JOURNAL OF MAGNETIC RESONANCE, Series B 107, 293–297 (1995)

The Impact of Direct Refinement against Proton Chemical Shifts on Protein Structure Determination by NMR

JOHN KUSZEWSKI, ANGELA M. GRONENBORN, AND G. MARIUS CLORE

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

Received March 9, 1995

Proton chemical shifts have long been recognized to contain highly valuable structural information (*1–10*). Recent developments in empirical models for chemical-shift calculations have shown that it is now possible to predict ^{1}H chemical shifts for nonexchangeable protons to within 0.23–0.25 ppm for proteins for which high-resolution crystal structures are available (*11, 12*). Further, it is generally observed that the shifts calculated from high-resolution crystal structures are significantly closer to the experimental ones than those calculated from the corresponding NMR structures (*13*). In this light, it seems of interest to incorporate ^{1}H chemical-shift restraints directly into the refinement of protein structures determined by NMR, in a manner analogous to that recently described for $^3J_{HN\alpha}$ coupling constants (*14*) and $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts (*15*). ^{1}H chemical-shift restraints were recently employed in the determination of the solution structure of carbonmonoxy myoglobin (*16*). However, the presence of a heme group that exerts very large ring-current shifts, together with the facts that the calculations were based on a relatively low number of experimental interproton distance restraints (an average of 8.5 per residue) and that the resulting structures were of relatively low precision, makes it difficult to assess the real impact of direct ^{1}H chemical-shift refinement for a non-heme-containing protein. In this Communication, we have examined the results of direct ^{1}H chemical-shift refinement using human thioredoxin (105 residues) as a model system. This particular system provides an ideal test case as a very-high-precision NMR structure, based on about 29 experimental NMR restraints per residue, is available (*17*).

The calculated ^{1}H chemical shift, σ_{calc}, can be decomposed into the four terms

$$\sigma_{calc} = \sigma_{random} + \sigma_{ring} + \sigma_{ani} + \sigma_E, \qquad [1]$$

where σ_{random}, σ_{ring}, σ_{ani}, and σ_E are the random coil, ring-current, magnetic-anisotropy, and electric-field shifts, respectively. To incorporate ^{1}H chemical shifts directly into the refinement procedure, we have simply added a ^{1}H chemical-shift pseudoenergy term, E_{prot}, of the form

$$E_{prot} = \sum_i k_{prot} (\sigma_{calc,i} - \sigma_{obs,i})^2, \qquad [2]$$

where k_{prot} is the force constant, and $\sigma_{obs,i}$ and $\sigma_{calc,i}$ are the observed and calculated ^{1}H chemical shifts, respectively, of proton i, to the simulated-annealing program XPLOR (*18, 19*).

In implementing the target function given by Eq. [2], our description and parameterization of the various terms in Eq. [1] describing σ_{calc} follows exactly that given by Williamson and Asakura (*12*). The value of σ_{ring} is given by

$$\sigma_{ring} = iBG(\mathbf{r}), \qquad [3]$$

where i is a ring-current intensity factor, B is a proportionality constant related to the susceptibility of the ring, G is a spatial term, and $\mathbf{r}$ is the vector from the observed proton to the ring. The description employed for $G(\mathbf{r})$ is that of the Haigh–Mallion theory (*20*), where

$$G(\mathbf{r}) = \sum S_{ij} (r_i^{-3} + r_j^{-3}), \qquad [4]$$

where r_i and r_j are the distances from the ring atoms i and j to the proton of interest, and S_{ij} is the area of the triangle formed by atoms i and j and the proton projected onto the plane of the ring; σ_{ani} is described by the M1 model and represents the sum of the anisotropies arising from the C′–O and C′–N bonds of the backbone and the side-chain functional groups of Asp, Glu, Asn, and Gln (*2, 15*),

$$\sigma_{ani} = \sum^{C'-N,C'-O} (3r_{X(j)H})^{-3} [\Delta\chi_{1,j} (3\cos^2\theta - 1) + \Delta\chi_{2,j}(3\cos^2\phi - 1)], \qquad [5]$$

where $r_{X(j)H}$ is the distance from the proton to the center of anisotropy X along the C–O or C–N bond j; $\Delta\chi_{1,j}$ and $\Delta\chi_{2,j}$ are constants that depend on the bond type j; θ is the angle between the X(j)H vector and the normal to the plane defined by the atoms of the bond of interest plus a third atom

1064-1866/95 $6.00

TABLE 1
Structural Statistics[a]

	$\langle SA_{pc} \rangle$	$\langle SA_c \rangle$	$\langle SA_{ori} \rangle$
	RMS deviation from ^1H shifts (ppm)[b]		
Overall (403)	0.25 ± 0.002	0.31 ± 0.003	0.31 ± 0.004
CαH (105)	0.28 ± 0.003	0.32 ± 0.004	0.31 ± 0.004
methyl (51)	0.20 ± 0.006	0.29 ± 0.013	0.29 ± 0.010
other (247)	0.24 ± 0.003	0.30 ± 0.005	0.31 ± 0.005
	RMS deviation from carbon chemical shifts (ppm)[c]		
$^{13}C\alpha$ (102)	1.08 ± 0.01	1.03 ± 0.03	1.23 ± 0.01
$^{13}C\beta$ (97)	0.99 ± 0.01	0.94 ± 0.02	1.13 ± 0.02
	RMS deviations from experimental restraints		
Interproton distances (Å) (2571)	0.015 ± 0.001	0.013 ± 0.001	0.008 ± 0
Torsion angles (degrees) (273)	0.280 ± 0.058	0.193 ± 0.038	0.169 ± 0.015
$^3J_{HN\alpha}$ (Hz) (89)	0.46 ± 0.015	0.42 ± 0.004	0.33 ± 0.008
E_{LJ} (kcal mol^{-1})[d]	−495 ± 9	−506 ± 9	−493 ± 7
	RMS deviations from idealized geometry		
Bonds (Å) (1653)	0.004 ± 0	0.004 ± 0	0.003 ± 0
Angles (degrees) (2985)	0.519 ± 0.008	0.506 ± 0.008	0.480 ± 0.003
Improper torsions (degrees) (838)	0.308 ± 0.010	0.273 ± 0.012	0.227 ± 0.003

[a] The notation of the structures is as follows: $\langle SA_{pc} \rangle$, $\langle SA_c \rangle$, and $\langle SA_{ori} \rangle$ are the ensemble of proton and carbon chemical-shift refined, carbon chemical shift refined, and original structures, respectively. There are 40 individual structures in each ensemble, and $\overline{SA}_{pc}$, $\overline{SA}_c$, and $\overline{SA}_{ori}$ are the mean coordinates obtained by averaging the coordinates of the individual structures in $\langle SA_{pc} \rangle$, $\langle SA_c \rangle$, and $\langle SA_{ori} \rangle$ ensembles, respectively. The $\langle SA_c \rangle$ and $\langle SA_{ori} \rangle$ coordinates are from (*15*) and (*17*), respectively. The number of terms for the various restraints is given in parentheses. All the structures were calculated using the simulated-annealing protocol of (*25*) with minor modifications using the program XPLOR (*18, 19*). The final values of the force constants for the various terms are as follows: 7.5 kcal mol^{-1} ppm^{-2}, 0.5 kcal mol^{-1} ppm^{-2}, and 1 kcal mol^{-1} Hz^{-2} for the harmonic proton chemical shift, carbon chemical shift, and coupling constant restraints, respectively; 30 kcal mol^{-1} Å^{-2} and 200 kcal mol^{-1} rad^{-2} for the square-well interproton distance and torsion angle restraints, respectively; 500 kcal mol^{-1} Å^{-2} and 500 kcal mol^{-1} rad^{-2} for the harmonic bond and angle (bond angles and improper torsions) terms, respectively; and 4 kcal mol^{-1} Å^{-4} for the quartic van der Waals repulsion term with the hard-sphere van der Waals radii set to 0.8 times their values in the CHARMM PARAM19/PARAM20 energy parameters (*26*).

[b] All exchangeable protons (backbone and side-chain amides) are excluded.

[c] All Cα and Cβ chemical shifts are included with the exception of the shifts for the two cysteine residues (Cys32 and Cys35) and the single histidine (His44) and tryptophan (Trp31) residues (*15*).

[d] E_{LJ} is the Lennard-Jones van der Waals energy calculated with the CHARMM empirical energy function (*26*). It is not included in the target function for simulated annealing.

(Cα, C′, and O for the C′–O bond, O, C′, and N for the C′–N bond, Cβ, Cγ, and Oδ1 for Asn, Cγ, Cδ, and Oε1 for Gln, Cβ, Cγ, Oδ1 and Cβ, Cγ, Oδ2 for Asp, and Cγ, Cδ, Oε1 and Cγ, Cδ, Oε2 for Glu); and ϕ is the angle between the X(j)H vector and the axis located in the plane of the three atoms that is perpendicular to the z axis defined by the bond of interest. Finally, σ_E is described by the E2 model (*21, 22*),

$$\sigma_E = \sum_{i=1}^{\text{Nheavyatom}} (Q_i / r_{iH}^2) \cos \theta_i, \qquad [6]$$

where i represents a heavy atom, θ_i is the angle between the i–H and C–H vectors, Q_i is the charge on i, and r_{iH} is the distance from i to H.

In order for the atomic forces to be calculated during simulated annealing (or conjugate gradient minimization), the partial derivatives of the proton shift energy with respect to the x, y, and z coordinates of each atom that contributes to a change in a proton shift must be calculated. Thus, for example, the partial derivative of E_{prot} with respect to the x coordinate of atom i is calculated by the chain rule as

$$\partial E_{prot}/\partial x_i = 2k_{prot}(\sigma_{ring} + \sigma_{ani} + \sigma_E + \sigma_{random} - \sigma_{obs}) \times (\partial\sigma_{ring}/\partial x_i + \partial\sigma_{ani}/\partial x_i + \partial\sigma_E/\partial x_i), \qquad [7]$$

where $\partial\sigma_{ring}/\partial x_i$, $\partial\sigma_{ani}/\partial x_i$, and $\partial\sigma_E/\partial x_i$ are the partial derivatives of σ_{ring}, σ_{ani}, and σ_E with respect to the x coordinate of atom i. Note that $\partial\sigma_{random}/\partial x_i$ and $\partial\sigma_{obs}/\partial x_i$ are zero,

TABLE 2
Atomic RMS Differences[a]

	Atomic RMS difference (Å)		
	Backbone atoms	All atoms	All ordered atoms[b]
Coordinate precision			
$\langle SA_{pc} \rangle$ vs $\overline{SA}_{pc}$	0.21 ± 0.03	0.61 ± 0.03	0.33 ± 0.03
$\langle SA_{c} \rangle$ vs $\overline{SA}_{c}$	0.23 ± 0.03	0.63 ± 0.04	0.35 ± 0.03
$\langle SA_{ori} \rangle$ vs $\overline{SA}_{ori}$	0.19 ± 0.02	0.60 ± 0.03	0.30 ± 0.02
Atomic RMS shifts			
$\overline{SA}_{pc}$ vs $\overline{SA}_{ori}$	0.32	0.39	0.36
$\overline{SA}_{pc}$ vs $\overline{SA}_{c}$	0.24	0.31	0.27
$\overline{SA}_{c}$ vs $\overline{SA}_{ori}$	0.19	0.27	0.22

[a] The notation of the structures is given in footnote *a* to Table 1.
[b] The list of ordered atoms is given in (*17*), which describes the original structures.

since these terms do not depend on the coordinates. The partial derivatives of the individual terms ($\partial\sigma_{ring}/\partial x_i$, etc.) are long expressions and were evaluated using Mathematica (*23*). In order to make them computationally feasible, subexpressions that the partial derivatives with respect to the varying atoms have in common were calculated first and saved as temporary variables. A similar procedure was undertaken recursively to simplify each common subexpression. The procedure was automated with the Mathematica program "Optimize" (*24*).

The results of direct refinement against ^{1}H chemical shifts are summarized in Tables 1 and 2. The starting structures comprised the original ensemble of 40 simulated-annealing structures calculated on the basis of 2933 experimental NMR restraints consisting of interproton distances, torsion angles, and $^3J_{HN\alpha}$ coupling constants (*17*). These structures also served as the starting structures in our study on the effect of direct refinement against ^{13}C chemical-shift restraints (*15*). The new ensemble of structures was calculated incorporating both ^{1}H and ^{13}C chemical shifts into the refinement using the simulated-annealing protocol of Nilges *et al.* (*25*) with minor modifications. This particular protocol which involves a period of dynamics at 1500 K followed by slow cooling to 100 K has a very large radius of convergence and has been designed to overcome large energy barriers on the path toward the global minimum region and to rigorously sample the conformational space consistent with the target function (*25*). Thus, for example, chains can readily interpenetrate each other during the early stages of the refinement. The quality of the nonbonded contacts (assessed independently by calculating the Lennard–Jones van der Waals energy) and the deviations from idealized covalent geometry remain unaffected by the incorporation of ^{1}H chemical-shift restraints. The agreement with the experimental NMR restraints is within experimental error and, with the obvious exception of the ^{1}H chemical-shift restraints, is comparable to those of the structures refined against $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts (Table 1).

The deviations between the calculated ^{1}H shifts for an ensemble of 40 random structures (with intact covalent geometry and no van der Waals overlap) and the observed

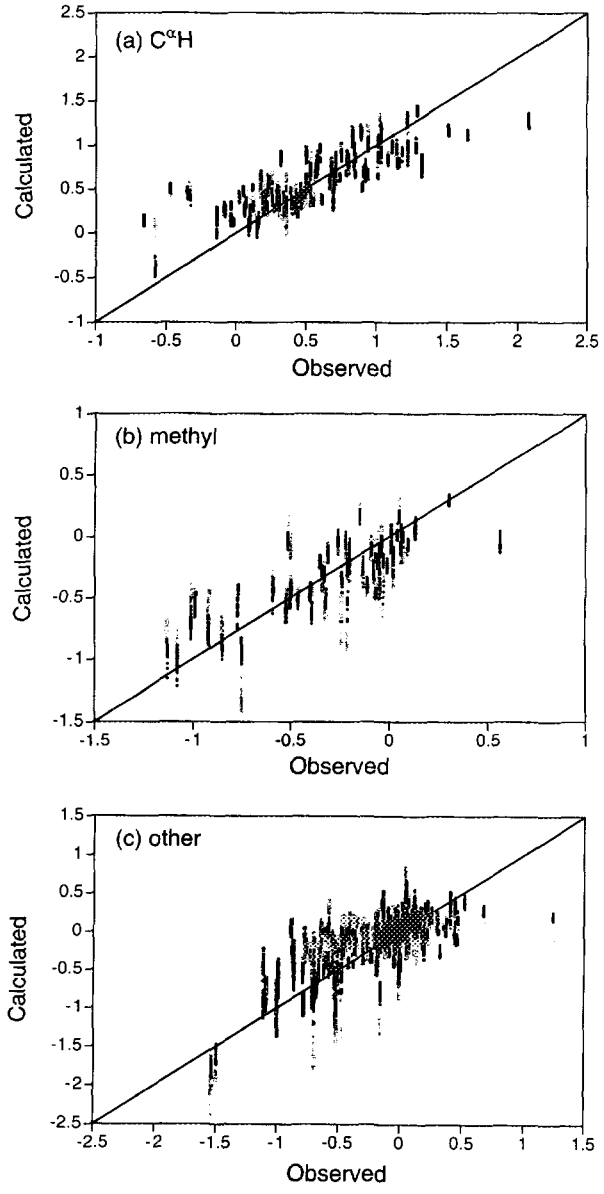

FIG. 1. Agreement between observed and calculated ^{1}H chemical shifts (ppm) before (gray circles) and after (black circles) ^{1}H chemical-shift refinement for (a) CαH protons, (b) methyl protons, and (c) all other nonexchangeable protons of all 40 simulated-annealing structures. For a linear fit to the observed versus calculated chemical-shift data, the correlation coefficients for the CαH, methyl, and all other nonexchangeable protons are 0.769, 0.628, and 0.599, respectively, before ^{1}H chemical-shift refinement, and 0.820, 0.838, and 0.757, respectively, after ^{1}H chemical-shift refinement.

shifts is 0.48 ± 0.02 ppm for all nonexchangeable protons, 0.57 ± 0.03 ppm for CαH protons, 0.44 ± 0.02 ppm for methyl protons, and 0.47 ± 0.03 ppm for all remaining nonexchangeable protons. If we consider that the error in the calculation of the ^{1}H shifts is 0.23 ppm [the value of the best RMS deviation between observed and calculated shifts obtained from an analysis of crystal structures (*12*)] then the fraction of further meaningful improvement, defined as $(RMS_{obs} - 0.23)/(RMS_{random} - 0.23)$, that could be obtained for the structures refined with ^{1}H chemical-shift restraints is 0.08 for all nonexchangeable protons, 0.15 for CαH protons, 0 for methyl protons, and 0.04 for the remaining nonexchangeable protons. The corresponding values for the two ensembles of structures calculated without ^{1}H chemical-shift restraints are 0.32, 0.24–0.26, 0.29, and 0.29–0.33, respectively. Hence, there is a large and highly significant improvement in agreement between observed and calculated ^{1}H chemical shifts upon refinement with ^{1}H chemical-shift restraints. This is clearly illustrated by the plots of observed versus calculated ^{1}H chemical-shifts shown in Fig. 1 and by the histograms of the distribution of errors shown in Fig. 2.

The precision of the structures is unaffected by ^{1}H chemical-shift refinement but is accompanied by a small atomic RMS shift in the coordinates (Table 2). The latter is a little larger than that obtained when only ^{13}C chemical shifts are used in the refinement and is most marked in the region of the active site due to the presence of a Trp (Fig. 3). Nevertheless, the three ensembles of structures still overlap in virtually all regions of the structure. The fact that a large and significant improvement in the agreement between observed and calculated ^{1}H chemical shifts can be achieved without affecting the precision of the coordinates and with only small atomic RMS shifts is a testament to the high quality of the starting ensemble of structures. It also suggests that while the atomic RMS shifts upon ^{1}H chemical-shift refinement are small they may be significant and lead to an increase in accuracy of the coordinates.

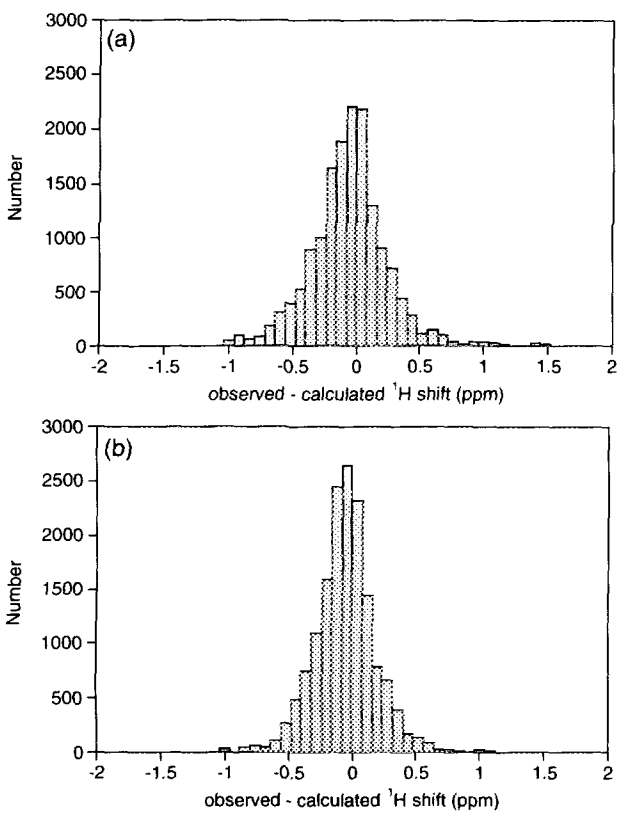

FIG. 2. Histogram of the distribution of errors between calculated and observed ^{1}H shifts (a) before and (b) after ^{1}H chemical-shift refinement for all nonexchangeable protons of all 40 simulated-annealing structures.

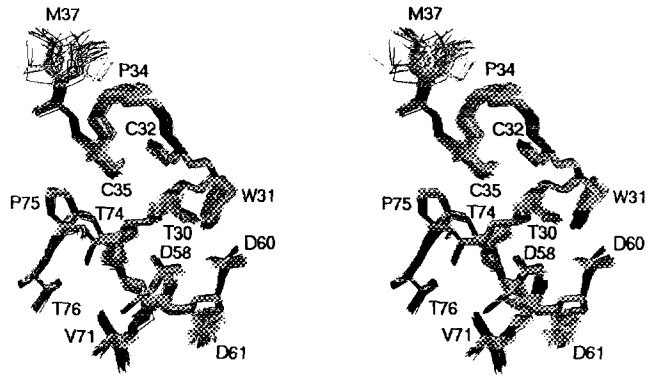

FIG. 3. Stereoview of the active site and neighboring regions of reduced human thioredoxin showing a superposition of the ensemble of 40 simulated-annealing structures before (gray) and after (black) refinement against ^{1}H chemical shifts.

ACKNOWLEDGMENTS

We thank Kirsten Frank and Ad Bax for useful discussions. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (G.M.C. and A.M.G.).

REFERENCES

1. J. L. Markley, D. H. Meadows, and O. Jardetzky, *J. Mol. Biol.* **27,** 25 (1967).
2. N. J. Clayden and R. J. P. Williams, *J. Magn. Reson.* **49,** 383 (1982).
3. C. Redfield, J. C. Hoch, and C. M. Dobson, *FEBS Lett.* **159,** 132 (1983).
4. M. A. Weiss and J. C. Hoch, *J. Magn. Reson.* **72,** 324 (1987).
5. A. Pastore and V. Saudek, *J. Magn. Reson.* **90,** 165 (1990).
6. D. S. Wishart, B. D. Sykes, and F. M. Richards, *J. Mol. Biol.* **222,** 311 (1991).
7. T. Asakura, E. Nakamura, H. Asakawa, and M. Demura, *J. Magn. Reson.* **93,** 355 (1991).
8. M. P. Williamson, T. Asakura, E. Nakamura, and M. Demura, *J. Biomol. NMR* **2,** 83 (1992).
9. K. A. Ösapay and D. A. Case, *J. Biomol. NMR* **4,** 215 (1994).
10. D. A. Case, H. J. Dyson, and P. E. Wright, *Methods Enzymol.* **239,** 392 (1994).

11. K. Ösapay and D. A. Case, *J. Am. Chem. Soc.* **113,** 9436 (1991).
12. M. P. Williamson and T. Asakura, *J. Magn. Reson. B* **101,** 63 (1993).
13. M. P. Williamson, J. Kikuchi, and T. Asakura, XVIth International Conference on Magnetic Resonance in Biological Systems, Veldhoven, The Netherlands, Abstract A86, 1994.
14. D. S. Garrett, J. Kuszweski, T. J. Hancock, P. J. Lodi, G. W. Vuister, A. M. Gronenborn, and G. M. Clore, *J. Magn. Reson. B* **104,** 99 (1994).
15. J. Kuszewski, J. Qin, A. M. Gronenborn, and G. M. Clore, *J. Magn. Reson. B* **106,** 92 (1995).
16. K. Ösapay, Y. Theriault, P. E. Wright, and D. A. Case, *J. Mol. Biol.* **244,** 183 (1994).
17. J. Qin, G. M. Clore, and A. M. Gronenborn, *Structure* **2,** 503 (1994).
18. A. T. Brünger, G. M. Clore, A. M. Gronenborn, and M. Karplus, *Proc. Natl. Acad. Sci. USA* **83,** 3801 (1986).
19. A. T. Brünger, "X-PLOR Version 3.0 Manual," Yale University, New Haven, Connecticut, 1992.
20. C. W. Haigh and R. B. Mallion, *Prog. NMR Spectrosc.* **13,** 303 (1980).
21. J. I. Musher, *J. Chem. Phys.* **37,** 34 (1962).
22. G. K. Hamer and W. F. Reynolds, *Can. J. Chem.* **46,** 3813 (1968).
23. S. Wolfram, "Mathematica: A System for Doing Mathematics by Computer," Addison–Wesley, Redwood City, California, 1991.
24. T. Robb. "Optimize," MathSource World Wibe Web Server (http://www.wri.com/MathSource), 1994.
25. M. Nilges, G. M. Clore, and A. M. Gronenborn, *FEBS Lett.* **229,** 317 (1988).
26. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4,** 187 (1983).