analysis of glycine, alanine and valine C^{α} shieldings thus provides a stringent test of our ability to predict chemical shifts in proteins, because the shift range is so large — over four times the C^{α} range previously investigated (de Dios et al., 1993). We

*To whom correspondence should be addressed.

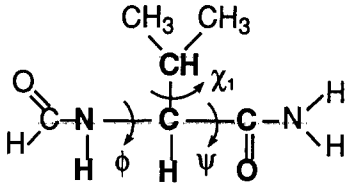
have used the following three 'model fragments' as representations of glycine-, alanine- and valine-containing peptides in proteins, since we felt they would successfully reproduce the main effects of ϕ,ψ (and χ) torsion angles in ab initio shielding calculations. Each fragment contains two amide groups:



Gly



Ala



Val

For the atoms shown in bold face we used a large basis set: 6-311++G(2d,2p), a triple zeta split-valence basis set with 2d, 2p polarization and additional diffuse functions, while for the other atoms we used the double-zeta basis set 6-31G (Krishnan et al., 1980), a locally dense approach (Chesnut and Moore, 1989). We used the charge field-perturbation gauge-including atomic orbital method (Ditchfield, 1972; Wolinski et al., 1990; de Dios and Oldfield, 1993) for shielding calculations, and incorporated charges of the remaining atoms in the protein using an electroneutral ENZYMIX charge set (Lee et al., 1993) for arginine, lysine, aspartate and glutamate, and an AMBER charge set (Weiner et al., 1986) for all other residues. Surface charges were not included, since their effects were expected to be negligible, based on the general observation that C^{α} shieldings in proteins are only weakly pH dependent, and surface charge fields are largely screened by solvent water.

Figure 1A shows experimental C^{α} chemical shifts (Ikura et al., 1990) versus computed chemical shieldings for glycine, alanine and valine residues in a recombinant *D. melanogaster* calmodulin (expressed in *Escherichia coli*). While there is some general agreement between theory and experiment, the 2.2-Å resolution of the X-ray data (Taylor et al., 1991) precludes accurate shielding calculations. However, a 1.7-Å resolution structure (bovine calmodulin expressed in *E. coli*; Chattopadhyaya et al., 1992) yields much better agreement, as shown in Fig. 1B. Thus, an accurate initial structure is essential for accurate shielding calculations. We then investigated a second protein, *S.* nuclease (SNase), and computed glycine, alanine and valine C^{α} shielding results are shown, together with calmodulin results, in Fig. 1C. There is evidence of a ~ 1–2 ppm bias between the two structures, which can be traced to a systematic increase in bond lengths in SNase.

Clearly, in order to compare results from different structures; it is necessary to relax individual structures, or energy minimize them, towards a uniform set of bond lengths (and bond angles) — taking into account necessary residual differences by use of the relevant shielding derivatives (manuscript in preparation). Figure 1D shows how structure relaxation (2000 steps of steepest descents with the AMBER force field, 5 Å solvent shell; Discover program, Biosym Technologies, San Diego, CA) greatly improves the slope (from −0.48 to −1.2) and correlation coefficient (from 0.69 to 0.97) for the alanine C^{α} sites in SNase, and similar improvements are seen when using either liganded (Loll and Lattman, 1989) or unliganded (Hynes and Fox, 1991) SNase structures. We then relaxed the bovine recombinant calmodulin structure and computed C^{α} shieldings for all

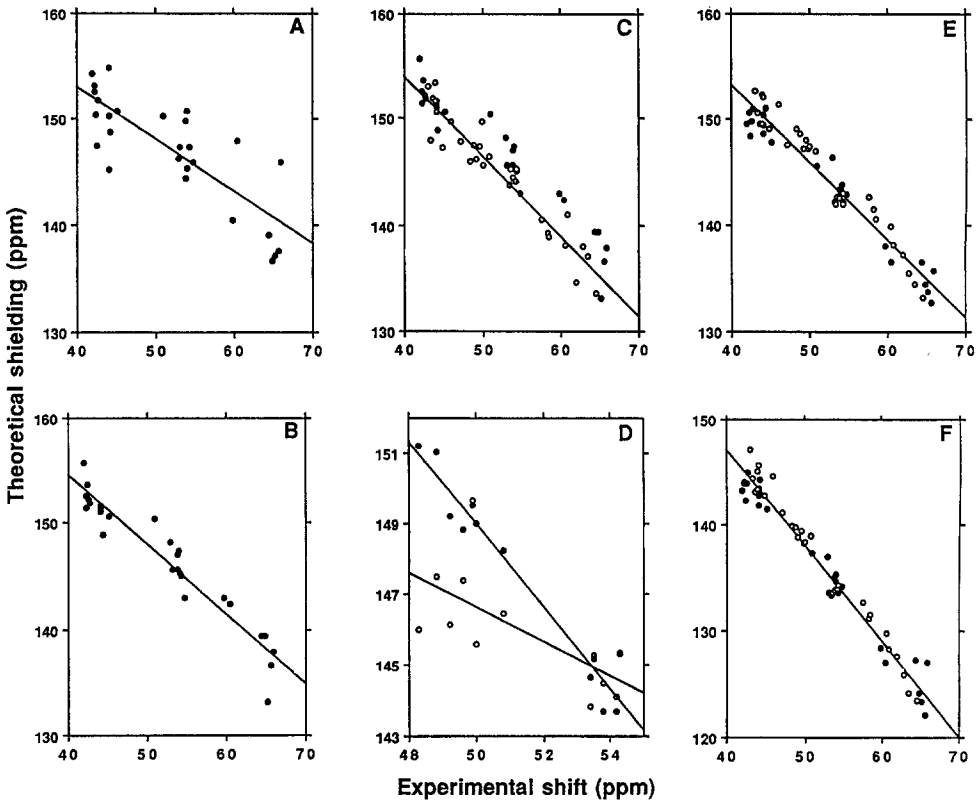


Fig. 1. Experimental chemical shifts vs computed chemical shieldings for C^{α} sites in glycine, alanine and valine residues in calmodulin and SNase. (A) recombinant *D. melanogaster* calmodulin NMR shifts (Ikura et al., 1990), 2.2-Å *D. melanogaster* X-ray structure of Taylor et al. (1991). (B) As in A, for the 1.7-Å recombinant bovine X-ray structure of Chattopadhyaya et al. (1992). (C) As in B (●) together with ligand-free SNase NMR shifts (Wang et al., 1992) of the 1.65-Å liganded SNase structure of Loll and Lattman (1989) (○). The 1.7-Å unliganded structure of Fox (Hynes and Fox, 1991) gives a similar correlation with experimental (liganded (Torchia et al., 1989; Wang et al., 1992) or unliganded (Wang et al., 1992)) shifts. There is a ~ 1–2 ppm bias towards increased computed shielding for all residues in calmodulin versus SNase, due to bias in bond lengths between the two X-ray structures. (D) Effect of structure relaxation on alanine C^{α} SNase shifts (Torchia et al., 1989). Unrelaxed structure (Loll and Lattman, 1989), ○, m = −0.48, R = 0.69; (●), 2000 steps of steepest descents energy minimization, AMBER force field, solvated structure from Fig. 1C. The slopes and correlation coefficients improve to m = −1.2, R = 0.97, on relaxation. (E) Effects of structure relaxation (> 10^4 steps) on glycine, alanine and valine C^{α} shielding in calmodulin (Ikura et al., 1990; Chattopadhyaya et al., 1992) (○) and SNase (Loll and Lattman, 1989; Wang et al., 1992) (●); overall m = −0.73, R = 0.97, r.m.s.d. = 1.9 ppm. (F) Effects of structure relaxation, bond length (In et al., 1987; Glowka, 1988; Bhandary and Kopple, 1991) and rovibrational corrections on glycine, alanine and valine C^{α} shifts in calmodulin and SNase; overall m = −0.91, R = 0.98, r.m.s.d. = 1.4 ppm (57 data points). Calculations were carried out in our laboratory by using a cluster of RISC computers (IBM corporation, Austin, TX) equipped with a total of 0.3 Gb of RAM, 29 Gb of disc, and with a peak theoretical speed of ~ 0.7 Gflops, as well as at the University of Illinois at Chicago on an IBM RISC 6000/Model 560 computer.

glycine, alanine and valine residues in both SNase and calmodulin, Fig. 1E. Considerable improvements in the quality of fit are obtained upon structure relaxation (or refinement), and the bias from the X-ray refinements is removed. However, the slope of Fig. 1E is only −0.73, which

is largely determined by the calculated chemical-shift differences between the different residues. The slope improves to −0.76 by assigning bond lengths from the AMBER force field to all residues; this correction was performed by using the shielding derivatives obtained from model calculations. Further small improvements are obtained by using bond lengths derived from X-ray structures of small peptides (In et al., 1987; Glowka, 1988; Bhandary and Kopple, 1991), giving a slope of −0.79.

The final correction we make arises because the measured shift is actually an average over all the configurations that a molecule samples during the time period of the NMR measurement. These 'corrections' arise from the nonrigidity and anharmonicity of molecules and have previously been found to be necessary in order to properly compare calculated with experimental chemical-shift values (Fowler et al., 1981; Lazzeretti et al., 1987; Jameson et al., 1991a,b). Specifically, corrections are needed in order to go from an equilibrium configuration (r_e) to the vibrational ground state (r_0). As an initial approximation, we take the rovibrational corrections for bond lengths and bond angles to be similar for sites in similar chemical environments. Differences will occur, of course, in the amplitude of these corrections due to the change in the number of heavy atoms as one goes from glycine to alanine to valine, but little is known regarding rovibrational corrections for molecules as large as proteins, and force fields currently available are not accurate enough for this purpose. The magnitude of the rovibrational corrections to shielding, however, do not only depend on the averaging of the coordinates, but also on the sensitivity of the shielding with respect to structural changes, and from studies performed on small molecules, it is known that such corrections to shielding are overwhelmingly dominated by bond-length contributions. Model calculations (manuscript in preparation) show that both alanine and valine have significantly more negative shielding derivatives with respect to bond-length changes than glycine. For example, the shielding derivatives with respect to the C^{α}-N bond are found to be −88 and −84 ppm/Å for alanine and valine, respectively, while for the C^{α}-C^{o} bond the values are −57 ppm/Å for alanine and −60 ppm/Å for valine. Glycine, on the other hand, has only −69 ppm/Å for C^{α}-N and −48 ppm/Å for C^{α}-C^{o}. If one assumes that the corrections are dominated by contributions from the C^{α}-N and C^{α}-C^{o} bonds and that ($r_0 - r_e$) values for these bonds in all three residues are similar, then it would be expected that the rovibrational corrections for glycine are ~ 20% less than those of alanine and valine, based on the fact that the derivatives for glycine are only 80% of those for alanine and valine.

Calculated values for the C^{α} shifts of alanine sites in SNase (de Dios et al., 1993) reveal a ~ 12 ppm offset when compared against experimental *absolute* shielding values (where an absolute shielding value of 186.43 ppm has been used for the reference compound, tetramethylsilane (Jameson and Jameson, 1987)). While this offset might be due to other than rovibrational effects, both the sign and magnitude of the offset are comparable to those found for small molecules (Fowler et al., 1981; Lazzeretti et al., 1987; Jameson et al., 1991a,b). Thus, if a −12 ppm rovibrational correction is assumed for alanine and valine, then glycine would have a −9.6 ppm correction (80% of 12 ppm), based solely on the shielding derivatives; that is, it would be ~ 2.4 ppm more shielded. These corrections, together with the peptide bond-length corrections, are incorporated in Fig. 1F, and it can be seen that there is a noticeable improvement over the results shown in Fig. 1E. In particular, the slope improves from m = −0.73 to m = −0.91, and R and r.m.s.d. values also both improve slightly (from R = 0.97 to R = 0.98; r.m.s.d. from 1.9 ppm to 1.4 ppm).

The results presented above show that solution NMR chemical shifts for C^{α} sites in glycine,

alanine and valine residues in proteins can now be quite accurately predicted from known (X-ray) structures using quantum-chemical methods. Highly refined initial structures are essential, but even these need to be improved upon by use of structure relaxation, bond-length/bond-angle corrections via use of shielding derivatives, and, as we suggest, rovibrational corrections. The ability to predict the entire ~ 24 ppm C^{α} shielding range without hydrogen-bond partners and with an overall r.m.s.d. of ~ 1.4 ppm augurs well for using ^{13}C NMR chemical shifts in testing, refining and potentially in predicting protein structure — both in solution and in the solid state. In particular, we find that the conformation around torsion angles ϕ,ψ and χ overwhelmingly dominate C^{α} shielding, with electrostatic field contributions being < 10% of the shielding range for each amino acid (a maximal value, since the dielectric constant ε has a value of 1). Also, by their very nature, quantum-chemical calculations provide an independent route to verification of chemical-shift assignments. For example, the assignment of points which fall off of a general trend by a substantial amount might, in general, bear close inspection.

ACKNOWLEDGEMENTS

We are grateful to Professor P. Pulay, Dr. K. Wolinski and Professor J.F. Hinton for providing their TEXAS 90 program, to Professor C.J. Jameson for helpful advice and for use of her IBM RISC 6000/Model 560 computer, and to Professor J. Markley for providing unpublished assignments on SNase. This work was supported by the United States National Institutes of Health (grants HL-19481 and GM-40426).

REFERENCES

Bax, A. (1989) *Annu. Rev. Biochem*, **58**, 223–256.
Bhandary, K.K. and Kopple, K.D. (1991) *Acta Crystallogr.*, **C47**, 1280–1283.
Billeter, M., Vendrell, J., Wider, G., Avilés, F.X., Coll, M., Guasch, A., Huber, R. and Wüthrich, K. (1992) *J. Biomol. NMR*, **2**, 1–10.
Chattopadhyaya, R., Meador, W.E., Means, A.R. and Quiocho, F.A. (1992) *J. Mol. Biol.*, **228**, 1177–1192.
Chesnut, D.B. and Moore, K.D. (1989) *J. Comput. Chem.*, **10**, 648–659.
de Dios, A.C. and Oldfield, E. (1993) *Chem. Phys. Lett.*, **205**, 108–116.
de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993) *Science*, **260**, 1491–1496.
Ditchfield, R. (1972) *J. Chem. Phys.*, **56**, 5688–5691.
Fowler, P.W., Riley, G. and Raynes, W.T. (1981) *Mol. Phys.*, **42**, 1463–1481.
Glowka, M.L. (1988) *Acta Crystallogr.*, **C44**, 1639–1641.
Hynes, T.R. and Fox, R.O. (1991) *Protein Struct. Funct. Genet.*, **10**, 92–105.
Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667.
In, Y., Doi, M., Ishida, T. and Inoue, M. (1987) *Acta Crystallogr.*, **C43**, 770–773.
Jameson, A.K. and Jameson, C.J. (1987) *Chem. Phys. Lett.*, **134**, 461–466.
Jameson, C.J., de Dios, A.C. and Jameson, A.K. (1991a) *J. Chem. Phys.*, **95**, 1069–1079.
Jameson, C.J., de Dios, A.C. and Jameson, A.K. (1991b) *J. Chem. Phys.*, **95**, 9042–9053.
Krishnan, R., Binkley, J.S., Seeger, R. and Pople, J.A. (1980) *J. Chem. Phys.*, **72**, 650–654.
Lazzeretti, P., Zanasi, R., Sadlej, A. and Raynes, W.T. (1987) *Mol. Phys.*, **62**, 605–616.
Lee, F.S., Chu, Z. and Warshel, A. (1993) *J. Comput. Chem.*, **14**, 161–185.
Loll, P.J. and Lattman, E.E. (1989) *Protein Struct. Funct. Genet.*, **5**, 183–201.
Oschkinat, H., Griesinger, C., Kraulis, P.J., Sørensen, O.W., Ernst, R.R., Gronenborn, A.M. and Clore, G.M. (1988) *Nature*, **332**, 374–376.
Taylor, D.A., Sack, J.S., Maune, J.F., Beckingham, K. and Quiocho, F.A. (1991) *J. Biol. Chem.*, **266**, 21375–21380.

Torchia, D.A., Sparks, S.W. and Bax, A. (1989) *Biochemistry*, **28**, 5509–5524.
Wang, J., Hinck, A.P., Loh, S.N., LeMaster, M. and Markley, J.L. (1992) *Biochemistry*, **31**, 921–936.
Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. (1986) *J. Comput. Chem.*, 7, 230–252.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Wolinski, K., Hinton, J.F. and Pulay, P. (1990) *J. Am. Chem. Soc.*, **112**, 8251–8260.