Journal of Biomolecular NMR, 5 (1995) 245–258
ESCOM

J-Bio NMR 242

Theory and application of the maximum likelihood principle to NMR parameter estimation of multidimensional NMR data

Roger A. Chylla and John L. Markley*

Department of Biochemistry, University of Wisconsin-Madison, 420 Henry Mall, Madison, WI 53706, U.S.A.

Received 2 August 1994
Accepted 4 October 1994

Keywords: Maximum likelihood; Data processing; Least-squares analysis; Parameter extraction; Model fitting

Summary

A general theory has been developed for the application of the maximum likelihood (ML) principle to the estimation of NMR parameters (frequency and amplitudes) from multidimensional time-domain NMR data. A computer program (ChiFit) has been written that carries out ML parameter estimation in the D − 1 indirectly detected dimensions of a D-dimensional NMR data set. The performance of this algorithm has been tested with experimental three-dimensional (HNCO) and four-dimensional (HN(CO)-CAHA) data from a small protein labeled with ^{13}C and ^{15}N. These data sets, with different levels of digital resolution, were processed using ChiFit for ML analysis and employing conventional Fourier transform methods with prior extrapolation of the time-domain dimensions by linear prediction. Comparison of the results indicates that the ML approach provides superior frequency resolution compared to conventional methods, particularly under conditions of limited digital resolution in the time-domain input data, as is characteristic of D-dimensional NMR data of biomolecules. Close correspondence is demonstrated between the results of analyzing multidimensional time-domain NMR data by Fourier transformation, Bayesian probability theory [Chylla, R.A. and Markley, J.L. (1993) *J. Biomol. NMR*, **3**, 515–533], and the ML principle.

Introduction

Extraction of relevant parameters (e.g., amplitudes and frequencies) from NMR time-domain data has been accomplished conventionally by examination of the absorption spectrum after Fourier transformation (FT). While extremely successful, the estimation of NMR parameters by traditional Fourier methods has two salient drawbacks: inaccurate amplitude estimation in overlapped regions of the frequency spectrum and poor frequency resolution when the sampled acquisition time is significantly shorter than the half-time for decay of the measured signals (truncation). The development of 3D and 4D NMR experiments with short acquisition periods along the indirectly detected dimensions has enhanced the truncation problem. In many 4D experiments, for example, as few as eight complex points may be acquired along one or more of the indirectly detected dimensions.

One approach to alleviating the poor frequency resolution inherent in the Fourier transform of truncated data has been to extrapolate the free induction decay (FID) by linear prediction (Barkhuijsen et al., 1985) prior to FT (Gesmar and Led, 1989; Zhu and Bax, 1990; Led and Gesmar, 1991; Kay et al., 1992). Although numerically stable and computationally efficient, linear prediction is sensitive to noise, can predict only a limited number of additional data points, and is reliable only if the number of frequency components is less than one quarter of the number of available data points (Kumaresan and Tufts, 1982).

*To whom correspondence should be addressed.

Abbreviations: FT, Fourier transformation; ML, maximum likelihood; MLD, minimum description length; FID, free induction decay.

Software for carrying out the multidimensional ML estimation is available from the National Magnetic Resonance Facility software section of the Internet GOPHER utility at gopher://gopher.nmrfam.wisc.edu/11/Software/Chifit.

Bayesian probability theory in one (Bretthorst, 1990a–c) and multiple dimensions (Chylla and Markley, 1993) offers an alternative approach to NMR parameter estimation. A general D-dimensional algorithm has been developed (Chylla and Markley, 1993) which extrapolates time-domain data along the constant-time dimensions of multidimensional NMR experiments. NMR data acquired with constant-time acquisition periods are particularly amenable to Bayesian analysis, because the mathematical models used to describe such data need not include decay rates.

Parameter estimation by modeling the time-domain data with an analytical function is not unique to the Bayesian probability approach. The maximum likelihood principle has been applied to NMR parameter estimation in the time domain for the special cases of one (Miller and Greene, 1989) and two (Miller et al., 1993) dimensions. As with Bayesian probability theory, the systematic portions of the data are modeled in the time domain by a linear combination of exponentially decaying sinusoids. The ML principle states that the 'best' estimates for a given number of sinusoids are given by the set of parameters that minimizes the variance between the measured FID and the parameterized signal model ('least-squares' model). An attractive feature of ML estimation is that established methods exist (Marquardt, 1963; Press et al., 1988) for searching the parameter space associated with the least-squares cost function.

We present here the theory and application of the ML principle to NMR parameter estimation for the general case of D dimensions. We compare the most likely frequency estimates obtained from application of the ML principle to those obtained from conventional Fourier analysis with prior extrapolation by linear prediction. The results from this comparison establish a close correspondence between Fourier analysis, the ML principle and Bayesian probability theory as applied to frequency estimation of time-domain data. Given prior knowledge of the phase and decay rates, the maximum likelihood estimates of the frequency parameters are identical to the most probable frequencies predicted by Bayesian probability theory.

The development of a computer algorithm (ChiFit) that implements the ML procedure is described here. The utility of the algorithm and the ML method in general are demonstrated by their application to the processing of multinuclear 3D and 4D time-domain NMR data along all of the indirectly detected dimensions. A comparison is made between the frequency resolution obtained by the two approaches: FT with prior linear prediction and ML analysis.

Methods

The experimental NMR data were acquired on a Bruker AM-500 spectrometer, highly modified (Mooberry et al., 1994) to accommodate extra channels and the ability to apply gradient pulses to the sample. Published pulse and phase cycling schemes were used in collecting the 3D HNCO (Kay et al., 1990) and 4D HN(CO)CAHA (Kay et al., 1992) data. The sample was a uniformly (99%) $^{13}C/^{15}N$-labeled 2.0–2.5 mM solution of subunit *c* of the H^+-transporting F_1F_o ATP synthase, also known as ATP synthase (Girvin and Fillingame, 1993). The protein was dissolved in a chloroform:methanol:water (4:4:1) solvent mixture containing 50 mM NaCl.

Data analysis was carried out on a Silicon Graphics Indigo (SGI) R4000 workstation running under IRIX 4.0.5H, a version of the UNIX operating system. Fourier processing, linear prediction, data storage, data display and the production of HPGL (Hewlett–Packard Graphics Language) files were performed by the commercial NMR software program FELIX (version 2.3, Biosym Technologies, San Diego, CA). In each case where linear prediction was performed, the number of 'poles' was set to 1/3 of the number of experimentally measured data points. Mirror-image linear prediction (Zhu and Bax, 1990) was implemented within FELIX through the execution of specialized macros. ML analysis was performed by computer programs (CHIFITa and CHIFITp) (Chylla, 1994) written in the C programming language. HPGL files produced by FELIX were imported into the CorelDRAW (Corel Software, Salinas, CA) program. Modifications were made in CorelDRAW; the figures were then converted to PostScript files. The figures in this paper were produced by a laser printer with a PostScript interpreter.

Theory

The log-likelihood of the experimental data

As with all data acquired by digital computers, NMR data consist of a series of discretely sampled data points:

$$y \equiv \{y_1, y_2, y_3, \ldots, y_i\} \qquad 1 \le i \le N \qquad (1)$$

The starting assumption of any approach that seeks to fit experimental data to a parametric model is that each experimentally measured data point (y_i) can be expressed as the sum (y_i) of a systematic (signal) portion (f_i) and a random (noise) portion (e_i):

$$y_i = f_i + e_i \qquad (2)$$

According to Eq. 1, the residual, i.e., the difference between the data and the systematic portion ($y_i - f_i$), can be used as an estimate of the noise. Since the noise portion is random, probability theory can be used to quantify the likelihood of obtaining a given set of noise values, $e \equiv \{e_1, e_2, e_3, \ldots, e_i\}$. If the number of sampled noise points is 'large', the distribution of the noise can be described by a Gaussian probability distribution:

$$l(e) = (2\pi\sigma^2)^{-(N/2)} \exp\left[\frac{-\sum_{i=1}^{N} e_i^2}{2\sigma^2}\right] \quad (3)$$

where σ^2 is the variance of the noise (assumed to be known for now). The likelihood of the noise can be expressed alternatively as the likelihood of the model (systematic portion), conditional upon the data by substitution of the residual $(y_i - f_i)$ for e_i:

$$l(f|y) = (2\pi\sigma^2)^{-(N/2)} \exp\left[\frac{-\sum_{i=1}^{N} (y_i - f_i)^2}{2\sigma^2}\right] \quad (4)$$

Thus, an expression for the log-likelihood is obtained by taking the logarithm of both sides of Eq. 4, yielding:

$$\log[l(f|y)] = (-N/2)\log(2\pi\sigma^2) - \left[\frac{\sum_{i=1}^{N} (y_i - f_i)^2}{2\sigma^2}\right] \quad (5)$$

Equation 5 provides a theoretical basis for the so-called 'least-squares' approach to model optimization. For any given parametric model, the log-likelihood of the model is maximized when the squared sum of the difference between the model and the data is minimized. The ML estimates of the parameters contained in the model are thus simply those parameters that minimize the numerator of Eq. 5. Given a set of parameters P, the log-likelihood can be made conditional upon the data and P. Discarding any terms that are independent of P, the log-likelihood, conditional on the data and P, is given by:

$$\log[l(f|y,P)] \propto \left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{N} y_i f_i(P) - \left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^{N} f_i(P)^2 \quad (6)$$

Model for the systematic portion of multidimensional NMR data

Mathematical models for 1D time-domain NMR data are derived from the classical description of Bloch et al. (1946). In this formulation, the signal portion of the data is modeled as a linear combination of decaying sinusoids, each characterized by its amplitude, Larmor frequency, phase, and decay rate (line width). For quadrature-detected, complex data, the mathematical model describing the signal is given by

$$f_i = \sum_{j=1}^{J} A_j V_{ji}(t_i|\omega_j,\theta_j,\alpha_j) \qquad 1 \le j \le J \quad (7)$$

$$V_{ji}(t_i|\omega_j,\theta_j,\alpha_j) = e^{i(\omega t_i + \theta_j)} e^{-a_j t_i} \quad (8)$$

where J is the total number of signals, A_j is the amplitude of the jth signal, $V_{ji}(t_i|\omega_j,\theta_j,\alpha_j)$ is the jth signal function at point i, and $[\omega_j,\theta_j,\alpha_j]$ are the respective frequency, phase, and decay rate of the jth signal.

Equation 8 can also be expressed in its Euler form:

$$V_{ji}(t_i|\omega_j,\theta_j,\alpha_j) \equiv [\cos(\omega t_i + \theta_j) + \mathbf{i}\sin(\omega t_i + \theta_j)]\, e^{-a_j t_i} \quad (9)$$

which emphasizes that this model includes known prior information about the nature of quadrature detection, i.e., that the receivers of the real and imaginary channels sample the data simultaneously but are out of phase by 90°. Only trivial changes in the phase of Eq. 9 (Chylla and Markley, 1993) are necessary to describe the systematic portion of the data for different methods of quadrature. $V_{ji}(t_i,\phi_i|\omega_j,\theta_j,\alpha_j)$ as defined in Eq. 9 thus suffices to describe the signal function of any 1D exponentially decaying sinusoid.

To extend the above model to an arbitrary number of dimensions D, consider a D-dimensional, rectangular matrix of time-domain NMR data in which the number of sampled data points along each dimension $d \equiv [1,2,3,...,D]$ is defined by $n_d \equiv [n_1,n_2,n_3,...,n_D]$, respectively. The total number of data points, N, is given by:

$$N = \prod_{d=1}^{D} n_d \qquad 1 \le d \le D \quad (10)$$

The D-dimensional data set can still be viewed as a linear array of data points, but now each data point i corresponds to a receiver time along each dimension given by $\hat{t}_i \equiv [t_{i1},t_{i2},...,t_{iD}]$. The correspondence between i and $\hat{t}_i$ is determined by the method of quadrature detection (complex or real data) and the number of data points collected along each dimension (Eq. 10).

With this notation, the D-dimensional model describing the systematic portion of the data can be written as:

$$f_i = \sum_{j=1}^{J} A_j V_{ji}(\hat{t}_i|\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j) \qquad 1 \le i \le N \quad (11)$$

$$V_{ji}(\hat{t}_i|\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j) \equiv \prod_{d=1}^{D} \cos(\omega_{jd} t_{id} + \theta_{jd})\, e^{-\alpha_{jd} t_{id}} \quad (12)$$

where $[\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j]$ are vector quantities that represent the respective frequency, phase, and decay rate of signal j along dimension d. The D-dimensional signal function of Eq. 12 is thus simply the product of each of the complex 1D signal functions (Eq. 9) along dimensions $d \equiv [1,2,3,...,D]$.

Separation of the phase parameter from the signal functions

The model for the systematic portion of the data (Eq. 11) is expressed as the product of the amplitudes, A_j, and signal functions, $V_{ji}(\hat{t}_i|\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j)$. The signal functions are dependent upon the nonlinear vector parameters $[\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j]$. The equality $\cos(\omega t + \theta) = \cos(\omega t)\cos(\theta) - \sin(\omega t)\sin(\theta)$,

however, suggests that the phase parameters can be separated from the other nonlinear parameters. Substitution of this equality into the systematic model of the data yields $K \equiv 2^D$ coefficients, $B_{jk}(A_j,\hat{\theta}_j)$, and K signal functions, $U_{jki}(\hat{t}_i|\hat{\omega}_j,\hat{\alpha}_j)$, for each sinusoid in the model. The K coefficients and K basis functions are given by the D-dimensional products

$$\sum_{k=1}^{K} B_{jk}(A_j,\hat{\theta}_j) \equiv A_j \prod_{d=1}^{D} (\cos\theta_{jd} - \sin\theta_{jd}) \qquad (13)$$

and

$$\sum_{k=1}^{K} U_{jki}(\hat{t}_i|\hat{\omega}_j,\hat{\alpha}_j) \equiv \prod_{d=1}^{D} [\cos(\omega_{jd}t_{id}) + \sin(\omega_{jd}t_{id})] e^{-\alpha_{jd}t_{id}} \qquad (14)$$

To eliminate the cumbersome use of double subscripts, a single subscript m can be used to represent all $M \equiv J \times K$ possible combinations of jk where

$$m \in \bigcup_{j=1}^{J} \bigcup_{k=1}^{K} jk \qquad (15)$$

With this notation, the form of the systematic data becomes

$$f_i = \sum_{m=1}^{M} B_m(A_m,\theta_m) U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m) \qquad (16)$$

Table 1 displays the relationships between the subscripts jk and m and the form of $B_m(A_m,\theta_m)$ and $U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)$ for the special cases of D = one, two and three dimensions and J = 2 sinusoids.

Log-likelihood of the frequency and decay rate parameters

Separation of the phase and amplitude parameters from the frequency and decay rate parameters allows one to express the log-likelihood of the frequency and decay rate parameters independent of the amplitude and phase parameters. The first step in formulating such an expression is to substitute the phase-independent systematic form of the data (Eq. 16) into the log-likelihood of the amplitude, frequency, phase, and decay rate parameters (Eq. 6):

TABLE 1
COEFFICIENTS AND PHASE-INDEPENDENT BASIS FUNCTIONS FOR ONE, TWO AND THREE DIMENSIONS[a]

D	K	j	k	m	B_m	U_{mi}
1	2	1	1	1	$A_1\cos(\theta_{11})$	$\cos(\omega_{11}t_{i1})$
		1	2	2	$-A_1\sin(\theta_{11})$	$\sin(\omega_{11}t_{i1})$
		2	1	3	$A_2\cos(\theta_{21})$	$\cos(\omega_{21}t_{i1})$
		2	2	4	$-A_2\sin(\theta_{21})$	$\sin(\omega_{21}t_{i1})$
2	4	1	1	1	$A_1\cos(\theta_{11})\cos(\theta_{12})$	$\cos(\omega_{11}t_{i1})\cos(\omega_{12}t_{i2})$
		1	2	2	$-A_1\sin(\theta_{11})\cos(\theta_{12})$	$\sin(\omega_{11}t_{i1})\cos(\omega_{12}t_{i2})$
		1	3	3	$-A_1\cos(\theta_{11})\sin(\theta_{12})$	$\cos(\omega_{11}t_{i1})\sin(\omega_{12}t_{i2})$
		1	4	4	$A_1\sin(\theta_{11})\sin(\theta_{12})$	$\sin(\omega_{11}t_{i1})\sin(\omega_{12}t_{i2})$
		2	1	5	$A_2\cos(\theta_{21})\cos(\theta_{22})$	$\cos(\omega_{21}t_{i1})\cos(\omega_{22}t_{i2})$
		2	2	6	$-A_2\sin(\theta_{21})\cos(\theta_{22})$	$\sin(\omega_{21}t_{i1})\cos(\omega_{22}t_{i2})$
		2	3	7	$-A_2\cos(\theta_{21})\sin(\theta_{22})$	$\cos(\omega_{21}t_{i1})\sin(\omega_{22}t_{i2})$
		2	4	8	$A_2\sin(\theta_{21})\sin(\theta_{22})$	$\sin(\omega_{21}t_{i1})\sin(\omega_{22}t_{i2})$
3	8	1	1	1	$A_1\cos(\theta_{11})\cos(\theta_{12})\cos(\theta_{13})$	$\cos(\omega_{11}t_{i1})\cos(\omega_{12}t_{i2})\cos(\omega_{13}t_{i3})$
		1	2	2	$-A_1\sin(\theta_{11})\cos(\theta_{12})\cos(\theta_{13})$	$\sin(\omega_{11}t_{i1})\cos(\omega_{12}t_{i2})\cos(\omega_{13}t_{i3})$
		1	3	3	$-A_1\cos(\theta_{11})\sin(\theta_{12})\cos(\theta_{13})$	$\cos(\omega_{11}t_{i1})\sin(\omega_{12}t_{i2})\cos(\omega_{13}t_{i3})$
		1	4	4	$A_1\sin(\theta_{11})\sin(\theta_{12})\cos(\theta_{13})$	$\sin(\omega_{11}t_{i1})\sin(\omega_{12}t_{i2})\cos(\omega_{13}t_{i3})$
		1	5	5	$-A_1\cos(\theta_{11})\cos(\theta_{12})\sin(\theta_{13})$	$\cos(\omega_{11}t_{i1})\cos(\omega_{12}t_{i2})\sin(\omega_{13}t_{i3})$
		1	6	6	$A_1\sin(\theta_{11})\cos(\theta_{12})\sin(\theta_{13})$	$\sin(\omega_{11}t_{i1})\cos(\omega_{12}t_{i2})\sin(\omega_{13}t_{i3})$
		1	7	7	$A_1\cos(\theta_{11})\sin(\theta_{12})\sin(\theta_{13})$	$\cos(\omega_{11}t_{i1})\sin(\omega_{12}t_{i2})\sin(\omega_{13}t_{i3})$
		1	8	8	$-A_1\sin(\theta_{11})\sin(\theta_{12})\sin(\theta_{13})$	$\sin(\omega_{11}t_{i1})\sin(\omega_{12}t_{i2})\sin(\omega_{13}t_{i3})$
		2	1	9	$A_2\cos(\theta_{21})\cos(\theta_{22})\cos(\theta_{23})$	$\cos(\omega_{21}t_{i1})\cos(\omega_{22}t_{i2})\cos(\omega_{23}t_{i3})$
		2	2	10	$-A_2\sin(\theta_{21})\cos(\theta_{22})\cos(\theta_{23})$	$\sin(\omega_{21}t_{i1})\cos(\omega_{22}t_{i2})\cos(\omega_{23}t_{i3})$
		2	3	11	$-A_2\cos(\theta_{21})\sin(\theta_{22})\cos(\theta_{23})$	$\cos(\omega_{21}t_{i1})\sin(\omega_{22}t_{i2})\cos(\omega_{23}t_{i3})$
		2	4	12	$A_2\sin(\theta_{21})\sin(\theta_{22})\cos(\theta_{23})$	$\sin(\omega_{21}t_{i1})\sin(\omega_{22}t_{i2})\cos(\omega_{23}t_{i3})$
		2	5	13	$-A_2\cos(\theta_{21})\cos(\theta_{22})\sin(\theta_{23})$	$\cos(\omega_{21}t_{i1})\cos(\omega_{22}t_{i2})\sin(\omega_{23}t_{i3})$
		2	6	14	$A_2\sin(\theta_{21})\cos(\theta_{22})\sin(\theta_{23})$	$\sin(\omega_{21}t_{i1})\cos(\omega_{22}t_{i2})\sin(\omega_{23}t_{i3})$
		2	7	15	$A_2\cos(\theta_{21})\sin(\theta_{22})\sin(\theta_{23})$	$\cos(\omega_{21}t_{i1})\sin(\omega_{22}t_{i2})\sin(\omega_{23}t_{i3})$
		2	8	16	$-A_2\sin(\theta_{21})\sin(\theta_{22})\sin(\theta_{23})$	$\sin(\omega_{21}t_{i1})\sin(\omega_{22}t_{i2})\sin(\omega_{23}t_{i3})$

[a] For every D-dimensional sinusoid, indicated by the subscript j, there are ($K=2^D$) coefficients (A_m) and K phase-independent basis functions (U_{mi}). The $k \equiv [1,2,3,...,K]$ coefficients are related to the *j*th amplitude and *j*th phase of the sinusoid, as shown in the table (Eq. 14 in the text). The $k \equiv [1,2,3,...,K]$ basis functions are related to the *j*th frequency and *j*th decay rate along dimensions $d \equiv [1,2,3,...,D]$, as shown in the table (Eq. 15 in the text). Subscript m is used to represent all ($M \equiv JK$) possible combinations of the pair of subscripts jk.

$$\log[l(f|y,P)] \propto \left(\frac{1}{\delta^2}\right)\sum_{i=1}^{N}\left[y_i\sum_{m=1}^{M}\left[B_m(A_m,\hat{t}_m)U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)\right]\right] - \left(\frac{1}{2\delta^2}\right)\sum_{i=1}^{N}\left[\sum_{m=1}^{J}\left[B_m(A_m,\hat{\theta}_m)U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)\right]\right]^2 \quad (17)$$

Upon interchanging subscripts i and m, Eq. 17 can be rewritten as:

$$\log[l(f|y,P)] \propto \left(\frac{1}{\delta^2}\right)\sum_{m=1}^{M}B_m(A_m,\hat{\theta}_m)\left[\sum_{i=1}^{N}y_i\,U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)\right] - \left(\frac{1}{2\delta^2}\right)\left[\sum_{m=1}^{M}B_m(A_m,\hat{\theta}_m)\left[\sum_{i=1}^{N}U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)\right]\right]^2 \quad (18)$$

Equation 18 is a set of linear equations in which B_m are the M linear coefficients. The B_m values that maximize the log-likelihood are solved for by setting derivatives of Eq. 18 with respect to B_m equal to zero. This procedure generates a system of M linear equations that can be expressed in terms of the matrix equation

$$\sum_{m=1}^{M}G_{ml}B_l = X_m \quad (19)$$

where X_m is the projection of the data upon the basis functions $U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)$:

$$X_m \equiv \sum_{i=1}^{N}y_i U_{mi}(\hat{t}_i|\hat{\omega}_j,\hat{\alpha}_j) \quad (20)$$

and where **G** is the interaction matrix given by:

$$G_{ml} = \sum_{i=1}^{N}U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)U_{li}(\hat{t}_i|\hat{\omega}_l,\hat{\alpha}_l) \quad (21)$$

If the basis functions are sampled over a uniform time space and if the D-dimensional frequencies are well separated (differences between the frequencies along each dimension are large compared to the line widths along each dimension), then the orthogonality property of cosine and sine functions ensures that G_{ml} approaches zero for all $m \neq l$. Under these conditions, the inverse of the interaction matrix is simply the reciprocal of each of its diagonal terms, and the ML values for $B_m(A_m,\hat{\theta}_m)$ are given by

$$B_m = \frac{X_m}{G_{mm}} \quad (22)$$

Upon substituting the definitions of X_m and G_{mm} from Eqs. 20 and 21, respectively, the log-likelihood of Eq. 18 can be written as:

$$\log[l(f|y,\hat{\omega}_m,\hat{\alpha}_m)] \propto \sum_{m=1}^{M}B_m X_m - \frac{1}{2}\sum_{m=1}^{M}\sum_{l=1}^{M}B_m B_l G_{ml} \quad (23)$$

If the assumption is made again that the sinusoids are well separated, then, since G_{ml} is negligible for all $m \neq l$, Eq. 23 reduces to:

$$\log[l(f|y,\hat{\omega}_m,\hat{\alpha}_m)] \propto \sum_{m=1}^{M}B_m X_m - \frac{1}{2}\sum_{m=1}^{M}(B_m)^2 G_{mm} \quad (24)$$

Substitution of X_m/G_{mm} for B_m (Eq. 22) results in a simple expression for the log-likelihood of the frequency and decay rate parameters which is independent of the amplitude and phase parameters:

$$\log[l(f|y,\hat{\omega}_m,\hat{\alpha}_m)] \propto \frac{1}{2}\sum_{m=1}^{M}\frac{(X_m)^2}{G_{mm}} \quad (25)$$

Weighted power spectrum and weighted squared absorption spectrum as maximum-likelihood estimators of frequency parameters

The log-likelihood of the frequency and decay rate parameters has special significance in that the M projections of the data, X_m, over the M basis functions, $U_{mi}(\hat{t}_i|\hat{\omega}_m,\hat{\alpha}_m)$, are related to the complex Fourier transform of the data. For example, in one dimension (D=1, K=2) for one signal (J=1) there are M=2 basis functions and X_m is given by:

$$X_m = \begin{cases} \sum_{i=1}^{N}y_i\cos(\omega t_i)e^{-\alpha t_i} & m = 1 \\ \sum_{i=1}^{N}y_i\sin(\omega t_i)e^{-\alpha t_i} & m = 2 \end{cases} \quad (26)$$

The summations appearing in Eq. 26 are simply the cosine and sine transforms of the data weighted by an exponential filter function, the decay rate of which is matched to the decay rate of the signal. The value of G_{mm} in one dimension is given by:

$$G_{mm} = \begin{cases} \sum_{i=1}^{N}(\cos(\omega t_i+\phi_i)e^{-\alpha t_i})^2 & m = 1 \\ \sum_{i=1}^{N}(\sin(\omega t_i+\phi_i)e^{-\alpha t_i})^2 & m = 2 \end{cases} \quad (27)$$

$$\approx \frac{1}{2}\sum_{i=1}^{N}e^{-\alpha t_i}$$

As evident from Eq. 27, G_{mm} is dependent on the decay rate α, but independent of the angular frequency ω when sampled over a uniform time interval.

Given that G_{mm} is a constant for a fixed decay rate α, the log-likelihood of the frequency parameter ω (Eq. 25) in one dimension is given by

$$\log[l(f|y,\hat{\omega}_m,\hat{\alpha}_m)] \propto \frac{\left(\sum_{i=1}^{N} y_i \cos(\omega t_i + \phi_i) e^{-\alpha t_i}\right)^2 + \left(\sum_{i=1}^{N} y_i \sin(\omega t_i + \phi_i) e^{-\alpha t_i}\right)^2}{\frac{1}{4}\sum_{i=1}^{N} e^{-2\alpha t_i}} \quad (28)$$

The numerator of the right-hand side of Eq. 28, however, is simply the weighted power spectrum evaluated at the angular frequency ω. To extend this result to the general case of D dimensions, the log-likelihood of a D-dimensional frequency, $\hat{\omega} \equiv [\omega_1,\omega_2,...,\omega_D]$, with prior knowledge of the decay rates, $\hat{\alpha} \equiv [\alpha_1,\alpha_2,...,\alpha_D]$, along each dimension, is proportional to the D-dimensional weighted power spectrum of the signal evaluated at $\hat{\omega}$. This relationship is valid under three conditions: (i) when the data are sampled uniformly along each dimension; (ii) when the data are either stationary (no decay rate) or weighted by $\hat{\alpha}$; and (iii) when the signal resonating at the D-dimensional frequency $\hat{\omega}$ is well separated from other signals in the spectrum.

If both the decay rates *and* phases of each signal are known, then the phase-dependent model for the systematic data (Eq. 12) can be used. The projection of the data over the phase-dependent basis function V_{ji} (Eqs. 11 and 12) is calculated for each signal in the data. This projection is equivalent to the absorptive portion of the Fourier transform (linear combination of the real and imaginary portions of the Fourier transform weighted by $\cos\theta_j$ and $\sin\theta_j$, respectively). For the special case in which the phases along each dimension are known, the log-likelihood of a D-dimensional frequency vector $\hat{\omega}$ is given by the squared weighted *absorption* spectrum evaluated at $\hat{\omega}$.

Algorithm

The theory developed in the previous section is used here to apply the maximum likelihood principle to the tasks of signal recognition and parameter estimation in the analysis of D-dimensional time-domain data. As discussed below, a separate termination criterion must be used for the third task of model selection. The algorithm described in this section has been translated into a computer program called ChiFit (Chylla, 1994).

In this section it is assumed that the following prior information is available about the acquired NMR data: (i) the phase of signal j along dimension d (θ_{dj}), which can be calculated from its frequency along dimension d (ω_{dj}); and (ii) estimated to within a factor of two of the true value, the mean decay rate, $\overline{\alpha_d}$, along dimension d. The phase and the approximate decay rates are generally known, since both are required to perform phase-sensitive Fourier processing. The phase information is needed for performing linear phase corrections and the decay rate information is needed for constructing an optimal apodization function for digital filtering prior to Fourier transformation.

Strategy for maximum likelihood analysis of time-domain NMR data

As with all methods that seek to describe data in terms of a parametric model, the analysis of NMR data using the ML principle requires the performance of three tasks: signal recognition, parameter estimation, and model selection. Signal recognition is the process of finding the 'most likely' signals in the time-domain data. Parameter estimation is the process of assigning the 'optimal' values to the parameters describing each signal. Model selection is the process of deciding whether a particular model is more or less likely than an alternative model.

The systematic portion of the data is described by a linear combination of approximately orthonormal basis functions. A consequence of this orthonormality is that the occurrence of one signal in the data does not drastically limit the ability to reliably detect and accurately estimate the parameters of other signals in the data. The models describing the data, therefore, can be assembled incrementally, i.e., the most likely (largest) signals in the data can be added to the model first and the smaller signals obtained from the residual can be added next.

The weighted absorption spectrum, given knowledge of the phases and a rough estimate of the decay rate of the signals, is approximately proportional to the log-likelihood of the frequency parameters. The weighted absorption spectrum thus represents a search over all frequency space for the most likely frequencies contained in the data. These considerations suggest the following strategy for performing ML analysis of D-dimensional time-domain data:

(i) Compute the difference between the measured data (y_i) and the current systematic portion of the data ($f_i(P)$), i.e., the residual. If the model contains no signals (J=0), then the residual is identical to the measured data.

(ii) Use prior information about the average decay rate and the phase correction along each of the D dimensions to compute the D-dimensional, zero-filled, exponentially weighted absorption spectrum of the residual. Select a suitable threshold and locate the positions of local maxima in the square of the absorption spectrum above this threshold to obtain both the number of new signals (ΔJ) and initial frequency estimates for each putative signal j to be added to the current model.

(iii) Use prior knowledge of the average decay rate and phase along dimension d, together with frequency estimates obtained from step (ii), to compute initial estimates for the amplitude (A_j) of signal j from the projection of the time-domain data over its respective signal function (V_j, Eq. 12).

(iv) Perform a nonlinear least-squares optimization of *all* of the amplitude, frequency and decay rate parameters, starting from the initial estimates obtained from prior knowledge of the phases, the values for each of the parameters of the current model, and the estimates for the putative parameters obtained from steps (ii) and (iii).

(v) Use some type of termination criterion to decide whether the putative model containing (J + ΔJ) signals and (P + ΔP) parameters ($f_i(P + \Delta P)$) is more or less likely than the current model containing J signals and P parameters ($f_i(P)$). If the putative model is 'more likely' than the current model, then replace $f_i(P)$ with $f_i(P + \Delta P)$ and proceed to step (i) to execute an additional cycle of the algorithm. If the putative model is less likely than the current model, then terminate the algorithm.

Each of these steps is explained in more detail below.

Computation of the D-dimensional weighted absorption spectrum of the residual

For a discretely sampled vector of N data points, $y = [y_1, y_2, \ldots, y_N]$, the discrete Fourier transform F(y,k) at a frequency of ω_k is given by the summation

$$F(y|\omega_k) = \sum_{i=1}^{N} y_i[\cos(\omega_k t_i) - i\sin(\omega_k t_i)] \qquad (29)$$

If the phase of the signal ϕ_k resonating at frequency ω_k is known, then the real absorptive portion of the transform $A(y|\omega_k,\theta_k)$ can be calculated from the appropriate combinations of the cosine and sine transforms:

$$\begin{aligned} A(y|\omega_k,\theta_k) = {} & \cos(\phi_k)\sum_{i=1}^{N} y_i \cos(\omega_k t_i) \\ & - \sin(\phi_k)\prod_{i=1}^{N} y_i \sin(\omega_k t_i) \end{aligned} \qquad (30)$$

The zero-padded absorption spectrum s(y) is represented by a series of summations given by Eq. 30 over an angular frequency grid of 2N points, covering the range $-\pi \le \omega < \pi$ (complex data):

$$s(y) = \bigcup_{k=1}^{2N} A(y|\omega_k,\theta_k) \qquad -\pi \le \omega_k \le \pi \qquad (31)$$

The form of Eq. 31 suggests that the calculation time of the discrete Fourier spectrum is proportional to $2N^2$. The use of the fast Fourier transform (FFT) on time-domain arrays that have lengths that are even powers of 2, however, allows calculation of Eq. 31 in a time proportional to $2N\log_2 N$.

When an exponential filter is applied to the data prior to zero-filling, Fourier transformation and phase correction, the result is the zero-filled, weighted absorption spectrum:

$$s(y|\alpha) = \bigcup_{k=1}^{2N}\left[\sum_{i=1}^{N} y_i[\cos(\phi_k)\cos(\omega_k t_i) - \sin(\phi_k)\sin(\omega_k t_i)]e^{-\alpha t_i}\right] \qquad (32)$$

The D-dimensional form of the real absorptive portion of the weighted Fourier transform is simply the summation of the data points over the product of the phase-corrected, exponentially weighted cosine and sine transforms along each dimension:

$$A(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha}) = \sum_{i=1}^{N} y_i\left[\prod_{d=1}^{D}[\cos(\theta_{kd})\cos(\omega_{kd}t_{id}) - \sin(\theta_{kd})\sin(\omega_{kd}t_{id})]\right] e^{-\alpha_d t_{id}} \qquad (33)$$

Because the basis functions along each dimension are approximately orthonormal, the summation of the data over the product of the basis functions can be calculated as the product of the summations along each dimension. In the ChiFit algorithm, the D-dimensional absorption spectrum is calculated as D consecutive series of one-dimensional transforms, given by Eq. 32.

Signal recognition from the absorption spectrum

Given knowledge of the phase and decay rates, the D-dimensional, exponentially weighted, absorptive portion of the Fourier transform $A(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha})$ is proportional to the log-likelihood of a D-dimensional frequency $\hat{\omega}_k$. The D-dimensional weighted absorption spectrum $s(y|\hat{\alpha})$ can thus be viewed as a search over the entire D-dimensional frequency space for the most likely frequencies contained within the data. If there are no signals in the data, i.e., if the FID represents purely random noise, the average power of the noise is estimated from the mean squared magnitude of the absorption spectrum $\bar{s}^2(y|\hat{\alpha})$, where

$$\bar{s}^2(y|\hat{\alpha}) = \frac{1}{2N}\sum_{k=1}^{2N} A^2(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha}) \qquad (34)$$

It is logical to assume, therefore, that sufficient evidence for a signal resonating at a particular D-dimensional frequency $\hat{\omega}_k$ exists if $A(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha})$ is significantly larger than $\bar{s}^2(y|\hat{\alpha})$.

The ChiFit algorithm uses a threshold criterion to determine whether or not $s(y|\hat{\alpha})$ contains evidence for a signal resonating at $\hat{\omega}_k$. A maximum at $\hat{\omega}_k$ is added to the putative model if the inequality

$$A^2(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha}) > \sigma^2\bar{s}^2(y|\hat{\alpha}) \qquad (35)$$

is satisfied, where σ is a constant set at 5. Definitive evidence for $\hat{\omega}_k$ is considered to exist when σ=5. If Eq. 35

is not satisfied for all $\hat{\omega}_k$, then the ChiFit algorithm adds one putative signal to the model, resonating at $\hat{\omega}_{max}$, the frequency at which $A^2(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha})$ reaches a global maximum.

If Eq. 35 with $\sigma=5$ is satisfied for more than one value of $\hat{\omega}_k$, an additional criterion must be met before a signal resonating at $\hat{\omega}_k$ is added to the model. One of the assumptions needed to equate $A^2(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha})$ with the log-likelihood of the frequency parameters was that each of the signals is well separated from other signals in the model. ChiFit employs a local maximum test to determine whether a frequency is well separated from other putative frequencies in the residual. A signal resonating at a frequency $\hat{\omega}_k$ that satisfies the threshold criterion of Eq. 35 is added to the model if $A^2(y|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha})$ is the largest value within a D-dimensional rectangular region of four grid points wide, relative to the position of $\hat{\omega}_k$.

Obtaining initial estimates for the amplitudes

The ΔJ putative signals added to the model in step (ii) of the algorithm are assigned decay rates equal to the average decay rate and phases as determined by the known frequency dependence of the phase and the frequency estimates along each dimension. The frequency, phase, and decay rate estimates obtained for each signal are sufficient parameters to calculate the basis functions given by Eq. 12. Amplitudes that satisfy the linear matrix equation

$$\sum_{j=1}^{J} H_{jk}A_k = L_j \tag{36}$$

are assigned to each of the ΔJ signals. In Eq. 36, L_j represents the projection of the data over the jth basis function V_{ji} (Eq. 12):

$$L_j = \sum_{i=1}^{N} y_i V_{ji}(\hat{t}_i|\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j) \tag{37}$$

and **H** is the interaction matrix defined by

$$H_{jk} = \sum_{i=1}^{N} V_j(\hat{t}_i|\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j)\ V_k(\hat{t}_i|\hat{\omega}_k,\hat{\theta}_k,\hat{\alpha}_k) \tag{38}$$

Vector A in Eq. 36 is solved for by inverting the interaction matrix by Gauss–Jordan elimination (Press et al., 1988), followed by forming its matrix dot product with the L projection vector.

Nonlinear least-squares optimization of the frequencies, amplitudes and decay rates

The maximum likelihood estimates for the frequencies, amplitudes and decay rates are obtained from nonlinear least-squares optimization of Eq. 6 by a modified Levenberg–Marquardt method (Marquardt, 1963). The method is iterative and, starting from initial estimates for each of the parameters of the model, it calculates parameter increments that lower the variance of the residual, Ψ^2, where

$$\begin{aligned}\Psi^2 &= \sum_{i=1}^{N}(y_i - f_i)^2 \\ &= \sum_{i=1}^{N}\left[y_i - \sum_{j=1}^{J} A_j V_{ji}(\hat{t}_i,\hat{\phi}_i|\hat{\omega}_j,\hat{\theta}_j,\hat{\alpha}_j)\right]\end{aligned} \tag{39}$$

The parameter increments are derived from calculations of the first and second derivatives with respect to each of the parameters $[A_j,\hat{\omega}_j,\hat{\alpha}_j]$. The algorithm terminates when two sets of parameter estimates yield Ψ^2 values that differ by less than 0.1%.

Assuming that the phases are known (calculated or determined from manual phasing of the Fourier transform), each sinusoid is defined by one amplitude and in each dimension by a frequency and a decay rate. Thus, a D-dimensional model containing J sinusoids is characterized by Q=J(1+2D) parameters. Fewer parameters are required if one or more of the dimensions is recorded with a constant-time acquisition period which does not require a decay rate parameter. The partial derivatives of Ψ^2 with respect to each of the Q parameters P_q are calculated from the projections of the residual against the analytical derivatives of f_i with respect to P_q:

$$\frac{\partial\Psi^2}{\partial P_q} \propto \sum_{i=1}^{N}(y_i - f_i)\frac{\partial f_i}{\partial P_q} \qquad P_q \in [A_j,\hat{\omega}_j,\hat{\alpha}_j] \tag{40}$$

$$V_{ji} = \prod_{d=1}^{D}(V_{ji})_d \tag{41}$$

$$\frac{\partial f_i}{\partial A_j} = V_{ji} \tag{42}$$

$$\frac{\partial f_i}{\partial \omega_{jd}} = A_j\prod_{\lambda=1}^{D}(V_{ji})_\lambda \frac{\partial(V_{ji})_d}{\partial\omega_{jd}} \qquad \lambda \neq d \tag{43}$$

$$\frac{\partial f_i}{\partial \alpha_{jd}} = A_j\prod_{\lambda=1}^{D}(V_{ji})_\lambda \frac{\partial(V_{ji})_d}{\partial\alpha_{jd}} \qquad \lambda \neq d \tag{44}$$

$$\frac{\partial(V_{ji})_d}{\partial\omega_{jd}} = -\sin(\omega_{jd}t_{id} + \theta_{jd} + \phi_{id})\left(e^{-\alpha_{jd}t_{id}}\right)t_{id} \tag{45}$$

$$\frac{\partial(V_{ji})_d}{\partial\alpha_{jd}} = \cos(\omega_{jd}t_{id} + \theta_{jd} + \phi_{id})\left(e^{-2\alpha_{jd}t_{id}}\right)t_{id} \tag{46}$$

The second partial derivative of Ψ^2 with respect to the parameters P_q and P_r is calculated from the approximation

$$\frac{\partial^2(\Psi^2)}{\partial P_q\partial P_r} \approx \sum_{i=1}^{N}\left[\frac{\partial f_i}{\partial P_q}\right]\left[\frac{\partial f_i}{\partial P_r}\right] \tag{47}$$

The Levenberg–Marquardt algorithm is modified by placing constraints on the allowed values for the decay

rates. Upper and lower bounds on the decay rates are necessary because the decay rates are poorly determined by the truncated data.

Termination criterion used in model selection

The least-squares or ML criterion can be used to perform parameter estimation, given the prior ability to perform signal recognition, but it cannot be used in the task of model selection. The primary question of model selection is whether, given the data and any prior information, a particular model that contains (P+ΔP) parameters, $f(\hat{t}_i, \hat{\phi}_i|P+\Delta P)$, and yields a residual variance of $\Psi^2_{P+\Delta P}$, is more or less probable than a model containing P parameters, $f(\hat{t}_i, \hat{\phi}_i|P)$, that yields a residual variance of Ψ^2_P. The mere fact that $\Psi^2_{P+\Delta P}$ may be less than $f(\hat{t}_i, \hat{\phi}_i|P)$ is not sufficient evidence to conclude that the model containing P+ΔP parameters is the more likely model. This model has ΔP greater degrees of freedom and is thus expected to produce a lower residual variance (χ^2), regardless of whether the addition of the extra signal(s) is statistically significant.

In the algorithm described above, the task of model selection becomes one of establishing a termination criterion for deciding when to stop adding signals to the model describing the systematic portion of the data. One can imagine at least three different criteria that can be employed:

(i) Prior knowledge of the number of expected signals can be used to determine the number of sinusoids in the model.

(ii) Because the algorithm adds stronger signals with large amplitudes prior to weaker signals with small amplitudes, a minimum threshold amplitude can be employed. The algorithm can stop adding sinusoids when the amplitudes of any nascent sinusoid(s) drop below this threshold.

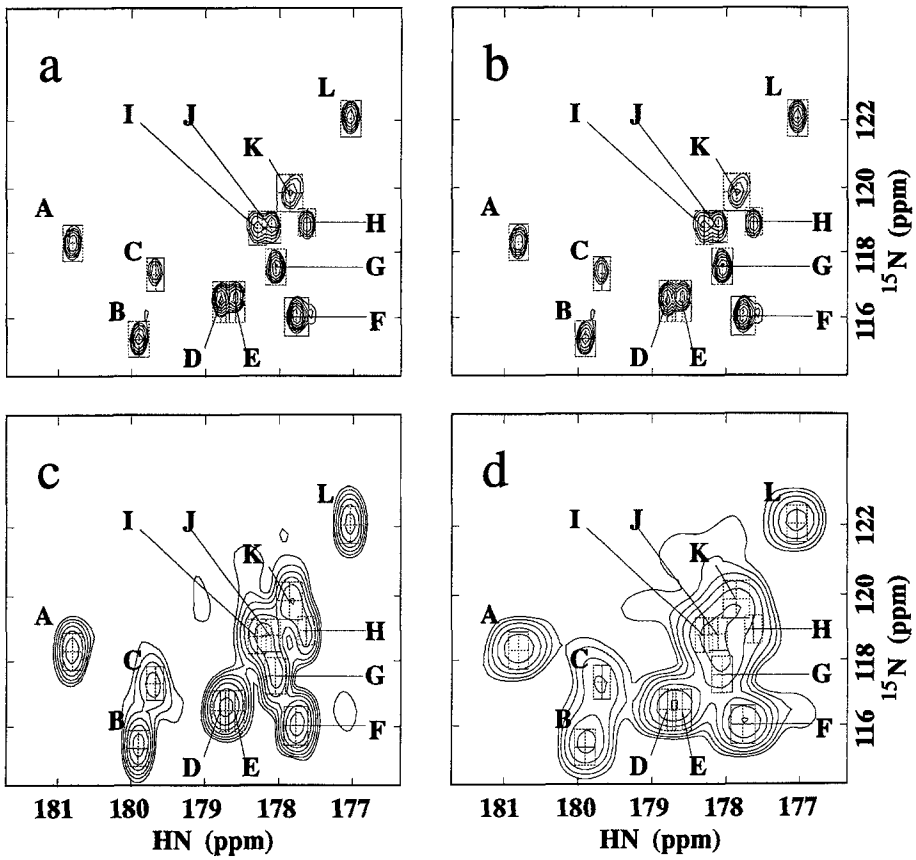

Fig. 1. Contour plots through a 3D HNCO data set processed by Fourier transformation of a 2D ^{13}C-^{15}N slice enhanced by linear prediction. The sample was a 2 mM solution of a subunit *c* of ATP synthase, labeled uniformly with ^{13}C and ^{15}N. The data, of variable resolution (as noted below), were first digitally filtered, zero-filled, Fourier transformed, and phased along the amide proton (acquisition, t_3) dimension. The two-dimensional t_1,t_2 slice was extracted from the absorptive portion of the spectrum at $\omega_3 = 8.59$ ppm. The digital resolution of the slice was (a) 128×64; (b) 64×64; (c) 64×32; and (d) 32×32 data points along the ^{13}C (t_1) and ^{15}N (t_2) dimensions, respectively. The t_1,t_2 slice was processed first by transformation of the ^{13}C dimension into the frequency (ω_1) domain. The ^{15}N (t_2) dimension was then extrapolated from n_2 to $2n_2$ points by mirror-image linear prediction, as described in the Methods section. A cosine-squared bell apodization function of width equal to $2n_2$ points was applied to the data, followed by zero-padding to a length of 64 complex points, Fourier transformation, phase correction, and discarding of the dispersion spectrum. The ^{13}CO (t_1) dimension was then transformed back into the time domain, extrapolated from n_1 to $1.5n_1$ points by linear prediction, apodized using a 32 complex point cosine-squared bell digital filter, zero-padded to a length of 128 complex points, and further processed analogously to the ^{15}N dimension. The footprints labeled A–L that are superimposed on the absorption spectra represent the widths and center positions of the most intense peaks present in the spectrum with the highest digital resolution (spectrum a).

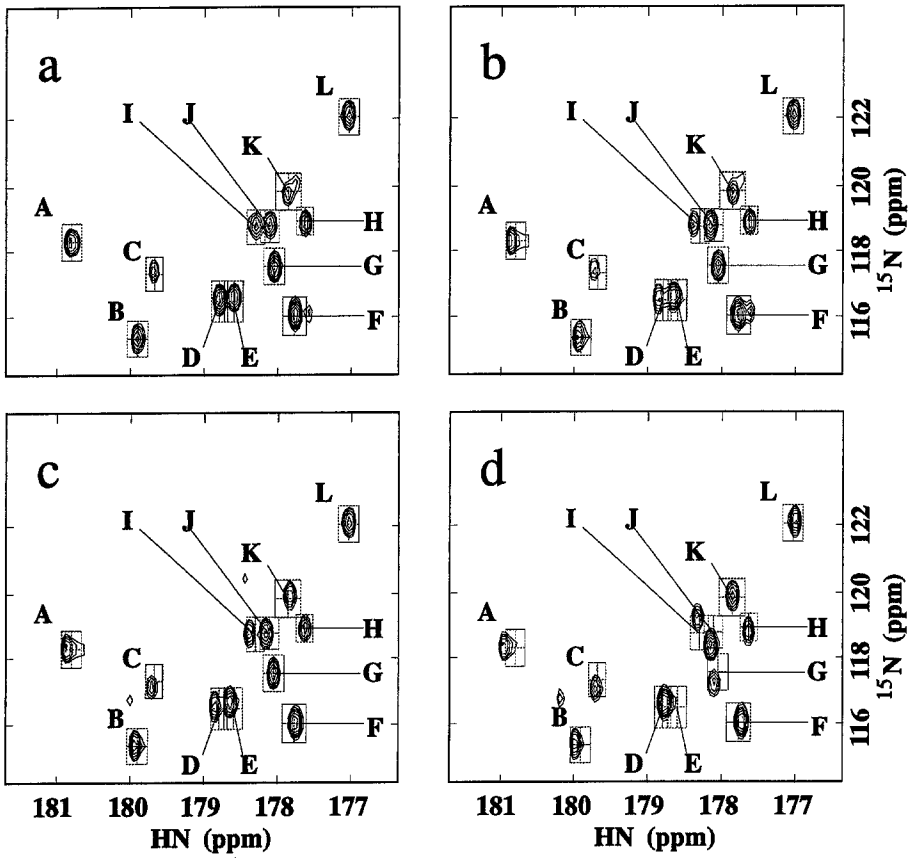

Fig. 2. Contour plots of a 2D ^{13}C-^{15}N slice, enhanced by maximum likelihood analysis. The data sets were the same ones as used in Fig. 1. The number, frequency, amplitude and decay rates of each of the signals in the 2D ^{13}C-^{15}N (t_1,t_2) slices at different digital resolution (a–d, as described in Fig. 1) were estimated by the maximum likelihood (ML) algorithm described in the text. The parameter estimates obtained from ML analysis were used to replace the experimental data with a synthetic data set of resolution 256×128 along the ^{13}C (t_1) and ^{15}N (t_2) dimensions, respectively. Each dimension was then apodized using a cosine-squared bell apodization function, zero-filled, Fourier transformed, and phased. After Fourier transformation and phase correction, the dispersive portions of each of the spectra were discarded. The footprints labeled A–L that are superimposed on the absorption spectra represent the width and center position of the most intense peaks present in the spectrum with the highest digital resolution (spectrum a).

(iii) Information theory can be used to determine whether $f(\hat{t}_i, \hat{\phi}_i|P+\Delta P)$ is a more likely model than $f(\hat{t}_i, \hat{\phi}_i|P)$. One statistic designed for this purpose is the minimum description length (MDL) criterion (Reddy and Birader, 1993):

$$MDL = N\ln(\Psi^2) + P\ln(N) \quad (48)$$

where N is the number of data points, Ψ^2 is the sum of the squares of the residual (χ^2), and P is the number of free parameters contained in the model.

ChiFit can use the third option with the ML statistic, i.e., the most likely model is the model that minimizes MDL. The ChiFit program stops adding signals to the model for the systematic portion of the data when a model $f(\hat{t}_i|P+\Delta P)$ with a χ^2 value of $\Psi^2_{P+\Delta P}$ has a higher MDL value than a model $f(\hat{t}_i|P)$ with a χ^2 value of Ψ^2_P.

Data

To compare the results of different processing techniques, synthetic data sets using exponentially decaying signals with added white noise are often chosen as example data sets. The weakness in this approach is twofold: (i) the decay rates of NMR signals are not always strictly exponential; and (ii) white noise is a simplistic model for the noise found in real NMR experiments. In addition to random noise, noise in NMR experiments arises from a variety of sources, e.g. incomplete cancellation of extraneous magnetization by phase cycling, initial time point artifacts, spectrometer instabilities, etc. Taking this into consideration, it is desirable to compare the results of applying different processing techniques to experimentally recorded data sets which have more 'realistic' properties.

To compare the frequency resolution obtainable by ML analysis to that obtainable by the present standard technique (linear-prediction-enhanced Fourier transformation), we have applied the processing methods to the enhancement of the indirectly detected dimensions of two experimental protein data sets: 3D HNCO and 4D HN(CO)CAHA data from subunit *c* of the F_1F_0 ATP

synthase (8 kDa), also known as the DDT binding protein (Girvin and Fillingame, 1993). The results of this comparison are displayed in Figs. 1 and 2. Each type of analysis was performed four times: first upon a high-resolution FID with a data point resolution of 128×64 (t_1,t_2) and subsequently on subsets of the same data set with decreasing resolution along t_1 and t_2. The rationale for processing the same data set with varying degrees of artificial truncation was to quantify how many of the signals resolved in the original data set were still resolved when both the signal-to-noise ratio and the frequency resolution of the time-domain data were decreased.

Figures 1 and 2 display contour plots of a 2D ^{13}CO-^{15}N (t_1,t_2) slice of the absorptive portion of the amide proton (acquisition) dimension at ω_3=8.59 ppm, extracted from the 3D HNCO data. Although the sample, ATP synthase, is a relatively small protein, it consists predominantly of α-helical regions and thus exhibits poor chemical shift dispersion in its amide proton and amide nitrogen chemical shifts. In Figs. 1 and 2, the resolution of the 2D FID prior to linear prediction was (a) 128×64; (b) 64×64; (c) 64×32; and (d) 32×32 data points along the carbonyl ^{13}C (t_1) and amide ^{15}N (t_2) dimensions, respectively. Figure 1 shows the implementation of the linear-prediction procedure (other processing details are given in the legend to the figure) and Fig. 2 displays an analogous series of contour plots which represent the absorption spectrum of exactly the same data sets after enhancement of the time-

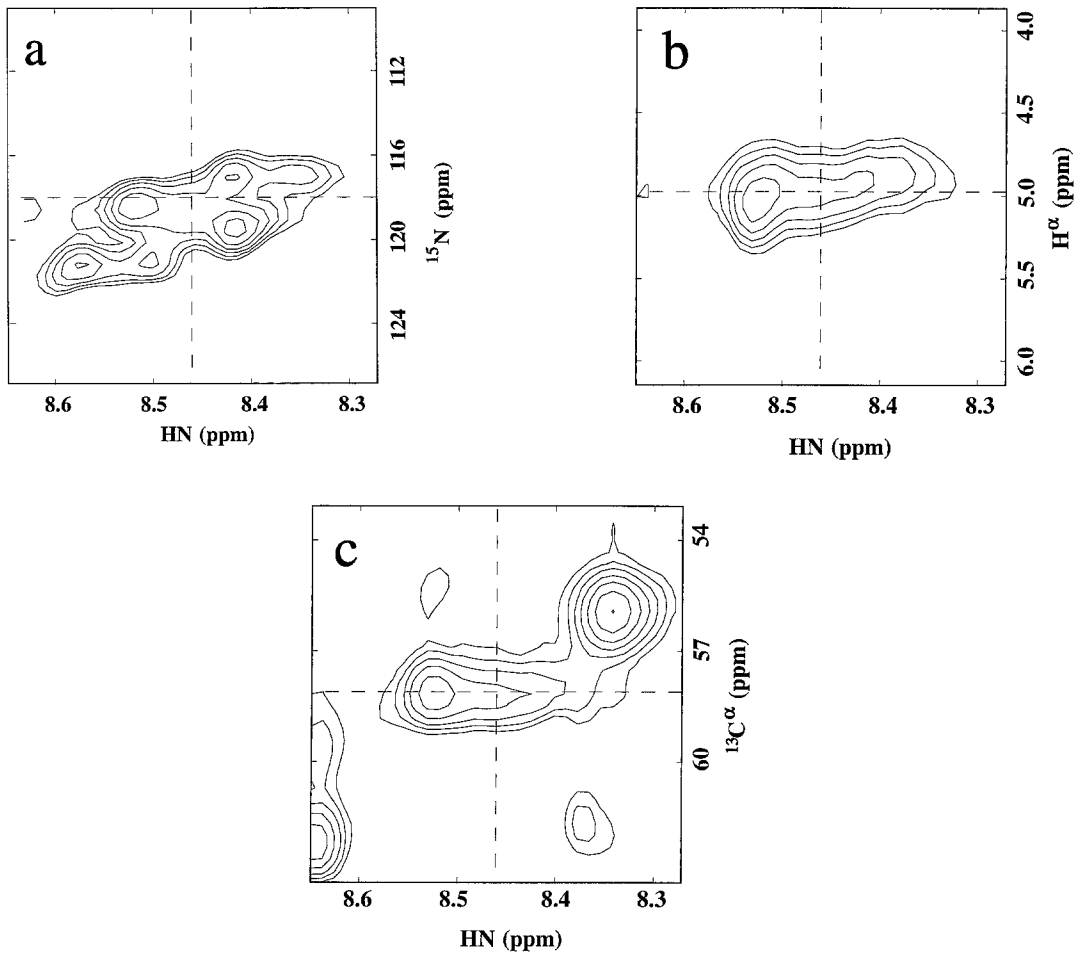

Fig. 3. Contour plots of slices through a 4D HN(CO)CAHA spectrum processed by Fourier transformation, enhanced by linear prediction. The protein sample was that described in Fig. 1. The three panels correspond to the (a) ^{1}HN-^{15}N; (b) ^{1}HN-H$^{\alpha}$; and (c) ^{1}HN-^{13}C$^{\alpha}$ planes intersecting the point [HN=8.46, ^{15}N=117.9, ^{1}H$^{\alpha}$=4.97, ^{13}C$^{\alpha}$=57.8] ppm in the 4D spectrum. The original 4D time-domain data (256×32×16×16 points) were first digitally filtered, zero-filled, Fourier transformed, and phased along the amide proton (acquisition, t_4) dimension. The dispersive portion of the amide proton dimension was discarded, yielding a matrix of 256×32×16×16 (Ω_4,t_1,t_2,t_3) data points. Each of the 256 three-dimensional t_1,t_2,t_3 slices was processed as follows. The ^{15}N (t_1) and H$^{\alpha}$ (t_2) dimensions were first transformed into the frequency domain, without apodization or zero-filling, yielding 16×8 complex points in the Ω_1 and Ω_2 dimensions, respectively. The ^{13}C (t_3) dimension was extended from 8 to 15 complex points by mirror-image linear prediction, as described in the Methods section. The t_3 dimension was then processed with cosine-squared bell apodization, zero-filled to 64 complex points, Fourier transformed, and phased. After phasing, the dispersive portion of the spectrum was discarded. The ^{1}HN (t_2) dimension was then inverse Fourier transformed and processed analogously to the t_3 dimension. Finally, the ^{15}N (t_1) dimension was inverse Fourier transformed and extrapolated from 16 to 24 complex points by linear prediction. Subsequent to linear prediction, the t_1 dimension was processed analogously to the other indirectly detected dimensions. The resolution of the final 4D (Ω_4,Ω_1,Ω_2,Ω_3) matrix was 256×64×64×64 data points, respectively.

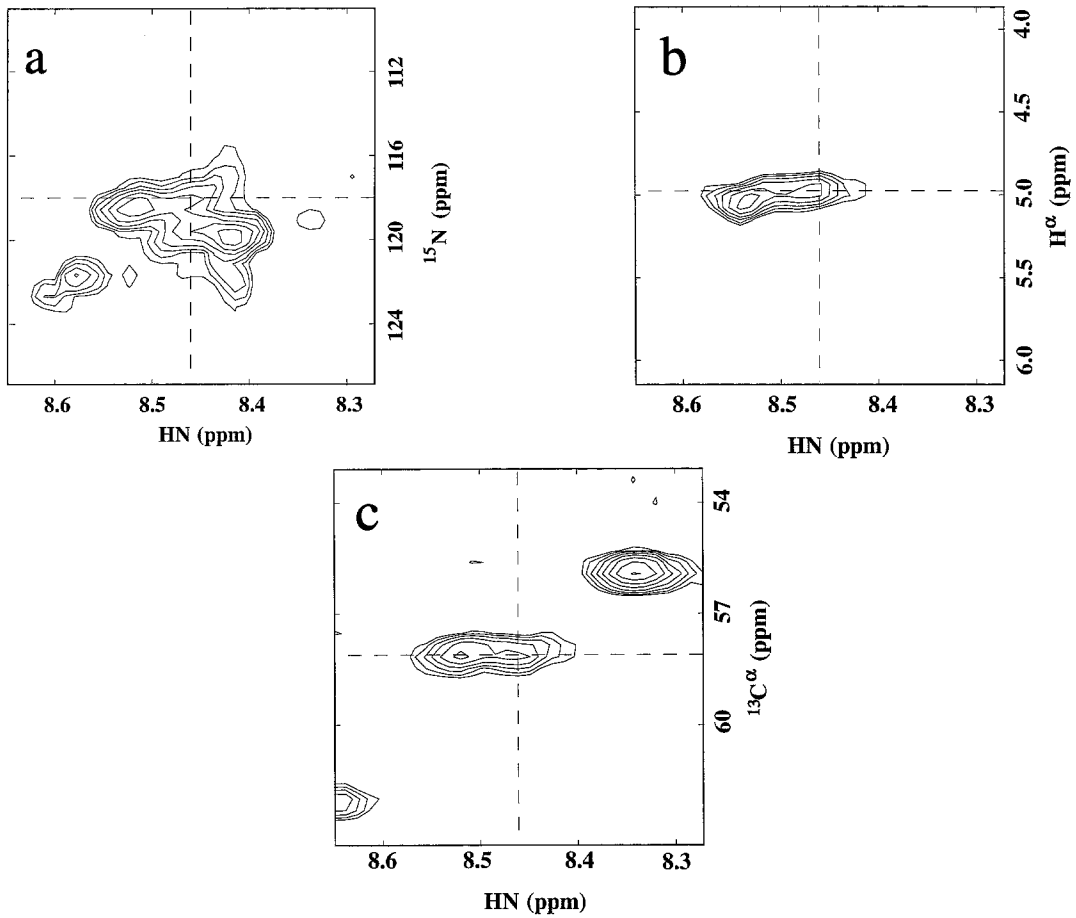

Fig. 4. Contour plots of slices through a 4D HN(CO)CAHA spectrum, enhanced by ML analysis. The 4D time-domain data used were the same as described in Fig. 3. The processing of the amide proton (acquisition) dimension was performed as described in Fig. 3. Each of the 256 three-dimensional t_1,t_2,t_3 (32×16×16) slices was processed as follows. The number, frequency, amplitude and decay rates of each of the signals in the 3D $^{15}NH^{\alpha}$-^{13}C slices were estimated by the maximum likelihood (ML) algorithm described in the text. The parameter estimates obtained from the ML analysis were used to replace the experimental data with a synthetic data set of resolution 64×64×64 along each of the indirectly detected dimensions. The dimensions were then each apodized using a 32 complex points cosine-squared bell apodization function, zero-filled to 64 complex points, Fourier transformed and phased. After phase correction, the dispersive portions were discarded. The resolution of the final 4D ($\Omega_4,\Omega_1,\Omega_2,\Omega_3$) matrix was 256×64×64×64 data points, respectively.

domain data by ML analysis, using the ChiFit algorithm described in the previous section.

From prior information about the phase corrections and the average decay rates along each dimension, ChiFit obtained the ML estimates for the number of sinusoids and their respective frequencies, amplitudes and decay rates along each dimension. Because the t_2 dimension was recorded with a constant-time acquisition period, the model describing the systematic portion of the data did not contain a decay rate for this dimension. The MDL statistic described in the previous section was used as the termination criterion for the algorithm. The parameter estimates obtained from the ML analysis were used to replace the experimental data with a synthetic data set of resolution 256×128 along the ^{13}C and ^{15}N dimensions, respectively. Each dimension was then processed by conventional Fourier processing methods, as described in the legend of Fig. 1.

A set of footprints labeled A–L are superimposed on the contour plots of both Figs. 1 and 2. The footprints represent the widths and center positions of the most intense peaks present in the best resolved spectra (Figs. 1a and 2a). The frequency positions of these signals are assumed to be the 'true' values in the truncated spectra. The existence of a maximum within a given footprint in a spectrum derived from a truncated data set was the criterion for considering that signal to be 'resolved' in that spectrum.

Visual inspection of the contour plots in Figs. 1 and 2 indicates that the ML-enhanced data sets have a greater number of resolved signals than the corresponding data sets that were processed by linear prediction. For example, although the frequency positions of signals D and I in Fig. 2b (64×64, t_1,t_2) are somewhat distorted from their 'true' positions, the two signals are completely unresolved in Fig. 1b. The cluster of five peaks (G-I-J-K-H) is very poorly resolved in Fig. 1d, whereas the same cluster is much better resolved in Fig. 2d.

The algorithm described in the previous section is general for any number of D dimensions. As an example of higher dimensional processing, ChiFit was used to perform a three-dimensional analysis of the $^{15}NH^{\alpha}$-^{13}C $[t_1,t_2,t_3]$ dimensions derived from the 4D HN(CO)CAHA data set from the same protein. The data were collected with a time-domain resolution of 256×32×16×16 data points along the respective $[t_4,t_1,t_2,t_3]$ dimensions. The H^{α} and ^{13}C dimensions had shared constant-time acquisition periods (Kay et al., 1992).

Figures 3 and 4 compare contour plots of the (a) HN-^{15}N; (b) HN-H^{α}; and (c) HN-^{13}C 2D absorption spectra derived from data processed by the two methods. The data in Fig. 3 were obtained by linear prediction extrapolation of each of the indirectly detected dimensions separately. The time required to process the entire 4D data set using linear prediction was 32 h on an SGI R4000 workstation. The series of three absorption spectra represent 2D cross sections of an unresolved HN(CO)CAHA correlation at [HN=8.46, ^{15}N=117.9, $^{1}H^{\alpha}$=4.97, ^{13}C=57.8] ppm. The data in Fig. 4 were obtained by using ML analysis to replace each of the 32×16×16 $^{15}NH^{\alpha}$-^{13}C $[t_1,t_2,t_3]$ regions with ML-derived synthetic data with a resolution of 64×64×64 data points. The time required to process the 256 $[t_1,t_2,t_3]$ regions was 18 h on the SGI R4000 workstation. An additional time period of 12 h was required to apply Fourier methods to the transformation of the time-domain data. The ML-enhanced spectrum (Fig. 4) contains evidence for a signal that is not at all resolved in the corresponding spectrum that was enhanced by linear prediction (Fig. 3).

Discussion and Conclusions

By using a Gaussian noise prior probability and Bayes' theorem, Bretthorst (1990c) has shown that for the case of one-dimensional NMR data consisting of stationary (no decay), well-separated sinusoids, the power spectrum (squared magnitude of the Fourier transform of the data) evaluated at a given frequency ω is proportional to the logarithm of the probability that the data contain a signal of frequency ω. This probability is independent of the amplitude or phase of the signal. This result has been extended to the general case of D dimensions (Chylla and Markley, 1993). Bayesian probability theory also states that the weighted power spectrum (squared magnitude of the Fourier transform of exponentially weighted data) is an optimal estimator of frequency when the line width of the well-separated decaying sinusoids in the data match the filter width of the exponential weighting function. Although they did not derive a relationship between the discrete Fourier transform and the log-likelihood of the noise, Miller and co-workers used the absorptive portion of the Fourier transform in one (Miller and Greene, 1989) and two dimensions (Miller et al., 1993) to obtain estimates of NMR parameters as an initial step of ML estimation. We have derived here, for the general case of D dimensions, a relationship between the weighted power spectrum and the log-likelihood of the frequency parameters, independent of the phase and amplitude parameters when prior information about the decay rates is available. We have also shown that the weighted absorption spectrum is proportional to the log-likelihood of the frequency parameters, independent of the amplitudes when the phases and decay rates of the signals are available from prior knowledge. In doing so, we have established a strong correspondence between Fourier analysis, Bayesian probability theory, and ML estimation as applied to the task of frequency estimation. When the amplitudes of the systematic portions of the data are constrained to their ML values, then Bayesian probability theory and ML estimation yield identical estimates for the most likely frequencies contained within the data.

As presented above, we have implemented the ML method in a computer program, ChiFit. The procedure assumes prior knowledge of the phases and an estimate of the mean decay rate along each dimension. The algorithm predicts the number, frequency, amplitude and decay rates of the signals in each of the indirectly detected dimensions after prior Fourier transformation of the acquisition dimension. The truncated experimental data along the indirectly detected dimensions are then replaced with extrapolated synthetic data, derived from the ML estimates of the corresponding experimental data. Subsequent Fourier transformation of the indirectly detected dimensions thus yields spectra with enhanced frequency resolution. The results presented here show that the approach can be implemented in a practical way on current computer workstations. Comparison of experimental triple-resonance 3D and 4D data demonstrates that the ML approach can resolve frequencies in truncated data that conventional processing techniques cannot.

In this work we have used the ML method in the same fashion that linear prediction is conventionally used, i.e., to extrapolate truncated time-domain data prior to digital filtering, zero-filling, and subsequent Fourier transformation. The final estimation of the number, frequency and amplitudes of each of the signals in the spectra, enhanced by both linear prediction and ML analysis, was performed by automatic 'peak picking' of all the peaks in the absorption spectrum above a certain threshold (results not given here). ML analysis was thus used to improve upon parameter estimation but was not, in the end, used to perform signal recognition or model selection. This limited use of ML estimation is by no means its most efficient or optimal implementation. Because ML estimation, in conjunction with Fourier transformation and information theory, is capable of performing the tasks of signal recognition, parameter estimation and model selection, the optimal use of ML estimation would be in carrying

out the task of predicting the number, frequency, amplitude, and decay rates of each signal in the entire multidimensional data matrix for all dimensions. Adapting the ML algorithm presented here to this more general task in a practical period of time is an active area of current research. The direct extraction of relevant NMR parameters from large, multidimensional data sets of proteins, in conjunction with robust automated assignment software (Olson and Markley, 1994), could provide a completely automated pathway from the acquisition of NMR data to full assignments of resonances and to structural analysis.

Acknowledgements

We thank Mark Girvin, Robert Fillingame, Frits Abildgaard and Ed Mooberry for allowing us to use their unpublished experimental 3D HNCO and 4D HN(CO)-CAHA data sets as illustrations in this study. This work was carried out at the National Magnetic Resonance Facility at Madison under support from NIH Grants LM04958 and RR02301. Equipment in the facility was purchased with funds from the University of Wisconsin, the NSF Biological Instrumentation Program (Grant DMB-8415048), the NIH Biomedical Research Technology Program (Grant RR02301), the NIH Shared Instrumentation Program (Grant RR02781) and the U.S. Department of Agriculture.

References

Barkhuijsen, J., De Beer, W.M., Bovee, W.M.M.J. and Van Ormondt, D. (1985) *J. Magn. Reson.*, **61**, 465–481.

Bloch, F., Hansen, W.W. and Packard, M.E. (1946) *Phys. Rev.*, **70**, 460–474.

Bretthorst, G.L. (1990a) *J. Magn. Reson.*, **88**, 533–551.

Bretthorst, G.L. (1990b) *J. Magn. Reson.*, **88**, 552–570.

Bretthorst, G.L. (1990c) *J. Magn. Reson.*, **88**, 571–595.

Chylla, R.A. (1994) unpublished documentation. Available through the National Magnetic Resonance Facility (NMRFAM) software section of the Internet GOPHER utility at gopher://gopher.nmrfam.wisc.edu/11/Software/Chifit.

Chylla, R.A. and Markley, J.L. (1993) *J. Biomol. NMR*, **3**, 515–533.

Gesmar, H. and Led, J.J. (1989) *J. Magn. Reson.*, **83**, 53–64.

Girvin, M.E. and Fillingame, R.H. (1993) *Biochemistry*, **32**, 12167–12177.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496–514.

Kay, L.E., Wittekind, M., McCoy, M.A., Friedrichs, M.S. and Mueller, L. (1992) *J. Magn. Reson.*, **98**, 443–450.

Kumaresan, R. and Tufts, D.W. (1982) *IEEE Trans. Acoust. Speech Signal Processing*, **30**, 833–840.

Led, J.J. and Gesmar, H. (1991) *J. Biomol. NMR*, **1**, 237–246.

Marquardt, D.W. (1963) *J. Soc. Ind. Appl. Math.*, **11**, 431–441.

Miller, M.I. and Greene, A.S. (1989) *J. Magn. Reson.*, **83**, 525–548.

Miller, M.I., Chen, S.C., Kuefler, D.A. and D'Avignon, D.A. (1993) *J. Magn. Reson. Ser. A*, **104**, 247–257.

Mooberry, E.S., Abildgaard, F. and Markley, J.L. (1994) *Methods Enzymol.*, **2**, 247–256.

Olson Jr., J.B. and Markley, J.L. (1994) *J. Biomol. NMR*, **4**, 385–410.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988) *Numerical Recipes in C*, Cambridge University Press, New York, NY, pp. 540–547.

Reddy, V.U. and Birader, L.S. (1993) *IEEE Trans. Signal Processing*, **41**, 2872–2881.

Zhu, G. and Bax, A. (1990) *J. Magn. Reson.*, **90**, 405–410.