JOURNAL OF MAGNETIC RESONANCE, Series B 101, 320–324 (1993)

Estimation of Interatomic Distances in Proteins from NOE Spectra at Longer Mixing Times Using an Empirical Two-Spin Equation

ASIF K. SURI AND RONALD M. LEVY*

Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855-0939

Received November 5, 1992; revised February 11, 1993

Two-dimensional nuclear Overhauser enhancement experiments have provided the structural information necessary to determine the three-dimensional structures of many proteins in solution (*1–5*). Spectra from such experiments yield information about the spatial proximity of nuclear spins (*6–10*). For small rigid molecules at short mixing times τ_m, the intensity of the NOE between protons i and j is proportional to the inverse sixth power of the distance between the two protons. In macromolecules, indirect magnetization transfer (spin diffusion) from nearby protons and intramolecular dynamics become important factors that affect the magnitude of the NOE between protons i and j and contribute to errors in estimating interproton distances from NOE cross peaks when using the two-spin approximation (*11–13*). To minimize spin-diffusion effects, experiments are often limited to short mixing times (~100 ms), for which the two-spin approximation is considered valid. We have found additional structural information in the longer-mixing-time experiments (*14*).

One way to extract this information is to perform a relaxation-matrix analysis to incorporate multispin effects (*15–20*). One such approach involves the construction of a hybrid NOE intensity matrix containing both experimental intensities and those calculated from a model structure. Inversion of this matrix creates the relaxation matrix from which interproton distances can be estimated. The existing model structure can then be improved with these new distances and a new hybrid NOE matrix generated (*21*). In a more computationally efficient variation of this procedure, the hybrid NOE intensity and relaxation matrices can be iterated to approach self-consistency before regenerating a new model structure (*22, 23*). The most computationally intensive relaxation-matrix analysis involves a refinement procedure where the difference between the calculated and experimental NOE matrices is minimized in a least-squares sense, subject to stereochemical constraints. The NOE matrix and its gradient are frequently recalculated from an evolving model structure (*20, 24, 25*). The improvement in the structures obtained depends on many factors (*26, 27*). In this Communication we suggest a simple empirical modification of the "two-spin equation" relating experimental NOE intensities to interproton distances, which incorporates spin-diffusion effects in an average way. Although our method for incorporating spin diffusion into structure refinement is in principle more approximate than the explicit refinement of calculated against experimental intensities, the intrinsic effects of motional dynamics limit the accuracy of the current structure-refinement procedures in any case.

The evolution of the NOE intensities as a function of time is described by the Bloch equation (*28*)

$$\frac{d}{dt}\mathbf{N}(t) = -\sigma\mathbf{N}(t), \qquad [1]$$

where $\mathbf{N}$ is the matrix of cross-peak intensities. The elements of the relaxation matrix σ contain the structural information

$$\sigma_{ij} = \left(\frac{\mu_0}{4\pi}\right)^2\left(\frac{\gamma_i^2\gamma_j^2\hbar^2}{10\langle r_{ij}^3\rangle^2}\right)[6J(2\omega) - J(0)] \qquad [2a]$$

$$\sigma_{ii} = E_i + \sum_{i\neq j}\sigma_{ij} \qquad [2b]$$

$$J(\omega) = \frac{\tau_c}{1 + \omega^2\tau_c^2}, \qquad [2c]$$

where σ_{ij} and σ_{ii} are the cross-relaxation and direct-relaxation rates, respectively. E_i, the external relaxation rate of spin i, is constant for all spins with a value of 0.5 s^{-1} in the simulations presented below. The spectrometer frequency is 500 MHz, and γ is the gyromagnetic ratio for ^1H nuclei. For the spectral density function, $J(\omega)$, we assume an isotropic tumbling model with an overall tumbling time τ_c. Given a structural and motional model from which the elements of the relaxation matrix are constructed, the Bloch equation

* To whom correspondence should be addressed.

1064-1866/93 $5.00

can be integrated analytically to obtain the cross-peak intensities at mixing time τ_m

$$\mathbf{N}(\tau_m) = \mathbf{U}^t\mathbf{e}^{-\lambda\tau_m}\mathbf{U}\mathbf{N}(0), \quad [3]$$

where $\mathbf{e}^{-\lambda\tau_m}$ is a diagonal matrix, λ is an eigenvalue of σ, and $\mathbf{U}$ is the matrix of eigenvectors. The elements $N_{ij}(\tau_m)$ depend on the geometric relation of protons i and j to all the other protons and not just the distance between protons i and j. However, in the linear two-spin approximation and in the absence of internal motion, the following relation is used to estimate r_{ij} from a measured (or simulated) cross-peak intensity $N_{ij}(\tau_m)$ (*29–31*):

$$\left[\frac{N_{ij}(\tau_m)}{N_{ref}(\tau_m)}\right]^{1/6} \approx \left(\frac{\sigma_{ij}}{\sigma_{ref}}\right)^{1/6} = \frac{r_{ref}}{r_{ij}}. \quad [4]$$

The N_{ref} and r_{ref} correspond to the intensity and distance of a calibrating proton pair; e.g., a common choice is the NOE between two methylene protons.

There is a systematic error introduced by spin diffusion when the two-spin equation (Eq. [4]) is used to extract distances from intensities (*19, 27*). The effect of spin diffusion at longer mixing times is most often to increase the NOE intensity between spin pairs that are further apart because additional spin density can flow between the spins via indirect pathways containing intervening spins. The estimated interproton distances are therefore shorter than the true distances. In addition, spin diffusion tends to reduce the intensity of peaks with short interproton distances, such as methylenes. Thus the relative effects of spin diffusion for reference spin pairs compared with experimental pairs contribute to the error in the estimation of interproton distances. It should be possible to compensate in an average way for errors in estimated distances based on Eq. [4], introduced through spin diffusion, by replacing the exponent 1/6 in Eq. [4] with an empirical adjustable parameter $a > 1/6$:

$$r_{ij} = \left(\frac{N_{ref}}{N_{ij}}\right)^a r_{ref}. \quad [5]$$

We have carried out simulations of spin-diffusion effects in order to optimize the exponent in Eq. [5]. In these simulations, NOE intensities at different mixing times were calculated for two proteins, crambin and pancreatic trypsin inhibitor (*32, 33*), by solving the Bloch equations (Eq. [1]) based on the minimized crystal structures of the proteins. From the simulated NOE intensities, sets of distances were estimated using different values for the exponent in Eq. [5] as described below.

Simulated NOE spectra for crambin were generated at mixing times τ_m of 20, 50, 200, and 400 ms, at each of two tumbling times τ_c of 2 and 5 ns. Calculated cross peaks with intensities below 0.003 were considered to fall below the "noise" threshold and were excluded from the subsequent analysis (*14, 16*). In addition, at each mixing time, the error analysis includes only those intensities that first exceed the noise threshold at that mixing time but at shorter mixing times fall below the threshold; that is, the error analysis is carried out on the new cross peaks not previously observed. The reference intensity used for calibration is the average of all glycine methylene cross peaks. The reference distance for this proton pair is 1.78 Å. In order to optimize the value of the exponent in Eq. [5] for which the average error between simulated distances (using Eq. [5]) and the actual distances (calculated directly from the minimized crystal structure) was a minimum, several sets of simulated distances were calculated as the exponent a was varied in increments of 0.001 between 0.150 and 0.600. Figure 1 illustrates the results corresponding to a mixing time of 200 ms and a tumbling time of 5 ns. Using the standard exponent $a = 1/6$ to calculate distances from new cross-peak intensities, the average error in the estimated distance is 38%. The average error for this mixing time and tumbling time is minimized when an empirical exponent $a \approx 1/4$ is used to estimate the distances from the cross-peak intensities. For this choice of exponent the average error in the estimated distances due to spin diffusion is reduced to ≈9%, a very significant reduction in the error as compared to the error associated with the standard exponent value.

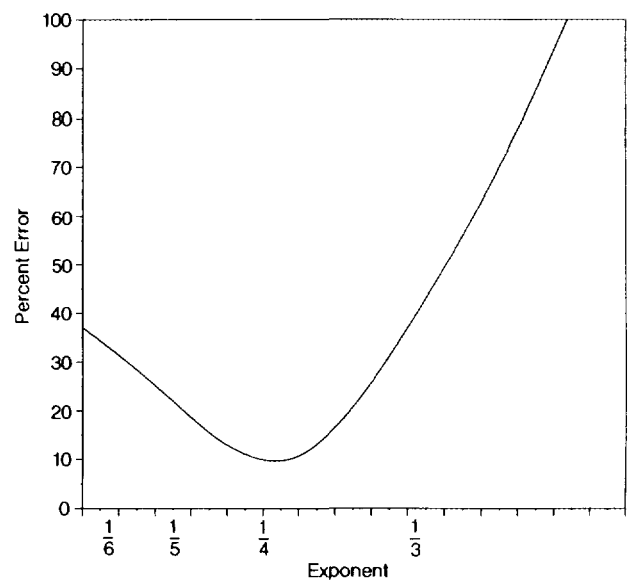

FIG. 1. Plot of percentage error vs exponent (Eq. [5]) at τ_m = 200 ms and τ_c = 5 ns. The exponent of the empirical two-spin approximation was incremented by a value of 0.001 and then used to calibrate the NOE spectrum of crambin. Average percentage errors between calculated and actual distances were computed after each increment. The exponent corresponding to the smallest percentage error was considered the optimized or "best exponent."

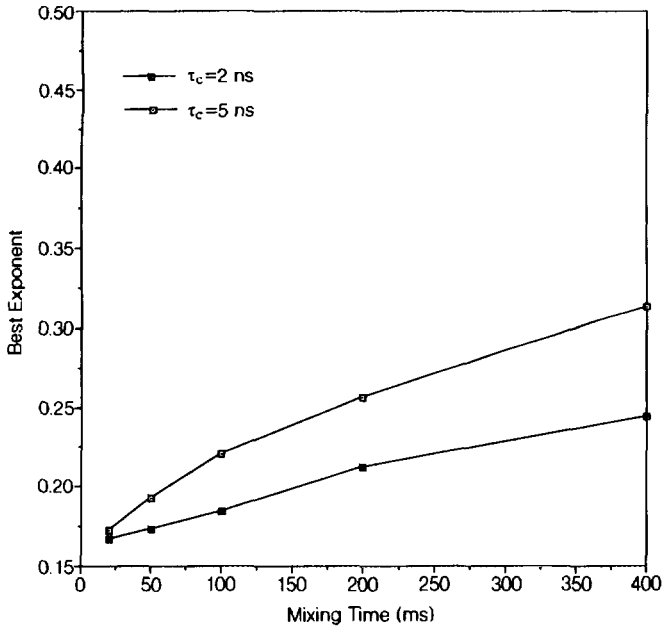

FIG. 2. Plot of the best exponent vs mixing time for crambin. The value of the exponent is shown to be sensitive to the overall tumbling time. As the shortest mixing times show, the value of the exponent converges to a value of 1/6.

In Fig. 2 we plot the optimized exponent as a function of mixing time for two different choices for the tumbling time of 2 and 5 ns. As the mixing time becomes very short, not surprisingly, the optimized exponent approaches 1/6 since this is the theoretically correct value as the mixing time approaches zero. The empirical best-fit exponent increases at larger mixing times, reflecting the larger spin-diffusion effects at the longer mixing times. At the longest mixing time and slowest tumbling time the optimized exponent value is 1/3.18. The figure also illustrates that the optimized exponent exhibits some sensitivity to the tumbling time as expected since spin diffusion is accentuated for large proteins with longer rotational correlation times.

Table 1 displays the average percentage error and rms deviation between calculated and exact distances as a function of mixing time, obtained when the distances are estimated from NOE intensities using the standard exponent $a = 1/6$. The corresponding errors and standard deviations obtained using the optimized exponents along with the optimized values of the exponents are shown in Table 2. It is clear that the use of an empirically adjusted exponent which incorporates spin-diffusion effects in an average way can greatly reduce the errors due to spin diffusion in interproton distances estimated from NOE intensities. The largest error in distances listed in Table 1 is 44.9% observed for the longest mixing time $\tau_m = 400$ ms and slowest tumbling time $\tau_c = 5$ ns. At this same mixing time and tumbling time, the average error in the estimated distances is greatly reduced to 10.3% using the optimized empirical exponent $a = 1/3.18$ to extract the distances from the intensities (Table 2). This error is of similar magnitude to the errors observed for a larger protein ($\tau_c = 5$ ns) even at a short mixing time ($\tau_m = 50$ ms) using the standard exponent (average error = 12.3%; see Table 1).

The basis for the empirical adjustment of the exponent in the two-spin equation relies on the assumption that spin-diffusion effects can be corrected for in an average way by increasing the exponent in Eq. [5] to compensate for the apparent shortening of internuclear distances at longer mix-

TABLE 1
Analysis of New Peaks That Were Not Observed at Previous Mixing Times: Crambin

Mixing time (ms)	Exponent	Percentage error (average)	Average deviation (Å)
	(a) $\tau_c = 2$ ns		
20	0.167	0.789	0.027
50	0.167	3.727	0.202
100	0.167	9.539	0.481
200	0.167	18.153	0.902
400	0.167	28.220	1.550
	(b) $\tau_c = 5$ ns		
20	0.167	2.395	0.146
50	0.167	12.287	0.602
100	0.167	21.468	1.113
200	0.167	32.220	1.885
400	0.167	44.949	2.998

TABLE 2
Analysis of New Peaks That Were Not Observed at Previous Mixing Times: Crambin

Mixing time (ms)	Best exponent	Percentage error (average)	Average deviation (Å)
	(a) $\tau_c = 2$ ns		
20	0.167	0.789	0.027
50	0.173	2.832	0.169
100	0.184	5.640	0.322
200	0.212	6.985	0.380
400	0.224	8.431	0.528
	(b) $\tau_c = 5$ ns		
20	0.172	1.888	0.126
50	0.193	6.422	0.342
100	0.221	7.207	0.422
200	0.256	9.667	0.636
400	0.314	10.342	0.807

ing times due to spin diffusion. A comparison of the errors in estimated distances calculated using the standard exponent (Table 1) with those calculated using the optimized exponents (Table 2) demonstrates that this is a very good assumption for crambin. To check the sensitivity of our results to the protein chosen for the model calculations we have repeated the simulations of NOE intensities and optimized exponents for PTI. The results are displayed in Fig. 3. The same trends are observed for PTI as for crambin; however, the optimized exponents at each mixing time seem to be greater for PTI than Crambin. This effect turned out to be due to the calibration procedure. For both proteins, calibration of intensities at each mixing time is carried out in the simulations by averaging the NOE intensities of all glycine methylene spin pairs. The calibration turned out to be slightly different for crambin and PTI because there are so few glycines in these small proteins and there are variations in the intensity value among different glycines due to spin diffusion. For example, at τ_m = 200 ms and τ_c = 5 ns, the calculated calibration intensity corresponding to a distance of 1.78 Å is N_{ref} = 0.309 for crambin, whereas for PTI at this mixing time and tumbling time, N_{ref} = 0.241. As shown in Fig. 3, when the same calibration intensity–distance relation is used for PTI as for crambin, the optimized exponents for PTI are almost identical to the values calculated for crambin. This suggests that spin-diffusion effects on protein NOE spectra are sufficiently similar for proteins of similar size and shape to permit derivation of optimized exponents from back-calculations on reference proteins which can be applied to proteins whose structures are not known.

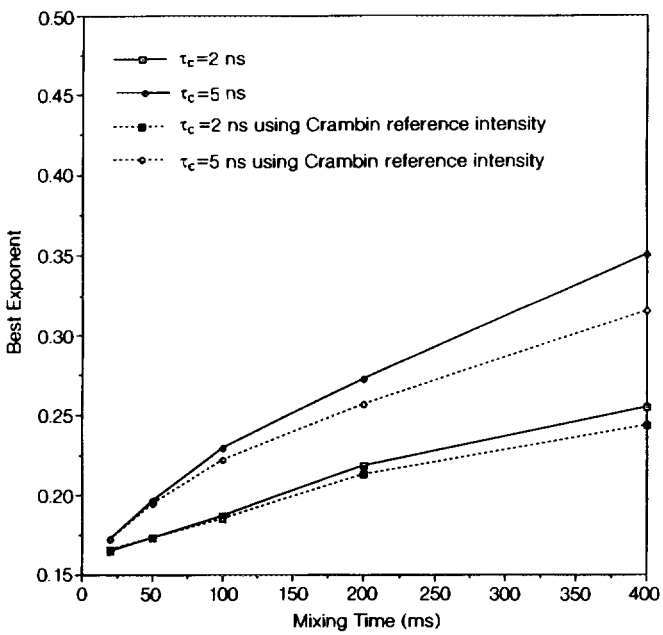

FIG. 3. Plots of the best exponent vs mixing time for PTI at both τ_c = 2 ns and τ_c = 5 ns using the average of all glycine methylene protons cross peaks within the protein as the reference intensity. The plots for PTI were redrawn after using the average crambin reference intensity for calibration (dashed lines) and were found to almost perfectly superimpose onto the corresponding crambin plots in Fig. 2.

NOE intensity data collected at longer mixing times (e.g., more than 200 ms) are not generally being used to generate protein structures because of the difficulty in correcting for spin-diffusion effects. There is strong motivation to develop simple and computationally fast methods to incorporate distances estimated from long-mixing-time experiments since typically there is a large increase in the number of new NOEs observed in protein spectra at mixing times above 200 ms (*14*). A comparison of the errors in distances estimated using the standard exponent (Table 1) with those obtained using the empirically adjusted exponents (Table 2) clearly shows that distance errors can be drastically reduced when spin diffusion is accounted for in an average way. Implementation of this method is no more difficult or time consuming than the use of the normal two-spin equation. We will report later on our development of empirical methods to estimate distance constraint ranges for NOEs measured at longer mixing times (unpublished results). It is hoped that these studies will stimulate more extensive use of longer-mixing-time data in protein structure determinations.

ACKNOWLEDGMENTS

This work has been supported by Grant GM-30580 from the National Institutes of Health. We acknowledge a grant of computer time from the National Cancer Institute.

REFERENCES

1. G. M. Clore and A. M. Gronenborn, *Crit. Rev. Biochem.* **24,** 479 (1989).
2. K. Wüthrich, "NMR of Proteins and Nucleic Acids," Wiley, New York, 1986.
3. K. Wüthrich, *Science* **243,** 45 (1989).
4. A. Bax, *Annu. Rev. Biochem.* **58,** 223 (1989).
5. P. E. Wright, *TIBS* **14,** 255 (1989).
6. I. Solomon, *Phys. Rev.* **92,** 411 (1953).
7. J. Noggle and R. Schirmer, "The Nuclear Overhauser Effect—Chemical Applications," Academic Press, New York, 1971.
8. D. Newhaus and M. Williamson, "The Nuclear Overhauser Effect in Structural and Conformational Analysis," VCH Publishers, New York, 1989.
9. P. Balaram, A. A. Bothner-By, and J. Dadok, *J. Am. Chem. Soc.* **94,** 4015 (1972).
10. F. A. L. Anet and A. J. R. Bourn, *J. Am. Chem. Soc.* **87,** 5250 (1965).
11. A. Kalk and H. J. C. Berendsen, *J. Magn. Reson.* **24,** 343 (1976).
12. E. T. Olejniczak, M. Karplus, C. M. Dobson, and R. M. Levy, *J. Am. Chem. Soc.* **106,** 1923 (1984).
13. A. N. Lane, *J. Magn. Reson.* **78,** 425 (1988).
14. Y. Chen, A. K. Suri, D. Kominos, A. M. Naylor, S. M. Pitzenberger, V. M. Garsky, R. M. Levy, and J. Baum, *J. Biomol. NMR,* submitted.
15. G. M. Clore and A. M. Gronenborn, *J. Magn. Reson.* **84,** 398 (1989).

16. B. A. Borgias and T. L. James, *J. Magn. Reson.* **79,** 493 (1988).
17. E. T. Olejniczak, R. T. Gampe, and S. W. Fesik, *J. Magn. Reson.* **67,** 28 (1986).
18. J. W. Keepers and T. L. James, *J. Magn. Reson.* **57,** 404 (1984).
19. C. B. Post, R. P. Meadows, and D. G. Gorenstein, *J. Am. Chem. Soc.* **112,** 6796 (1990).
20. P. Yip and D. Case, *J. Magn. Reson.* **83,** 643 (1989).
21. R. Boelens, T. M. G. Konig, and R. Kaptein, *J. Mol. Struct.* **173,** 299 (1988).
22. B. A. Borgias and T. L. James, *J. Magn. Reson.* **87,** 475 (1990).
23. E. P. Nickonowicz, R. P. Meadows, and D. G. Gorenstein, *Biochemistry* **29,** 4193 (1990).
24. J. D. Baleja, R. T. Pon, and B. D. Sykes, *Biochemistry* **29,** 4828 (1990).
25. J. E. Mertz, P. Güntert, K. Wüthrich, and W. Braun, *J. Biol. NMR* **1,** 257 (1991).
26. P. D. Thomas, V. J. Basus, and T. L. James, *Proc. Natl. Acad. Sci. USA* **88,** 1237 (1991).
27. D. Kominos, A. K. Suri, D. B. Kitchen, D. Bassolino, and R. M. Levy, *J. Magn. Reson.* **97,** 398 (1992).
28. A. Abragam, "The Principles of Nuclear Magnetism," Oxford Univ. Press, Oxford, 1978.
29. G. M. Clore, A. M. Gronenborn, A. T. Brünger, and M. Karplus, *J. Mol. Biol.* **186,** 435 (1985).
30. R. A. Bell and J. K. Saunders, *Can. J. Chem.* **48,** 1114 (1970).
31. A. D. Kline, W. Braun, and K. Wüthrich, *J. Mol. Biol.* **204,** 675 (1988).
32. W. A. Hendrickson and M. M. Teeter, *Nature* **290,** 107 (1981).
33. A. Wlodawer, J. Deisenhofer, and R. Huber, *J. Mol. Biol.* **193,** 145 (1987).