JOURNAL OF MAGNETIC RESONANCE **98,** 192–199 (1992)

Two-Dimensional Linear Prediction for Signals Truncated in Both Dimensions

GUANG ZHU* AND AD BAX

*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892; and *Chemical Physics Program, University of Maryland at College Park, College Park, Maryland 20742*

Received December 31, 1991; revised February 12, 1992

Resolution in multidimensional NMR is generally limited by the relatively short acquisition times in the indirectly detected dimensions. For most types of 2D experiments, signals acquired for long t_1 durations, on the order of T_2 or larger, contribute less to the signal-to-noise (S/N) ratio of the final 2D spectrum than signals acquired for short t_1 durations. As a consequence, the highest sensitivity per unit of measuring time is obtained if the t_1 acquisition time is kept shorter than T_2. Therefore, one frequently limits the number of t_1 increments to as small a number as is sufficient to provide the required spectral resolution in the F_1 dimension. Note that in the detected dimension, long acquisition times, up to several times T_2, may be used without adding to the total time needed for recording the spectrum and without adversely affecting S/N. As a consequence, in 2D NMR, truncation of the time-domain signal is a problem only in the t_1 dimension.

A variety of slightly different linear prediction techniques have been proven successful at alleviating the problems associated with truncated time-domain signals. Such techniques have been applied both to 1D signals and 2D (*1–7*), 3D (*8, 9*), and 4D (*10, 11*) data matrices. In applications where the linear prediction is used to extend the time-domain signal prior to Fourier transformation (*3, 6–11*), it reduces the effect of truncation and thereby minimizes the need for strong apodization of the *acquired* data in the t_1 dimension. As a consequence this procedure improves both sensitivity and resolution.

In 3D and 4D NMR, truncation of the time-domain data is generally unavoidable. In particular for 4D NMR experiments (*10–13*), truncation can be very severe because the acquisition times are frequently much shorter than T_2. In several of the indirectly detected dimensions, commonly only 8 or 16 increments are used. In these cases, it becomes crucial to avoid truncation and enhance resolution in several dimensions simultaneously. Previously, we used a sequence consisting of Fourier transformation (t_2), 1D linear prediction (t_1), Fourier transformation (t_1), inverse Fourier transformation (t_2), linear prediction (t_2), and Fourier transformation (t_2) to accomplish this (*11*). The use of such an elaborate procedure is necessitated by the fact that the number of frequency components present in each 1D cross section through the multidimensional time-domain data generally exceeds the number of data points available

0022-2364/92 $5.00

in the cross section (*1, 9*). However, a closer study of this procedure indicates that it can cause significant lineshape distortions and artifacts. Even for noise-free real data, linear prediction can extend the time-domain accurately only if the number of frequency components is not larger than one-quarter of the available number of data points. This condition frequently is violated. Previously published methods for 2D linear prediction (*2, 4*) derive the linear prediction coefficients from 1D cross sections through the data matrix and therefore have similar limitations with regard to the maximum number of frequency components in the spectrum. Here we demonstrate a new method for linear prediction that predicts the truncated time domain in two dimensions simultaneously and that does not suffer from the problem mentioned above.

Our 2D linear prediction (LP) method is closely analogous to the 1D method, but differs in several points. Before discussing the 2D technique, we first briefly outline the 1D procedure. In 1D linear prediction, a time-domain series consisting of K exponentially damped sinusoids, represented by N real data points, $x(1), \ldots, x(\mathrm{N})$, can be extended by assuming that each data point can be expressed as a linear combination of M ($M \geqslant 2K$) previous ones:

$$x(n) = \sum_{k=1}^{M} c_k x(n-k). \qquad [1]$$

The linear prediction coefficients, c_k, can be determined by solving the $N - M$ equations of type [1] (provided $N - M > M$). If the total number of data points is a few hundred or less, this least-squares problem can be solved very efficiently by singular-value decomposition. Once the linear prediction coefficients are known, the future of the time domain can be calculated in a stepwise manner, using Eq. [1]. Alternatively, the frequencies and damping factors of the decaying sinusoids may be determined by calculating the roots of the polynomial,

$$z^M + c_1 z^{M-1} + \cdots + c_M = 0, \qquad [2]$$

where the time-domain signal is given by

$$x(n) = \sum_{k=1}^{M} A_k \exp(i\phi_k)(z_k)^n \qquad [3]$$

and A_k is the amplitude of the kth sinusoid and ϕ_k its phase. For real data, the roots occur as complex conjugate pairs, and one frequency component accounts for two of the polynomial roots.

In the application of linear prediction to multidimensional NMR, we usually do not attempt to calculate the amplitudes and phases of the NMR signals, using Eq. [3]. In the presence of noise, we find that such a procedure frequently leads to results that are less accurate than those obtained by peak picking of the Fourier transform of the time-domain data which have been extended by a modest number ($\leqslant N$) of predicted data points. Note that in the latter application the predicted data points are strongly attenuated by the use of digital filtering; i.e., small errors in the linear prediction coefficients do not give rise to dramatic distortions. Rooting of polynomial Eq. [2] is desirable, however, to allow root reflection of frequency components with $|z_k| > 1$,

i.e., of signal components that increase with time (9). Replacing these z_k values by $z_k/|z_k|^2$ and recalculation of the prediction coefficients ensure that the predicted portion of the signal is damped.

In the next section, we will describe how true 2D LP can be applied to a time-domain matrix of $N_1 \times N_2$ real data points. In our new 2D LP method, Eq. [1] is replaced by

$$x(m,n) = \sum_{k=1}^{M} \sum_{l=1}^{P} c_{kl} x(m-k, n-l) \qquad [4a]$$

or

$$x(m,n) = \sum_{k=1}^{M} \sum_{l=0}^{P-1} c_{kl} x(m-k, n-l). \qquad [4b]$$

The coefficients, c_{kl}, are obtained from these linear equations using singular-value decomposition, in exactly the same manner as mentioned above for the 1D LP case. The total number of frequency components in the 2D time-domain signal now must be smaller than $(M \times P)/4$. The number of equations, $(N_1 - M) \times (N_2 - P)$, must be larger than the number of unknown coefficients ($M \times P$ for Eq. [4a] and $M(P + 1)$ for Eq. [4b]). The relation between $x(m, n)$ and the matrix with data points $x(m - k, n - l)$ is graphically illustrated in Fig. 1. When using Eq. [4a], the matrix can be extended in both the horizontal and the vertical dimensions with a single set of linear prediction coefficients, c_{kl} (Fig. 1a). When using Eq. [4b], the matrix can be extended only in the vertical dimension (Fig. 1b), and a second matrix of prediction coefficients must be calculated for linear prediction in the horizontal dimension. The latter solution is slightly more robust for low signal-to-noise data and is used in the present work; a detailed comparison of the two methods will be presented elsewhere.

With an $M \times P$ prediction matrix, **C**, we can only predict data points $x(m, n)$ for which $m > M$ and $n > P$ (Eq. [4a]) or $n > P$ (Eq. [4b]). As a consequence, without further information it is not possible to predict data point $x(N_1 + 1, n)$ if $n < P$ (Fig. 1a). However, extra information about the NMR signal in the indirectly detected dimension is always available, permitting the data to be extended into the negative time domain (Fig. 1c). To avoid any discontinuities between the negative time-domain data and the acquired data, special precautions must be taken. First, since the time domain is only a few milliseconds long, and signals decay by only a small fraction during this time, the damping factor (T_2) can be approximated satisfactorily by a single number for all frequency components, and the damping can be removed by multiplication of the time domain with an increasing exponential, prior to extending the data into the negative time dimension. Second, the phase of the signal is generally known accurately and can be adjusted experimentally to zero at time zero. Using the fact that cosine signals are symmetric about time zero, and assuming that sampling is delayed by exactly half a dwell time, the acquired data matrix may be extended into the negative time domain [$x(-n + 1) = x(n), n = 1, 2, \cdots$] without any discontinuity in the phase of the signal. The shape of the extended matrix is depicted in Fig. 1c.

As mentioned above, in 1D linear prediction, the stability of the procedure can be improved by root reflection. In the 2D method described here, this procedure cannot

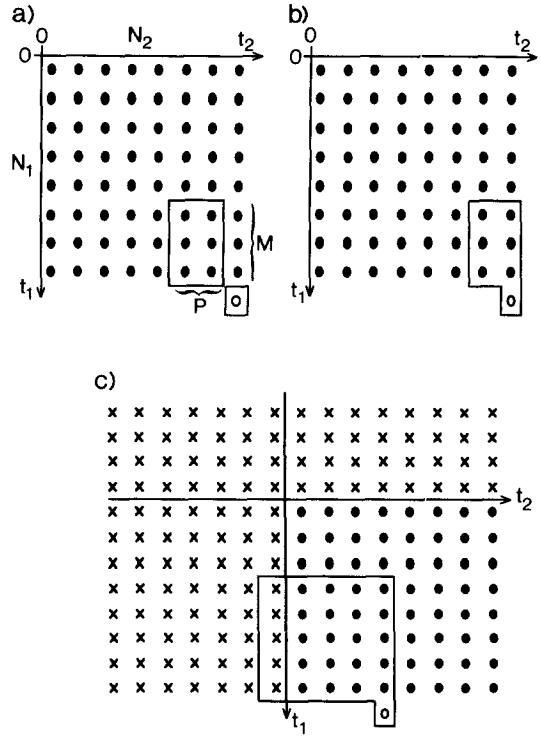


FIG. 1. Shape of the $N_1 \times N_2$ data matrix used in the 2D linear prediction procedure. Solid circles correspond to acquired data points; open circles indicate predicted data. The predicted data are calculated from the adjacent enboxed acquired data points, using linear prediction Eq. [4a] (a) or [4b] (b, c). Data points marked "×" in (c) have been obtained by reflecting the acquired data into the negative time domain.

easily be applied. Instead, we use a simple alternative approach: If a predicted data point, $x'(m, n)$, has an absolute value larger than the largest acquired data point, $|x_{max}|$, its magnitude is decreased according to

$$x(m, n) = x^2_{max}/x'(m, n). \quad [5]$$

Although, in principle, such a simple reflection procedure might be expected to yield artifacts for 2D time-domain signals that contain closely spaced antiphase resonances, in practice this does not appear to be a problem for the NOE data sets for which we have used it. In the absence of the reflection procedure, the predicted data frequently have higher noise levels, particularly for those 2D cross sections in which there are no signals.

For simplicity, the 2D LP procedure has been discussed for real data only. Because of the desirable folding properties of complex data (*14*) we strongly prefer to acquire the data in the hypercomplex format (*15*). Although somewhat more tedious to program, in the examples illustrated below, the linear prediction has been applied in this hypercomplex manner. For the case where sampling is started at time zero, this hypercomplex approach gives results that are identical to those obtained by converting

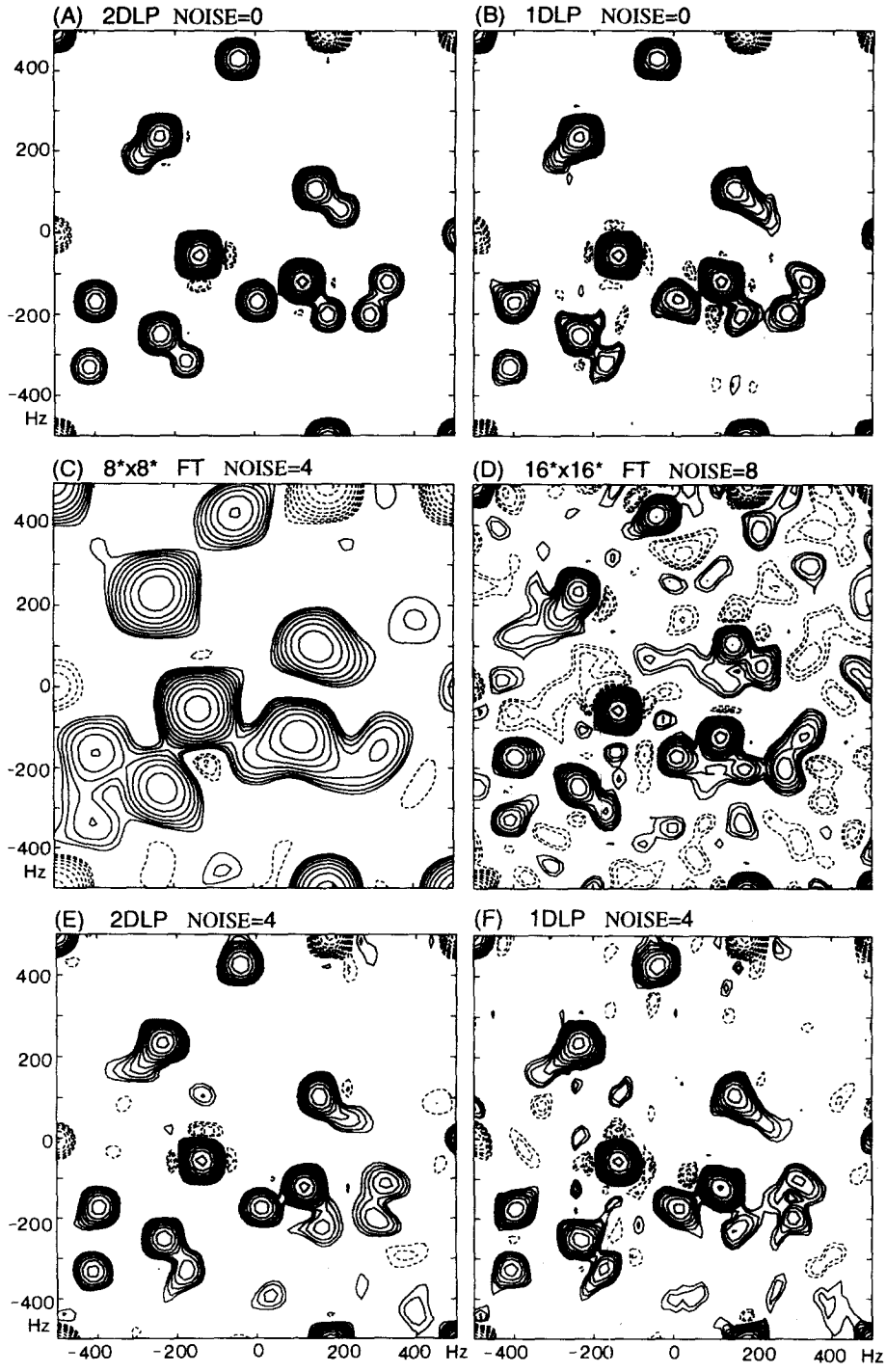
the hypercomplex $N \times N$ signal into a real $2N \times 2N$ time-domain signal, followed by the real 2D LP procedure, outlined above. Note that if sampling is not started at time zero, complex time-domain data cannot readily be converted to real data.

Figure 2 shows the application of 2D LP to a synthesized 8* × 8* time-domain signal, where 8* refers to eight complex data points. In the absence of noise, 2D LP (Fig. 2A) yields results that are indistinguishable from the Fourier transform of the 16* × 16* noiseless synthesized data (not shown). In contrast, the combination of Fourier transformation, 1D linear prediction, and inverse Fourier transformation is not capable of generating an accurately linear predicted data set (Fig. 2B). The reason for this is that the number of frequency components, in any cross section to which 1D linear prediction is applied, is larger than one-quarter the number of data points in this cross section.

Figure 2C shows the Fourier transform of the 8* × 8* matrix in the presence of white noise, with no linear prediction. Figure 2D shows the Fourier transform for a synthesized 16* × 16* matrix, with twofold lower S/N (note that four times the number of experiments would have to be conducted to generate the larger data matrix, resulting in the twofold lower S/N). Two-dimensional LP can be successfully used on the noise-containing 8* × 8* matrix (Fig. 2E), resulting in a spectrum that is of higher quality than either of the spectra in Figs. 2C and 2D. Although the combination of two 1D linear predictions with Fourier transform and inverse Fourier transform (Fig. 2F) improves the spectrum over that in either Fig. 2C or Fig. 2D, the quality is significantly lower than that obtained with 2D LP.

As a final example we illustrate the 2D LP method for experimental data. Figure 3A shows an F_1/F_3 cross section through the $^{13}C/^{13}C$-separated 4D NOESY spectrum (*10, 13*) of ^{13}C-enriched calmodulin complexed with a 26-residue unlabeled peptide. The spectrum was recorded at 600 MHz ^1H frequency, using a 1.5 mM sample concentration and a total accumulation time of 1.5 days. The size of the acquired time-domain matrix was 8*(^{13}C, t_1) × 16*(^1H, t_2) × 8*(^{13}C, t_3) × 512*(^1H, t_4). The cross section has been taken perpendicular to the $^1H/^1H$ NOESY planes, at ^1H coordinates F_2 = 0.85 ppm and F_4 = 0.43 ppm. Consequently, this cross section shows the cross peaks between the ^{13}C nuclei that are attached to the pairs of protons that contribute to the 0.85/0.43 ppm NOESY cross-peak density. The contour level is taken at 0.3% of the intensity of the diagonal methyl resonances, and the tails of these nearby intense

FIG. 2. Spectra obtained from simulated time-domain data. The simulated time domain consists of 8* × 8* data points and is extended to 16* × 16* (in the positive time domain) by linear prediction. Cosine-squared bell filtering (with the null adjusted at data point 17) and zero filling to 64* are used in both dimensions prior to Fourier transformation. Time domain signals range in amplitude from 1 to 10. The peak-to-peak amplitude of the time-domain noise equals 4. Broken contour lines correspond to negative intensity. (A) In the absence of noise, using 2D LP with a 6* × 6* prediction matrix, on a data matrix quadrupled in size (to 16* × 16*) by reflecting about time zero (Fig. 1C). Data corresponding to negative time are discarded after 2D LP, prior to 2D FT. (B) In the absence of noise, using 1D LP (six coefficients) in each dimension, with Fourier transformation and inverse Fourier transformation in the orthogonal dimension, as described in the text. (C) Fourier transform of the original 8* × 8* matrix in the presence of noise. (D) Fourier transform of the simulated 16* × 16* matrix, with twice the noise amplitude in (C). (E) Using 2D LP in the presence of noise [the same noise level as that used for (C)], generated in the same manner as (A). (F) 1D LP in the presence of noise, using the same scheme as that used for (B).


(A) 2DLP NOISE=0
(B) 1DLP NOISE=0
(C) 8*x8* FT NOISE=4
(D) 16*x16* FT NOISE=8
(E) 2DLP NOISE=4
(F) 1DLP NOISE=4
400
200
0
-200
-400
Hz
-400 -200 0 200 400 Hz

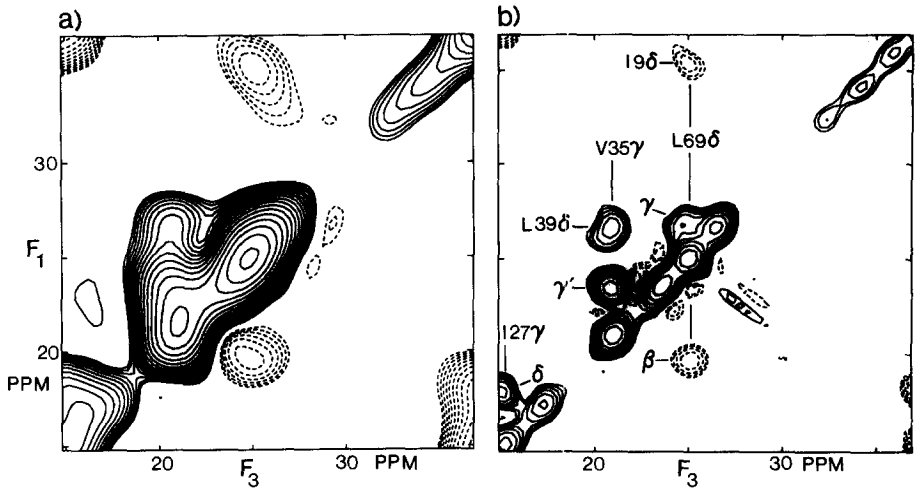


FIG. 3. Two-dimensional F_1/F_3 cross section (at F_2/F_4 = 0.85/0.43 ppm) through the 4D $^{13}C/^{13}C$-separated NOESY spectrum of calmodulin, complexed with a 26-residue peptide. (a) In the absence of linear prediction, using cosine-squared bell filtering and zero filling from 8* × 8* to 64* × 64* prior to 2D FT. (b) With 2D LP, using a 6* × 6* prediction matrix and extending the data matrix into the negative time domain (resulting in 16* × 16*) prior to 2D LP. After 2D LP the negative time-domain data are discarded, and the remaining 16* × 16* is subjected to cosine bell filtering and zero filling to 64* × 64*, prior to 2D FT. Except for the I9δ/L69δ and V35γ/L39δ cross peaks, all cross peaks correspond to intraresidue interactions. The artifact near F_3/F_1 = 28/23 ppm results from the fact that in the 2D LP procedure no proper root reflection can be applied. Broken contour lines correspond to negative intensity.

diagonal peaks give rise to the spurious diagonal resonances observed in Fig. 3. Figure 3a has been obtained without linear prediction and Fig. 3b illustrates the resolution enhancement obtainable with 2D linear prediction. Prior to calculating the mirror image data, signal decay was removed by multiplying the time domain with $e^{[(t_1+t_3)/T_2]}$, with T_2 = 20 ms and $t_{1_{max}} = t_{3_{max}}$ = 2.4 ms. The length of the time domain was extended to twice its original value (4.8 ms) in both the t_1 and the t_3 dimensions. The improved resolution afforded by the linear prediction clearly reveals six NOE interactions in this cross section, in addition to nearly a dozen spurious diagonal peaks resulting from the nearby intense diagonal peaks in the $^1H/^1H$ planes.

The 2D LP method works very well, provided that the data are severely truncated in both dimensions. This means that it is only of real use for the analysis of 3D and 4D data sets, where this situation commonly occurs. The price to be paid for 2D LP is the enormous amount of computing time needed. For example, in the present case, where the method is applied to an 8* × 8* data set, 36 hypercomplex coefficients need to be calculated from 220 equations, requiring 19 seconds on an IBM 6000/530 workstation. This process is repeated twice, to obtain predicted data in both dimensions (Eq. [4b]). Application of this type of procedure to a 16* × 16* acquired data matrix, using a 12* × 12* prediction matrix, requires about 20 minutes. Considering that this type of procedure must be applied to all cross sections through the 3D or 4D data set, the required computational time presently limits the general applicability of the 2D LP method. The large amount of computation time required also presents the

main stumbling block for extending the procedure outlined here to three or more dimensions. However, extrapolating the increase in computer power witnessed over the past decade, it may be anticipated that these applications will become feasible in the foreseeable future.

ACKNOWLEDGMENTS

We thank Mitsuhiko Ikura for providing the assignments for the cross section shown in Fig. 3 and Stephan Grzesiek and Geerten Vuister for useful comments during the preparation of the manuscript. G.Z. is supported by a fellowship from the Cooperative Graduate Program in Biophysics, sponsored by the Foundation for Advanced Education in the Sciences and the Graduate School of the University of Maryland at College Park. This work was supported by the Intramural AIDS-targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

REFERENCES

1. A. E. SCHUSSHEIM AND D. COWBURN, *J. Magn. Reson.* **71,** 371 (1987).
2. J. GORCESTER AND J. H. FREED, *J. Magn. Reson.* **78,** 292 (1988).
3. C. F. TIRENDI AND J. F. MARTIN, *J. Magn. Reson.* **81,** 577 (1989).
4. H. GESMAR AND J. J. LED, *J. Magn. Reson.* **83,** 53 (1989).
5. J. HOCH, *in* "Methods in Enzymology" (N. Oppenheimer and T. L. James, Eds.), Vol. 176, p. 216, Academic Press, San Diego, 1989.
6. Y. ZENG, J. TANG, C. A. BUSH, AND J. R. NORRIS, *J. Magn. Reson.* **83,** 473 (1989).
7. J. L. LED AND H. GESMAR, *J. Biomol. NMR* **1,** 237 (1991).
8. E. T. OLEJNICZAK AND H. L. EATON, *J. Magn. Reson.* **87,** 628 (1990).
9. G. ZHU AND A. BAX, *J. Magn. Reson.* **90,** 405 (1990).
10. G. M. CLORE, L. E. KAY, A. BAX, AND A. M. GRONENBORN, *Biochemistry* **30,** 12 (1991).
11. L. E. KAY, M. IKURA, G. ZHU, AND A. BAX, *J. Magn. Reson.* **91,** 422 (1991).
12. L. E. KAY, G. M. CLORE, A. BAX, AND A. M. GRONENBORN, *Science* **249,** 411 (1990).
13. E. R. P. ZUIDERWEG, A. M. PETROS, S. W. FESIK, AND E. T. OLEJNICZAK, *J. Am. Chem. Soc.* **113,** 370 (1991).
14. A. BAX, M. IKURA, L. E. KAY, AND G. ZHU, *J. Magn. Reson.* **91,** 174 (1991).
15. D. J. STATES, R. A. HABERKORN, AND D. J. RUBEN, *J. Magn. Reson.* **48,** 286 (1982).