Application of the Linear Prediction Method to NMR Spectroscopy

JENS J. LED* and HENRIK GESMAR*

Department of Chemistry, University of Copenhagen, The H. C. Ørsted Institute, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

Received May 6, 1991 (Revised Manuscript Received July 10, 1991)

Contents

I. Introduction

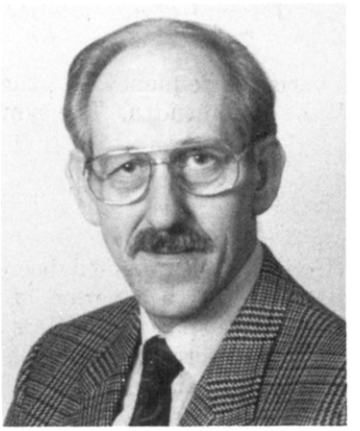
In the last few years linear prediction (LP) has proved valuable as a computational method for the accurate estimation of information contained in complicated NMR spectra. This holds in particular because the LP method does not suffer from a series of drawbacks that characterizes the discrete Fourier transform (DFT) method, routinely used in NMR spectroscopy since the introduction of the pulse NMR technique.[1,2] The LP method can be used either for spectral improvements in combination with the DFT method; i.e. it can be used qualitatively to remidy spectral artifacts that, otherwise, may arise from the discrete Fourier transformation of an experimental time-domain signal (the free induction decay, FID); or it can be used for spectral analysis and, thus, provide quantitative estimates of the spectral information contained in the FID; i.e. estimates of the frequencies, the line widths, the intensities, and the phases of the individual signals of the corresponding spectrum. In this case it is an alternative to the DFT method.

Starting with a brief exposition of the DFT method and its virtues and drawbacks, this review focuses on the scope and limitations of the LP method, and the qualitative and quantitative applications of the method to NMR spectroscopy. Other computational methods, such as the maximum entropy method[3] (MEM) and the maximum likelihood method[4] (MLM) are also being used for improved analyses of experimental NMR data, but will not be covered here. For a more exhaustive theoretical description of the LP and MEM methods the reader is referred to the excellent review by D. S. Stephenson.[5]

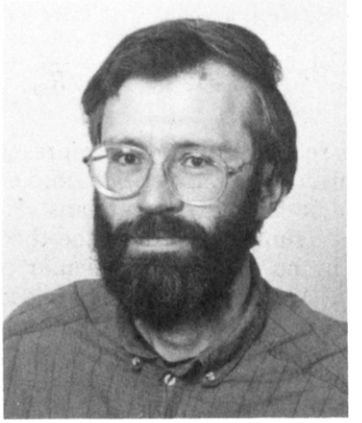
II. The Discrete Fourier Transformation of the Free Induction Decay

On the assumption that the sampled time-domain NMR signal (FID) is a sum of exponentially damped sinusoids the kth data point of the FID is given by

$$F_k = F(\Delta t k) = \sum_{j=1}^{p} I_j \exp((i2\pi\nu_j - R_{2j})(k\Delta t + T_{in}) + i\varphi_j) \quad (1)$$



Jens J. Led received a M.S. in chemical engineering in 1963 and a Ph.D. in physical chemistry in 1968 from The Technical University of Denmark. The Ph.D. study was carried out with Børge Bak at The University of Copenhagen, Denmark. In 1968 he joined the Department of Chemistry, The University of Copenhagen as a Lecturer in physical chemistry. From 1971 to 1973 he was a research associate with David M. Grant at the Department of Chemistry, The University of Utah. He received the Dr.Scient. degree from the University of Copenhagen in 1988 on the basis of a series of NMR spectroscopic studies of paramagnetic metal complexes of biological importance. He has currently a position as Senior Lecturer in physical chemistry at the Department of Chemistry, The University of Copenhagen and is head of the Protein NMR Center at The University. His main research interest includes NMR spectroscopic studies of proteins and the development of NMR spectroscopic and computational methods.



Henrik Gesmar was born in Copenhagen, Denmark, in 1948. He received his M.S. in chemistry from the University of Copenhagen in 1983. He is also a B.A. in Spanish. In 1988 he received his Ph.D. in chemical physics based on the thesis: "Some Quantitative Methods in Liquid State NMR Spectroscopy". This work was financially supported by a grant from the Carlsberg Foundation. At present he is a lecturer in physical chemistry at the University of Copenhagen and associated to the Protein NMR Center at the University. His primary research interest is the application of NMR spectroscopy to the study of proteins, focusing on spectral analysis and computational methods.

0009-2665/91/0791-1413$09.50/0

where Δt is the sampling interval, p the number of sinusoids, I_j the amplitude of the jth resonance, ν_j the chemical shift, R_{2j} the transverse relaxation rate, and ϕ_j the phase. Further, T_{in} is the initial delay between the rf pulse and the first data point, $F(0)$. As shown previously[6,7] the discrete Fourier transformation of eq 1 results in a discrete NMR spectrum:

$$\hat{S}(\nu) = \sum_{j=1}^{p} A_j \frac{1 - \exp((i2\pi(\nu_j - \nu) - R_{2j})T_{aq})}{1 - \exp((i2\pi(\nu_j - \nu) - R_{2j})/SW)} \quad (2)$$

Here ν is the variable frequency, T_{aq} the acquisition time, and SW the sweep width. The symbol A_j is defined as

$$A_j = I_j \exp((i2\pi\nu_j - R_{2j})T_{in} + i\varphi_j) \quad (3)$$

As it appears from eq 2 the obtained discrete spectrum is a periodic function with a period of $SW = \Delta t^{-1}$. Therefore, the discrete spectrum differs significantly from the theoretical spectrum, given as a sum of nonperiodic complex Lorentzians

$$s(\nu) = \sum_{j=1}^{p} I_j \exp(i\varphi_j)/R_{2j} \frac{1}{1 - i2\pi(\nu_j - \nu)/R_{2j}} \quad (4)$$

However, the discrete spectrum in eq 2 approximates the theoretical spectrum if (1) $T_{aq} \gg 1/R_{2j}$ and (2) SW $\gg R_{2j}$, $\nu_j - \nu$; i.e. (1) if all sinusoids decay completely during the acquisition period, and (2) if the period of the frequency function in eq 2 is much larger than the signal line widths as well as the frequency range of the spectrum. Whereas the former of the two last mentioned conditions are often fulfilled the latter is rarely so. Nonetheless, for all three conditions fulfilled eq 2 reduces to

$$\hat{S}(\nu) = SW \sum_{j=1}^{p} A_j/R_{2j} \frac{1}{1 - i2\pi(\nu_j - \nu)/R_{2j}} + \sum_{j=1}^{p} A_j/2 \quad (5)$$

as shown in ref 7. Equation 5 represents the ideal discrete Fourier transform spectrum; i.e. a frequency sampled sum of complex Lorentzians. However, even this ideal spectrum differs from the theoretical spectrum given by eq 4 at two particular points: (1) a constant, $\sum_{j=1}^{p} A_j/2$, has been added to each data point, and (2) each one of the p Lorentzians have been multiplied by the factor $\exp((i2\pi\nu_j - R_{2j})T_{in})$.

Here the first deviation results in a pseudobaseline in the DFT spectrum. According to eq 5 the level of this pseudobaseline is equal to half of the sum $\sum_{j=1}^{p} A_j$ which in turn, is given by the first point in the FID. Therefore, the pseudobaseline can be removed by multiplying the first point in the FID by 1/2 before performing the Fourier transformation. In particular in the case of two dimensional spectra this multiplication is important since here the pseudobaselines lead to F_1 and F_2 ridges.[8]

The second deviation, i.e. the multiplication of each data point in the spectrum by the factor $\exp((i2\pi\nu_j - R_{2j})T_{in}$, results in a frequency and relaxation dependent phase distortion of the DFT spectrum, as it appears from eq 3. This phase distortion does not vary linearly with the frequency, ν, in a continuous way. Instead it depends linearly only on the center frequencies of the individual Lorentzians. This implies that the same phase correction, i.e. the phase correction corresponding to the centers of the Lorentzians, must be used across a given signal irrespective of its line width. Thus, in effect the phase correction is nonlinear. Even so, a linear phase correction suffice in most cases with modest phase distortions and narrow and well-separated signals. If, however, extreme phase distortions occur or the signals are broad and poorly separated, or if small signals are situated on the tail of a larger signal, the spectrum can be correctly phased only at the center of each signal, and baseline undulation will occur.[9] Thus, the true spectrum cannot be produced even under ideal sampling conditions, i.e. when the time-domain signal is allowed to relax completely during acquisition and the sweep width is significantly larger than the width of the spectrum. Obviously, this effect increases with increasing values of the initial delay, T_{in}. Therefore, the value of T_{in} is normally kept as small as possible even though nonlinear effects of the amplifier, distortions due to electronic filtering, and overflow in the AD converter may reduce the quality of the first few data points of the FID, and thereby give rise to new baseline distortions. However, as discussed below these distortions can be remedied by linear prediction.

If the ideal sampling conditions are not fulfilled additional distortions occur. Thus in case of a trunction of the FID ($T_{aq} \simeq 1/R_{2j}$) the jth Lorentzian is overlaid by a new periodic expression, $\exp((i2\pi(\nu_j - \nu) - R_{2j})T_{aq})$, with the period of one data point (T_{aq}^{-1}). This, in turn, results in an apparently misphased signal.[6] Furthermore, the relative signal intensity is no longer given by the area between the curve and the baseline, because the pseudobaseline level is also altered by the factors $\exp((i2\pi(\nu_j - \nu) - R_{2j})T_{aq})$; i.e. the area between the baseline and the resonance curve depends on the frequency and the relaxation rate. As seen from eq 5, this deviation can be reduced by increasing the sweep width, SW, to a size significantly larger than the actual spectral width.

If the FID is zero filled before the discrete Fourier transformation in order to increase the digital resolution, the exponential term in the numerator of eq 2 causes even more pronounced distortions in the resulting spectrum. This can be seen by letting m assume fractional values in $\nu = mT_{aq}^{-1}$. "Wiggles" with a period of T_{aq}^{-1} will then appear in the spectrum in connection with the signal whose components in the FID have been truncated. It should be emphasized that in this case a correction of the baseline by multiplying the first data point of the FID by a factor of 1/2 will not work since the pseudobaseline is here modulated by the term $\exp(-i2\pi\nu T_{aq})$.

Finally, it should be emphasized that if the sweep width is not significantly larger than the actual frequency range covered by the spectrum (SW $\simeq \nu_j - \nu$) the periodic behavior of the denominator of eq 2 will give rise to the well-known effects of aliasing or zoning.[10]

As demonstrated above an experimental DFT spectrum differs from the theoretical spectrum to a large extent and may even differ significantly from the optimal DFT spectrum. As will be demonstrated in the following sections several of the drawbacks of the DFT

spectrum can be alleviated or even remedied by linear prediction calculations.

III. Spectral Improvements

It is evident from the previous section that a series of artifacts, often observed in DFT spectra, are due to defective or incomplete time-domain signals. This holds, in particular, for baseline distortions due to corruption of the first few data points of the FID, large first-order phase corrections caused by lack of information about the initial part of the FID, and truncation artifacts that result from incompletely decayed and zero-filled FID's. Consequently, these artifacts can be remedied or alleviated if the FID can be repaired or completed. As originally demonstrated by Tirendi and Martin[11] and Marion and Bax,[12] such time-domain signals can be generated from the experimental FID's by linear prediction calculations. In this section the principle of the LP calculations that allow repair and extrapolation of the FID's will be briefly reviewed together with a series of examples that illustrate the application of the LP method for these purposes.

A FID signal, F_k, consisting of a sum of exponentially damped sinusoids (eq 1), sampled at regular time intervals, has the following property[13]

$$F_k = \sum_{m=1}^{p} f_m F_{k-m} \tag{6}$$

disregarding the effect of the noise. Here f_m is the mth forward prediction coefficient. Since f_m is independent of k it can be determined from the experimental data points, and subsequently used to calculate the extension of the FID's. To determine the prediction coefficients, f_m, from a set of linear equations of the type given by eq 6, various techniques are being used[11,14-19] depending on the size and the nature of the problem. Some of these techniques were recently discussed by Gesmar et al.[7] An interesting alternative to direct extrapolation of the FID, the LPZ method, has been suggested by Tang and Norris.[20]

Further, the prediction coefficients are related to the frequencies, ν_j, and the relaxation rates, R_{2j}, through the characteristic polynomial

$$z^p - \sum_{m=1}^{p} f_m z^{p-m} = P(z) \tag{7}$$

since the roots, C_j, of this polynomial are given by the equation[13]

$$C_j = \exp((i2\pi\nu_j - R_{2j})\Delta t) \tag{8}$$

In general the precise number of signals contained in the FID is unknown a priori. Therefore a number of prediction coefficients must be chosen, that is sufficiently large ($>p$) to assure the determination of all the involved complex exponentials. This results in an excess of roots of the characteristic polynomial in eq 7. However, as shown by Kumaresan,[21] the roots are divided into two classes: one class containing the p signal roots that conform to eq 8, and another class containing the extraneous roots. Moreover, in the case of backward prediction, i.e.

$$F_k = \sum_{m=1}^{p} b_m F_{k+m} \tag{9}$$

the extraneous roots fall inside the unit circle, whereas the signal roots, C_j, are situated outside the unit circle, thus making possible a separation of the two classes. In the case of forward prediction (eq 6) both classes fall inside the unit circle. When the coefficients are to be used for an extrapolation of the FID this has, in principle, no consequences since both classes of coefficients must be used for the extrapolation. In practice, however, the presence of noise may occasionally result in roots that fall outside the unit circle corresponding to noise components with negative relaxation rates.

The first example of spectral improvement through a combination of the LP and DFT methods concerns the effect of a corruption of the first few data points of the FID. As shown in Figure 1 severe damage of these points results in a strongly curved baseline of the spectrum. However, a reconstruction of the first data points of the FID from the following experimental data points by backward LP using eq 9 results in a DFT spectrum with a completely smooth baseline.

Another powerful application of backward LP is for phase correction of soft-pulse spectra, such as soft-COSY and soft-NOESY spectra[22] or soft-HMQC spectra.[23] Due to the length of the selective pulses used in soft-pulse experiments, a considerable frequency-dependent dephasing of the magnetization occurs during the pulses. As discussed earlier, a linear phase correction of the size needed in this case would lead to severe distortions of the spectrum. Consequently, the resulting spectra are characterized by large first-order phase distortions, unless special measures are taken to refocus the magnetization, such as refocusing 180° pulse immediately after the soft-pulse.[22] Recently, Geen and Freeman[24] have developed and elegant "top-hat" shaped composite pulse with "phase purity"; i.e. a band selective pulse without phase distortion. This pulse, however, acts only if the magnetization is initially along the z axis.[24] It can, therefore, be used only as an excitation pulse and not as a general band-selective pulse.

An experimentally simple way to correct the phase distortion caused by the dephasing of the magnetization during semiselective soft pulses, is to extrapolate the phase distorted experimental FID backward to $t = 0$, that is to the time at which all the sinusoids of the FID are in phase. For this end backward LP is the ideal computational method as originally suggested by Marion and Bax,[12] and newly demonstrated by Led et al.[25] using the soft-HMBC experiment as an example. The pure-absorption-mode version of this experiment, developed by Davis,[23] includes two soft-pulses

$$\begin{array}{lccc} {}^1\mathrm{H}: & (\pi/2)_x\,(\mathrm{soft}) - \Delta - | - t_1/2 - \pi_y\,(\mathrm{soft}) - t_1/2 - | - \Delta - \mathrm{acq}\,(t_2) & \\ & | & | & \\ \mathrm{X}: & (\pi/2)_\varphi & (\pi/2)_x & \end{array} \tag{10}$$

Here, first-order phase distortions in the F_2 dimension can be avoided by using the $\pi/2$ "top-hat" excitation pulse[24] with "phase purity" for the proton excitation. However, the semiselective π pulse applied in the middle of the evolution period inevitably introduces a lower limit to the time, t_1, given by the length of the π pulse and thus gives rise to a considerable phase distortion in the F_1 dimension. In ref 25 Led et al. used the experiment to correlate the carbonyl carbons ($^{13}C'$) and the α protons of the individual leucine residues in human growth hormone (hGH), using a sample of hGH in which the carbonyl carbons of all 26 leucine residues

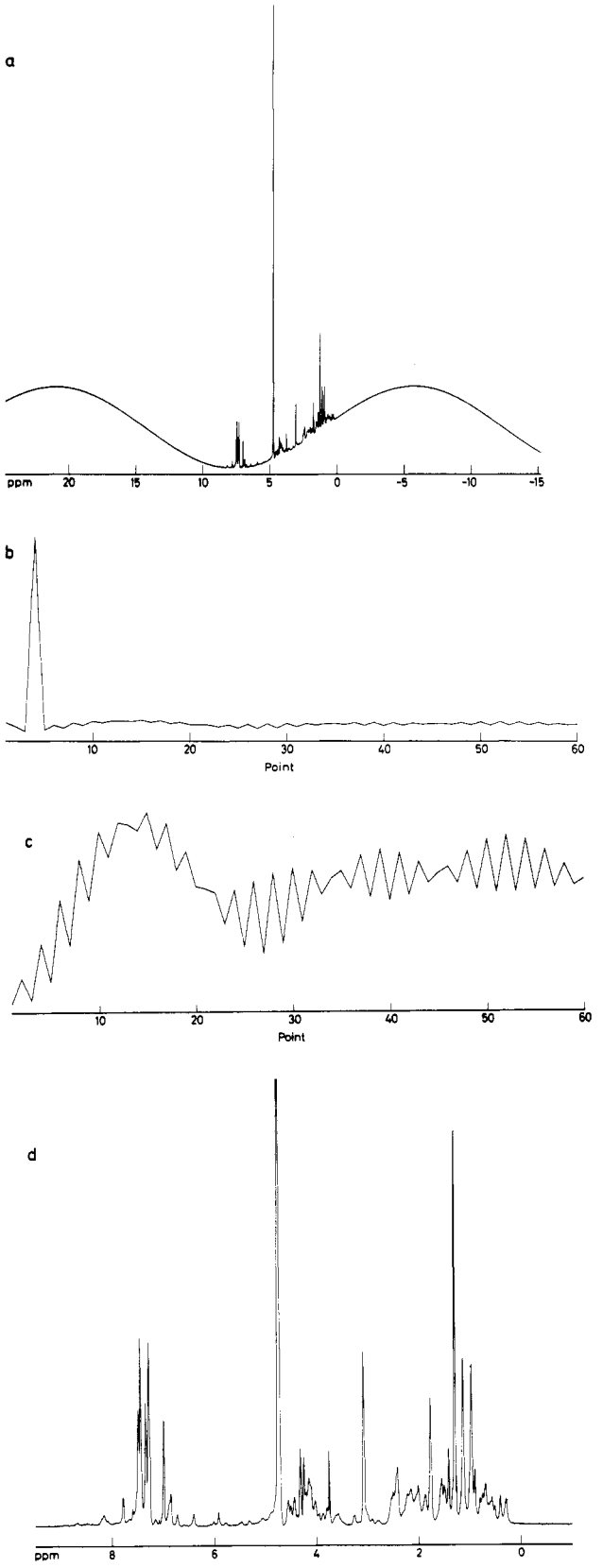


Figure 1. The first slice along the F_2 dimension in a two-dimensional HOHAHA spectrum of insulin in D_2O: (a) the DFT spectrum and (b) the first 60 data points of the FID. Notice the fourth data point in the FID is severly damaged. In c the first 6 data points of the FID were deleted and recalculated from 18 LP coefficients by using eq 9. The LP coefficients were determined from 36 equations of the type of eq 9, k ranging from 7 to 42. Notice that the vertical scale is changed. In d the DFT spectrum of the repaired FID is shown. Notice that both scales are changed.

were 95% ^{13}C enriched. The length of the applied Gaussian π semiselective pulse in the middle of the t_1 period was 2 ms. For the applied Δt period of 40 μs this corresponds to a delay of the t_1 acquisition of 50 data points, or a linear phase distortion of 9000 degrees across the sweep width. Although, in principle, there is no limit to the size of the first-order phase correction one can apply, there is, in practice, a limit due to the nonlinear nature of the phase distortion, discussed above in section II. This is immediately apparent from Figure 2a that shows a contour plot of the $C^{\alpha}H-^{13}C'$ region of the soft-pulse HMBC spectrum of hGH and from Figure 3a that shows a slice taken along the ^{13}C axis at the position indicated by the arrow in Figure 2a. Thus, although the spectrum basically is in pure absorption mode in both dimensions, negative signals occur despite an optimized phase correction. As revealed in Figure 3a the negative signals are due to baseline undulation around the signals as discussed in the previous section. When the experimental two-dimensional FID is extended 50 data points backward in the t_1 domain i.e. to the point where all sinusoids are in phase, the artifacts disappear and the discrete Fourier transform results in a spectrum almost free of negative signals and baseline undulation as illustrated in Figures 2b and 3b. Further spectral improvements can be obtained by forward extrapolation of the FID as shown in Figures 2c and 3c and discussed in detail in the following.

A third application of LP extrapolation, that has proved very useful, is forward extrapolation of truncated FID's with the purpose of reducing or eliminating the trunction errors, and increasing the resolution and the sensitivity of the DFT spectrum. This was originally demonstrated by Tirendi and Martin[11] for the case of a NOESY spectrum of a small protein. Forward extrapolation is particularly valuable in cases where the resolution and the sensitivity are limited by the time necessary to obtain the data file, such as three- or higher dimensional spectra[27] or natural abundance two-dimensional heteronuclear correlation spectra.[28] In both cases the time-domain data for the slowest incrementing time dimension is normally truncated. For the three- or higher dimensional spectra the truncation is due to accumulation time constraints rather than to the lack of signal. In contrast, lack of signal is mainly the problem in the natural abundance heteronuclear correlation spectra. Here the sensitivity problem can be alleviated by confining the data acquisition in the t_1 domain to only the first and most intense part of the FID, if the associated truncation errors and reduced resolution can be eliminated by forward LP extrapolation of the FID. In both cases the reliability of the extrapolation depends on the number and the quality of the experimental data used for the extrapolation. This can be realized by considering a noiseless FID containing N sinusoids. Here only $4N$ data points would be necessary to determine the $4N$ parameters that are involved (i.e. the frequencies, the line widths, the intensities, and the phases of the N signals) and, thus, reproduce the entire FID. When noise is present, as in experimental data, the number of data points and LP coefficients used for the extrapolation must be increased considerably. Thus, as demonstrated by Gesmar and Led[19] for a one-dimensional spectrum, 128 LP coefficients calculated from 512 data points were nec-

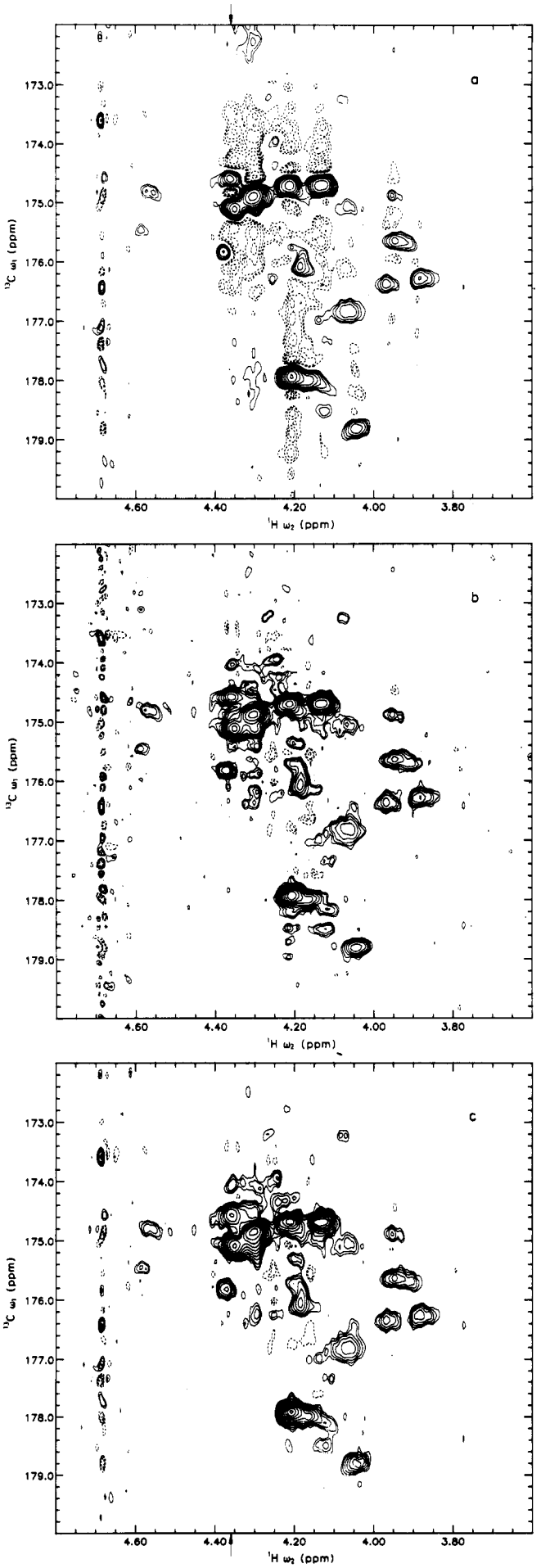


Figure 2. Contour plots of a soft-pulse absorption mode HMBC spectrum of human growth hormone (hGH) with 95% ^{13}C labeling of all 26 Leu carbonyl carbons. The spectrum correlates the carbonyl carbons with the α protons. Full contour lines indicate positive signals; broken contour lines indicate negative signals. The spectrum was obtained from a 0.8 mM solution of hGH in D_2O, 50 mM phosphate buffer, pD 2.32, and 305 K, with use of a Bruker AM 500 NMR spectrometer equipped with an inverse probe and a selective excitation unit. The pulse scheme in eq 10 was used for the experiment. A detailed description is given in ref 25. The initial (t_1,t_2) data set consisted of 512 × 2048 real data points acquired over a period of 25 h with acquisition times of 20.5 ms (t_1) and 102.4 ms (t_2), 128 acquisitions per t_1 increment and a delay of 1.84 s between acquisitions. During this delay the residual HOD signal was saturated with a weak DANTE pulse.[26] The sweep width was 12 500 Hz in F_1 and 10 000 Hz in F_2. After Fourier transformation the F_2 domain was truncated to 1250 Hz resulting in an absorption mode spectrum represented by 1024 data points. In a, a first-order phase correction of 9000° has been applied in order to correct the phase in the F_1 dimension. Before Fourier transformation the t_1 domain was zero filled to 2048 data points. In b, the t_1 domain has been backward extrapolated 50 points by using linear prediction and zero filled to a total of 2048 points. No first-order phase correction was necessary in order to correct the phase in the F_1 dimension. In c, the FID has, furthermore, been extrapolated to 2048 points (reprinted from ref 25; copyright 1991 Academic Press, Inc.).

essary for an optimum reproduction of four signals in a spectrum with a signal-to-noise ratio of approximately 1:10.

As mentioned above, the noise may also give rise to roots outside the unit circle. Since these roots correspond to noise components with negative relaxations, an extrapolation that included the corresponding LP coefficients would result in an extrapolated FID that increased instead of decayed.[29] In order to remove these roots and assure the decay of the extrapolated FID, a more stable solution should be obtained by including more experimental information.[29] If this is not possible, the roots may simply be reflected to fall inside the unit circle.[12] Therefore, it is mandatory to apply eq 7 in order to check for roots outside the unit circle and, if necessary, to construct a modified set of roots from which a new set of LP coefficients can be calculated. Without this check artifactual peaks may creep into the analysis.

Forward LP extrapolation, as outlined above, was recently applied to a proton-detected natural abundance heteronuclear multiple quantum coherence (HMQC) spectrum that correlates the protonated carbon in an insulin analogue (B9(Asp) insulin) to the attached protons.[28] The combination of a short t_1 acquisition time and LP extrapolation of the FID in the t_1 dimension is particularly favorable in this case due to the short R_2 relaxation rate induced by the self-aggregation of insulin.

In Figure 4 are shown various modifications of an experimental FID along the t_1 dimension in the methyl region of the C–H correlation spectrum of insulin. The original t_1, t_2 data set was Fourier transformed in t_2, and the FID in Figure 4 is a sum of a group of proton signals modulated by the frequencies of the attached ^{13}C nuclei. The extrapolation shown in Figure 4c was based on 150 forward LP coefficients derived from the experimental data points by using 354 equations (eq 6). The spectral resolution obtained by different types of data treatment, and the enhancement that can be obtained by the LP extrapolation, appear from the comparison in Figure 5. Thus, while the Fourier transform of the 512 ex-

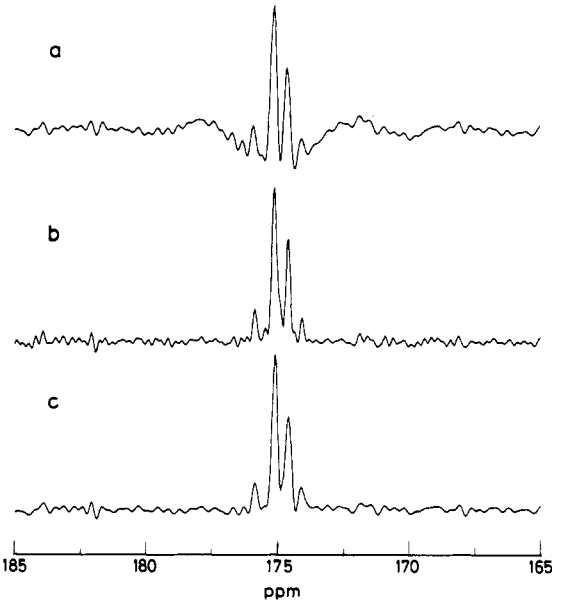


Figure 3. Traces along the ^{13}C axis in the soft-pulse HMBC spectrum in Figure 2. The traces in a, b, and c correspond to contour plots a, b, and c in Figure 2 and are taken at 4.36 ppm as indicated by the arrows in Figure 2. In c the FID has, furthermore, been forward extrapolated to 2048 points. (reprinted from ref 25; copyright 1991 Academic Press, Inc.).

perimental data points alone results in an unacceptable low digital resolution as shown in Figure 5a, zero filling (Figure 4a) gives rise to truncation errors in the form of a sinc function ("wiggles") superimposed on the signal as shown in Figure 5b. Although this artifact can be removed by digital filtering (Figure 4b) it can be done only by further sacrifice of resolution as seen in Figure 5c. Only if the FID is LP extrapolated to an almost complete decay (Figure 4c) is the information, inherent in the experimental FID, retained as shown in Figure 5d. The increased resolution obtained by the LP extrapolation is further illustrated in Figure 6 that shows the contour plot of the methyl region, corresponding to Figure 5 parts c and d.

The full potential of LP extrapolation for enhancing heteronuclear correlation spectra is displayed in Figure 7. This figure shows the trace in a heteronuclear single quantum correlation (HSQC) spectrum corresponding to the trace in the HMQC spectrum shown in Figure 5. Unlike the HMQC spectrum the HSQC spectrum has no splitting in the F_1 dimension due to passive J couplings to adjacent protons. Consequently, the line broadenings induced by these couplings in the HMQC spectrum are absent in the HSQC spectrum. A comparison of Figures 5d and 7b reveals that this is indeed the case. Also, a comparison of Figure 5c and 7a shows that the differences in line width in the two spectra are completely hidden by the extra line width imposed on the signals by the digital filtering, necessary in the case of truncations. Therefore, under experimental conditions where the FID must be truncated, such as in the case of natural abundance heteronuclear correlation spectra of proteins, the higher resolution of the HSQC spectra as compared to the resolution of the corresponding HMQC spectra can be observed only by applying LP extrapolation.

In conclusion it should be stressed that *backward* prediction is normally a "safe" modification, considering

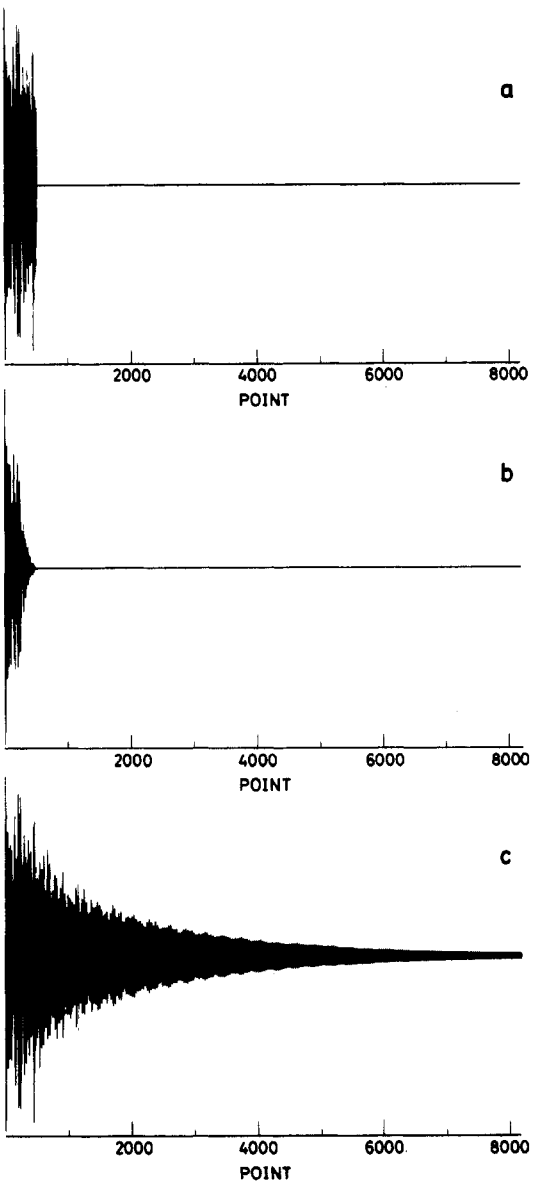


Figure 4. An experimental FID along t_1 (512 data points, t_1 acquisition time 10.2 ms), corresponding to the trace indicated by an arrow in the HMQC C–H correlated two-dimensional spectrum in Figure 6: (a) zero filled to 8192 data points; (b) zero filled to 8192 data points and digital filtered by using a $\cos^2$ window function; and (c) LP extrapolated to 8192 data points by using eq 7 (reprinted from ref 28, copyright 1991 ESCOM Science Publishers B.V.).

the relatively small number of extrapolated data points as compared with the number of LP coefficients that can be calculated, and the large surplus of experimental data points from which the LP coefficients can be derived. Thus, normally we apply a number of LP coefficients that is a factor of three larger than the number of backward predicted points. Further, the number of equations used to determine the LP coefficients should be at least twice the number of coefficients. When *forward* prediction is applied more care should be exercised. In practice an extrapolation that exceeds the number of experimental data points by several factors may often be desirable in order to assure a complete decay. In such cases the number of complex LP coefficients must at least be equal to the number of resonances in the FID, and preferably exceed it by several factors, the surplus necessary depending on the sig-

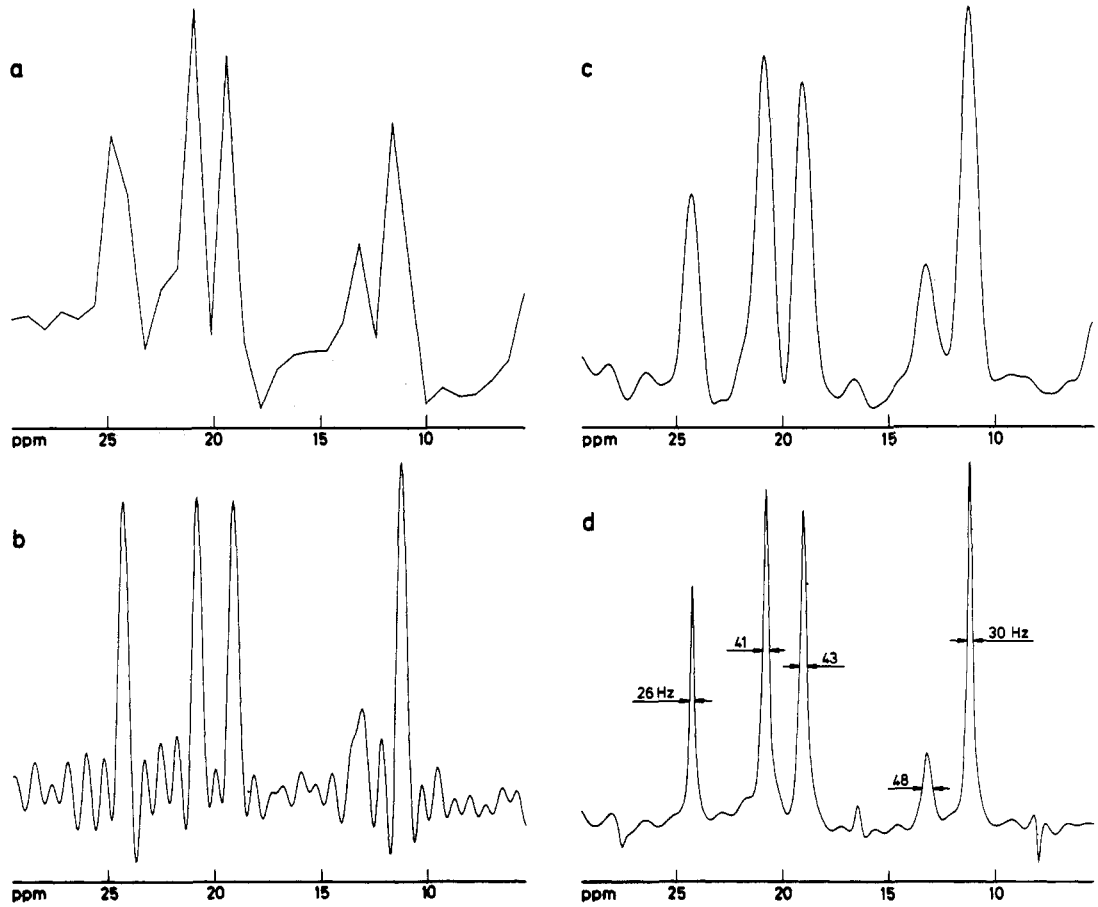


Figure 5. Comparison of four versions of the F_1 trace indicated by an arrow in the HMQC C-H correlated two-dimensional spectrum in Figure 6: (a) Fourier transform of the 512 experimental data points in Figure 4a, alone, and (b) the entire zero-filled FID; (c) Fourier transform of the digital filtered ($\cos^2$ filter) FID in Figure 4c; and (d) Fourier transform of the LP extrapolated FID in Figure 4c (c and d are reprinted from ref 28; copyright 1991 ESCOM Science Publishers B.V.).

nal/noise ratio. Should this condition not be fulfilled, the accuracy by which the frequencies are reproduced by the extrapolation will be too low, making the extrapolation less valuable or even meaningless if carried too far. In such cases a shorter extrapolation in combination with a window function may be appropriate.[27]

IV. Spectral Analyses

The ultimate use of linear prediction in the field of NMR spectroscopy is for spectral analyses; i.e. as an alternative to the DFT method. This was originally suggested by Barkhuijsen et al.,[14] using the singular value decomposition method for determining the LP coefficients (LPSVD). This approach was based on the work of Baron de Prony[30] and Kumaresan and Tufts.[31] The procedure for the analysis is to estimate the frequencies, ν_j, and the transverse relaxation rates, R_{2j}, of the involved sinusoids from the experimental time-domain signal by using eqs 7–9, as mentioned in the previous section. Subsequently, the intensities, I_j, and the phases, ϕ_j, are recovered from the FID by a linear least-squares calculation with use of eq 1. Therefore, unlike the DFT method that produces the frequency spectrum corresponding to the time-domain signal, the LP analysis results in a complete set of parameters describing the Lorentzians of the frequency spectrum. As a consequence, all the artifacts of the DFT method such as truncation errors, baseline shifts, baseline modulations, intensity distortions, and other drawbacks described in section II, and laboriously remedied in section III, vanish. From the parameters obtained through the LP analysis one can simulate a noiseless spectrum without the artifacts caused by the DFT (in the following named the LP spectrum). This, however, does not necessarily mean that the LP analysis provides a more accurate result than the DFT analysis. In the LP analysis the accuracy is merely given by the uncertainties of the parameters that result from the least-squares calculation, whereas in the DFT analysis the uncertainties are primarily reflected in the noise of the DFT spectrum.

The important difference, however, is that the LP analysis provides the parameters of interest *directly*, whereas the DFT spectrum must be subjected to further analysis in order to yield these parameters. Moreover, in the LP analysis of the FID all of the experimental data points are involved in the determination of the parameters, thus providing the most reliable estimates. A corresponding analysis of the DFT spectrum consists in a nonlinear least-squares analysis of the entire frequency domain signal; i.e. the absorption as well as the dispersion mode spectrum.[6] Although, in principle, this is possible it is, in practice, unfeasible for more complex spectra since the DFT method, unlike the LP method, does not provide information about the number of signals in the spectrum. Recently, however, Kumaresan et al.[32] proposed a new iterative method, requiring only very simple starting values although it works in the frequency domain. Preliminary results seem to indicate the applicability of the method to NMR data.

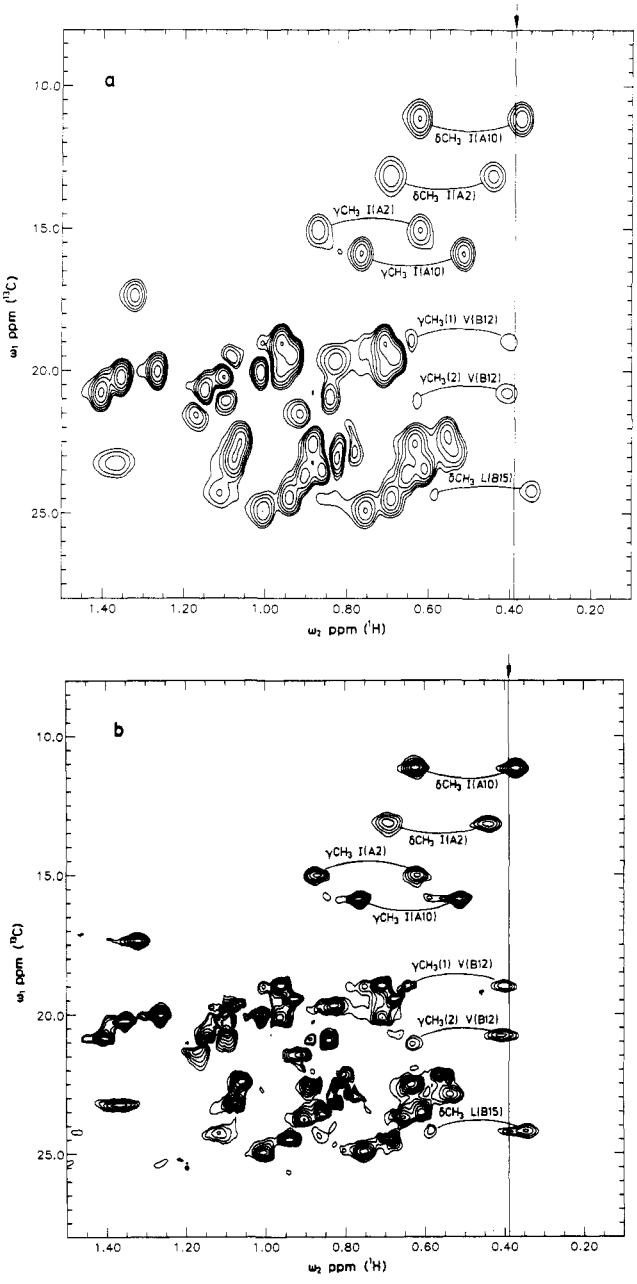


Figure 6. Comparison of two contour plots of the methyl region in the one-bond HMQC C-H correlated spectrum of B9(Asp) insulin: (a) without and (b) with LP extrapolation applied. The vertical line at 0.39 ppm (ω_2) indicates the trace shown in Figure 5. In a, a $\cos^2$–filter function was used in the t_1 domain while an exponential multiplication corresponding to a line broadening of 3 Hz was used in t_2. In b no digital filtering was used either in t_1 or in t_2 (reprinted from ref 28; copyright 1991 ESCOM Science Publishers B.V.).

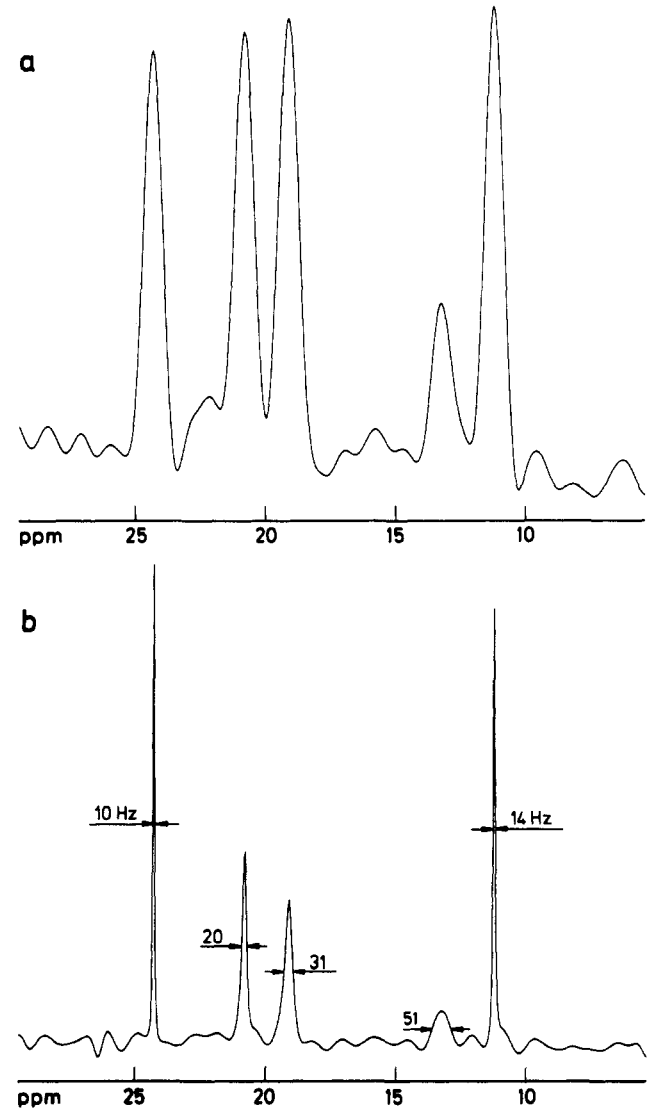


Figure 7. Comparison of two versions of the same F_1 trace in a HSQC C–H correlated two-dimensional spectrum, corresponding to the trace in the HMQC spectrum shown in Figure 5, parts c and d. The t_1 FID consists of 512 experimental data points and was (a) zero filled to 8192 data points and (b) linear prediction extrapolated to 8192 data points. A $\cos^2$ filter was used in a; no digital filter was used in b. The HSQC spectrum was obtained from the same sample as the HMQC spectrum, and under the same experimental conditions except for the difference in the pulse sequences used to obtain the two spectra (reprinted from ref 28; copyright 1991 ESCOM Science Publishers B.V.).

The major drawback of the LP analyses is the computing time necessary. Thus the complete analysis of a one-dimensional spectrum of a protein, like the analysis of the ^{13}C spectrum of insulin (vide infra), takes from a fraction of an hour to a few hours on a VAX workstation, depending on the model and the configuration of the workstation and the efficiency of the algorithm. This should be compared with a fraction of a second to a few seconds for a DFT analysis of the same spectrum. Still, the LP analysis seems worthwhile considering the results. Also, the fact that the time necessary to record the spectrum may be several hours puts the time for the LP analysis in perspective. Moreover, development of more efficient numerical procedures and faster computers will undoubtedly reduce this disadvantage in the near future.

A. One-Dimensional NMR Spectra

Using the approach outlined above it was shown by Led and Gesmar[19] that spectral parameters can be estimated from complicated signals, such as the ^1H FID of the heptapeptide Gly-Phe-Phe-Tyr-Thr-Pro-Lys in D_2O (8.9 mM, pD = 4.4). The conventional phase-corrected DFT spectrum is shown in Figure 8. The LP calculation was performed with 2400 real backward coefficients, using all 8 k data points except the initial 32 that were deleted in order to suppress the effects of electronic filtering and fast decaying transients caused by the rf pulse. Thus eq 9 was solved as 5760 equations in 2400 unknowns. The solution of the 2400 degree

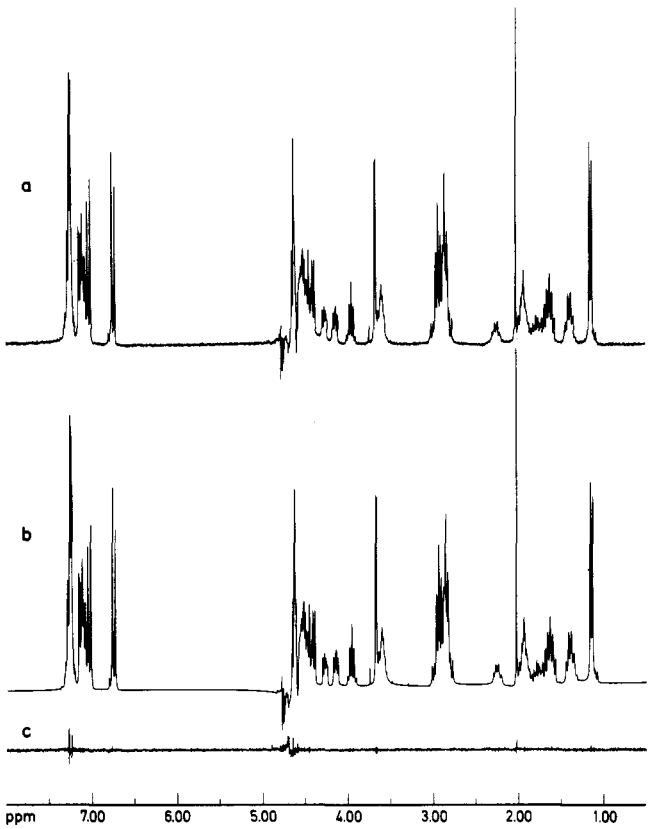


Figure 8. (a) ^{1}H FT NMR absorption spectrum of a 8.9 mM solution of the heptapeptide Gly-Phe-Phe-Tyr-Thr-Pro-Lys in D_2O at pD = 4.4; the FID was recorded at 250.1 MHz in 8 *k* using a sweep width of 2500 Hz; a relatively intense HDO signal was removed by presaturation. (b) LP spectrum of the heptapeptide, displaying the result of an LP calculation based on 2400 LP coefficients. (c) The cosine FT of the residual of the LP estimation (reprinted from ref 19; copyright 1988 Academic Press, Inc.).

polynomial equation corresponding to eq 7 gave 2163 extraneous roots (i.e. roots inside the unit circle) thus leaving 237 intensities and 237 phases to be determined by eq 1. Following this final least-squares calculation, signals with intensities smaller than the corresponding standard deviation were deleted. A total of 78 signals were removed by this procedure. The result of the calculation is displayed graphically in Figure 8b. In Figure 8c the cosine FT of the difference between the observed and the recalculated FID's is shown. Expansions of the aromatic region (6.60–7.45 ppm) are shown in Figure 9. As can be seen from Figures 8 and 9 and Table I the reproduction of the observed signals is excellent, every essential feature of the FT spectrum being reproduced. Small variations that cannot be classified as white noise appear in Figure 8c. Apart from the residual of the HDO signal at 4.696 ppm, caused by the HDO presaturation, these variations are all very narrow (one point) and as such represent very little intensity. They are residuals from narrow signals of modest intensity in which case even very small systematic errors in the FID may appear in the Fourier transform of the difference between the observed and the recalculated signal. Thus errors caused, for example, by the finite digital resolution or field inhomogeneity are likely to be visible in Figure 8, because of the good signal-to-noise ratio.

Gesmar and Led[19] also demonstrated that the LP method is able to extract information from signals with a poor signal-to-noise ratio. This particular capability

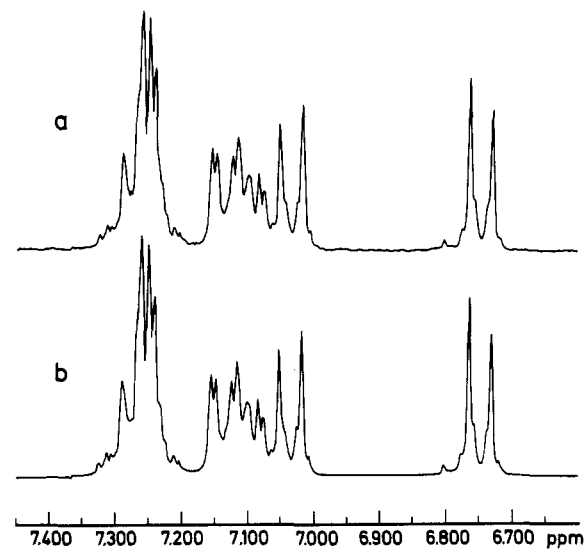


Figure 9. (a) Expansion of the aromatic region (6.60–7.45 ppm) in the DFT spectrum of the heptapeptide (Figure 8a). (b) Expansion of the aromatic region in the LP spectrum of the heptapeptide (Figure 8b). The corresponding parameters estimated by the LP calculation are shown in Table I (reprinted from ref 19; copyright 1988 Academic Press, Inc.).

TABLE I. Estimated Parameters[a] Corresponding to the Expanded LP Spectrum in Figure 11b

j	ν_j, ppm	$\Delta\nu_{(1/2)j}$	I_j[b]	$\sigma(I_j)$[c]	ϕ_j,[b] deg	$\sigma(\phi_j)$,[c] deg
1	6.7206	1.80	22	(3)	−157	(8)
2	6.7307	1.23	188	(3)	−119	(1)
3	6.7393	1.21	36	(3)	−138	(4)
4	6.7564	1.31	48	(3)	−120	(3)
5	6.7650	1.24	232	(3)	−131	(1)
6	6.7765	1.30	17	(3)	−127	(8)
7	6.8035	1.28	12	(3)	−145	(12)
8	7.0065	1.44	23	(3)	−152	(7)
9	7.0174	1.56	256	(4)	−128	(1)
10	7.0251	1.03	27	(3)	−86	(6)
11	7.0341	3.05	62	(7)	22	(6)
12	7.0447	0.77	13	(2)	157	(10)
13	7.0525	1.48	189	(3)	−162	(1)
14	7.0656	1.83	32	(4)	152	(7)
15	7.0764	1.52	83	(4)	−174	(3)
16	7.0843	1.76	134	(5)	−157	(2)
17	7.0977	0.77	9	(3)	130	(17)
18	7.1001	3.68	332	(11)	−144	(2)
19	7.1148	2.64	310	(7)	−126	(1)
20	7.1236	1.19	72	(3)	−109	(3)
21	7.1326	0.83	7	(2)	−159	(17)
22	7.1421	0.47	6	(2)	49	(15)
23	7.1478	1.80	140	(4)	−157	(2)
24	7.1560	1.73	169	(4)	−151	(1)
25	7.2038	0.48	4	(1)	−171	(22)
26	7.2126	1.53	23	(3)	180	(7)
27	7.2239	0.76	13	(2)	−133	(8)
28	7.2320	1.03	39	(2)	−159	(4)
29	7.2422	1.09	179	(3)	174	(1)
30	7.2505	2.28	724	(7)	−174	(1)
31	7.2585	2.56	784	(6)	−105	(1)
32	7.2675	0.14	9	(1)	−110	(5)
33	7.2712	0.38	23	(1)	133	(3)
34	7.2903	2.55	249	(7)	−171	(2)
35	7.2953	1.63	41	(5)	−26	(7)
36	7.3068	1.05	13	(3)	179	(13)
37	7.3124	1.45	30	(3)	−132	(7)
38	7.3250	0.63	9	(2)	−168	(9)
39	7.3658	0.43	1	(1)	120	(49)

[a] The estimates of frequency ($\nu_j = \omega_j/2\pi$), line width ($\Delta\nu_{(1/2)j} = R_{2j}/\pi$), intensity ($I_j$), and phase ($\phi_j$), resulting from a 2400 order LP calculation. Equations 9 and 1 were both solved using 8 *k* data points. [b] The values of I_j and ϕ_j refer to the intensity and phase at the time of the rf pulse. [c] Standard deviation calculated from the inverse of the normal equation matrix corresponding to eq 1.

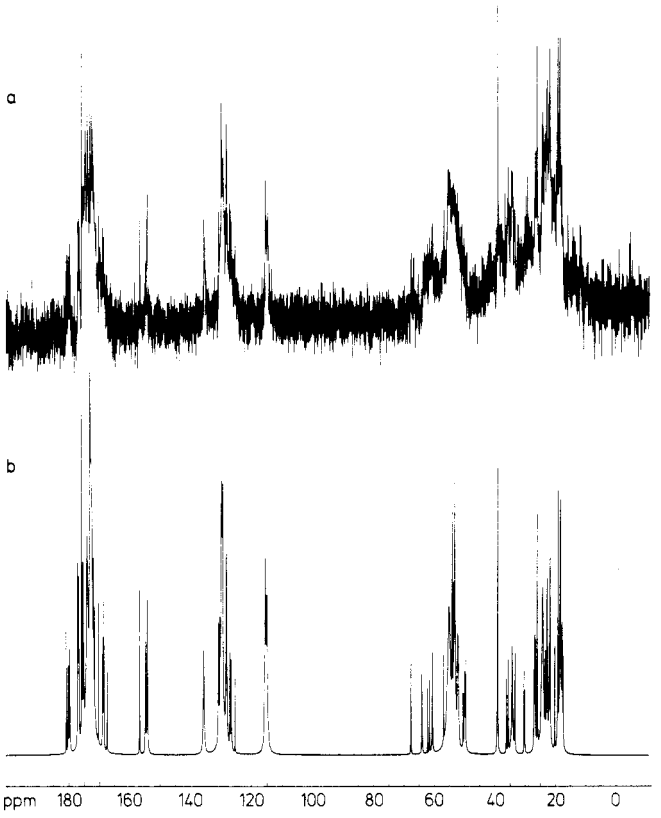


Figure 10. ^{13}C FT NMR (a) and ^{13}C LP NMR (b) absorption spectra of human insulin. The spectra were derived from the same FID. No digital filtering was applied to the FID. The LP spectrum was calculated from the spectral parameters obtained in an analysis with 2400 backward linear prediction coefficients. Signals smaller than three times their standard deviations, and two signals entirely out of phase with the adjacent signals, were deleted, leaving 85 lines in the LP spectrum. In the spectrum displayed all phases are set to zero (reprinted from reference 33, copyright 1988 American Chemical Society).

was exploited by Led et al.[33] who analyzed the differences between the ^{13}C FID of human and porcine insulin using the LP procedure. The frequency spectrum calculated from the parameters obtained in the LP analysis of the ^{13}C FID of the human insulin is shown in Figure 10 together with the conventional FFT spectrum, derived from the same experimental FID. Clearly, most of the signals found by the LP analysis are seen in the DFT spectrum. But unlike the DFT, the LP analysis gives quantitative measures of all the resolved signals in terms of values of the spectral parameters. In particular ν_j and R_{2j} are of interest as potential sources of information about differences between the structures and the dynamics of the two insulins. However, in the LP spectrum in Figure 10b most of the resolved signals are still superpositions, corresponding to carbon atoms in slightly different environments. Hence, a reliable physical interpretation of the parameters derived can be given only for a few of the signals.

In connection with estimation of spectra with very low signal-to-noise ratio a special procedure has been suggested by Delsuc et al.[34] Although its efficiency is difficult to assess, as only few linear prediction coefficients (viz. 140) were applied in the calculation, it looks very interesting and promising.

As far as analyses of low-sensitive, complex ^{13}C spectra of proteins are concerned, proton-detected two-dimensional natural abundance ^{13}C–^{1}H correlation spectra are now accessible, as it appears from the previous section. Also, as reviewed in the following, LP analyses of two-dimensional spectra are now feasible. Considering the enhanced resolution provided by both of these approaches, a combined use of these methods holds great promise as a source of detailed information about dynamics, structure and function of proteins.

B. Two-Dimensional NMR Spectra

In the case of two-dimensional NMR data the limitation of the DFT method mentioned in section II are specifically pronounced, since in practice, truncation in the t_1 dimension is always necessary because of the time consumed during sampling in this dimension. Moreover, as the two-dimensional DFT procedure cannot provide an absorption-like line shape in both dimensions, unless special phase cycling procedures are applied,[35,36] the linear prediction method should provide an effective improvement to two-dimensional spectral analysis. Yet, the time consumption of one-dimensional LP calculations, as described in the previous section, suggests that the extension to two dimensions might not be feasible, regarding the computational power normally available in NMR spectroscopy laboratories at present. Therefore some shortcuts have been proposed.

Schussheim and Cowburn[37] follow a scheme in which a Fourier transformation is applied in the t_2 dimension after which a LP calculation is performed on the t_1 FID's for each value of t_2. Manassen et al.[38] suggest a linear least-squares procedure, according to which the decay rates and resonance frequencies are determined prior to the two-dimensional experiment from a one-dimensional experiment. The amplitude of two-dimensional signals are hereafter calculated from the two-dimensional FID by a least-squares calculation. However, for complicated spectra this is not feasible. Therefore these authors reduce the calculation to several smaller ones by applying the Fourier transformation to the t_2 dimension. In the following the application of the LP principle to two-dimensional NMR free induction decays is described, following the treatment of Gesmar and Led.[39]

In the two-dimensional case, a phase modulated FID has the form

$$f(t_1,t_2) = \sum_{j,h} I_{hj} \exp(i\varphi_{hj}) \exp((i2\pi\nu_h^{(1)} - R_{2h}^{(1)})t_1) \exp((i2\pi\nu_j^{(2)} - R_{2j}^{(2)})t_2) \quad (11)$$

assuming the decay is exponential. Here $\nu_j^{(2)}$, $\nu_h^{(1)}$, $R_{2j}^{(2)}$, and $R_{2h}^{(1)}$ are the frequencies and transverse relaxation rates in the t_2 and t_1 dimension. I_{hj} is the corresponding amplitude and φ_{hj}, the phase. When the signal is sampled at regularly spaced intervals in both dimensions, the following discrete function is obtained:

$$F_{l,k} = \sum_{j=1}^{p_2} D_{lj} \exp((i2\pi\nu_j^{(2)} - R_{2j}^{(2)})\Delta t_2 k) \quad (12)$$

where

$$D_{lj} = \sum_{h=1}^{p_j} I_{hj} \exp(i\varphi_{hj}) \exp((i2\pi\nu_h^{(1)} - R_{2h}^{(1)})\Delta t_1 l) \quad (13)$$

Here Δt_1 and Δt_2 are the sampling intervals in the t_1 and t_2 dimensions, respectively. The value of p_2 is the

number of resonances in the t_2 dimension, and p_j is the number of correlations of the jth resonance. For each value of l, the linear-prediction principle, as stated in eq 9, can be applied to eq 12. Since, furthermore, $\exp((i2\pi\nu_j^{(2)} - R_{2j}^{(2)})\Delta t_2)$ does not depend on l, neither do the backward prediction coefficients. Therefore, the following equation is fulfilled for all k and l:

$$F_{lk} = \sum_{m=1}^{p_2} b_m F_{l,k+m} \tag{14}$$

Equation 14 shows that the entire two-dimensional FID can be included in the determination of the backward coefficients belonging to the t_2 dimension. Consequently the set of linear equations that determine the b_m's is only solved once for a given two-dimensional data set. Again, the linear-prediction order is chosen to be larger than the expected number of resonances in the t_2 dimension. For a detailed description of the solution to eq 14 the reader is referred to refs 19 and 39.

Having found the b_m's, the values of the frequencies, $\nu_j^{(2)}$, and transverse relaxation rates, $R_{2j}^{(2)}$, connected with the t_2 dimension, are determined from

$$C_j = \exp((-i2\pi\nu_j^{(2)} + R_{2j}^{(2)})\Delta t_2) \tag{15}$$

the backward equivalence of eq 8. Again, the rooting of the characteristic polynomial in eq 7 is only performed once for a given two-dimensional data set. Thus, the time used for the most time-consuming parts of the procedure, corresponds closely to the time used in the one-dimensional case. As mentioned in the preceding section the extraneous roots, originating from the excess of linear-prediction coefficients have a tendency of falling inside the unit circle. Because backward prediction is used, the signal roots, C_j, are situated outside the unit circle, due to the positive values of $R_{2j}^{(2)}$ in eq 15. Contrary to the case where forward prediction is used, the two classes of roots are thus easily separated, and the extraneous roots are excluded from the following part of the calculation.

Since the C_j's are known at this point, the D_{lj}'s in eq 12 can be evaluated from another overdetermined system of linear equations, viz.

$$F_{lk} = \sum_{j=1}^{p_2} D_{lj} C_j^{-k} \tag{16}$$

The calculation must be carried out for each value of l, but as the coefficients C_j are independent of l, the number of operations involved in the solution can be reduced considerably.

According to eq 13, D_{lj} itself is a sum of decaying complex exponentials, and as such the linear prediction principle applies once again

$$D_{lj} = \sum_{m=1}^{p_j} B_{jm} D_{l+m,j} \tag{17}$$

B_{jm} representing the t_1 backward coefficients belonging to the jth resonance. For each value of j, a complete linear-prediction calculation and subsequent least-squares determination of the intensity are applied, and thus estimates of the frequencies, $\nu_h^{(1)}$, and relaxation rates, $R_{2h}^{(1)}$, in the t_1 dimension are produced together with the value of intensity, I_{hj}, and phase, φ_{hj}, connecting the frequencies in the two dimensions. *A complete determination of the spectral parameters of the two-dimensional free-induction decay has thus been achieved.*

In connection with the final t_1 calculation, a few details should be noted. As in the former applications of the linear-prediction principle, the number of LP coefficient must be chosen somewhat larger than the expected number of resonances, since D_{lj} is corrupted by noise. The finite variance of D_{lj} is not only caused by the white noise contribution to the individual FID's in the t_2 dimension, but it is also affected by the t_1 noise in the general case. Because of the limited number of frequency correlations for each t_2 resonance, the determination of the spectral parameters can be accomplished involving only few t_1 values. This is an important advantage over the DFT spectrum in which "wiggles" and "ripples", originating from the truncation, would reduce the value of the spectral estimate as described in section II. Still, the resolution obtained in the two-dimensional linear-prediction analysis is further enhanced when the number of applied t_1 values is increased.

The applicability of the quantitative two-dimensional LP procedure has been demonstrated by Gesmar and Led.[39,40] The pairs of spectral parameters $(\nu_j^{(2)}, \nu_h^{(1)})$, $(R_{2j}^{(2)}, R_{2h}^{(1)})$, and the corresponding intensities, I_{hj}, and individual phases, φ_{hj}, were estimated from the ^{1}H COSY free-induction decay of threonine in D_2O and DMSO, the Fourier transform of which is shown in Figure 11a. In the t_2 dimension, 800 real backward prediction coefficients, b_m, were applied, and all 172 rows of the data matrix were included in the normal equations corresponding to eq 14. The first 16 points of each row were deleted in order to suppress possible fast, nonexponential decaying transients. The solution of the characteristic polynomial eq 7 resulted in 114 roots outside the unit circle. As the backward coefficients are real valued, these roots fall in pairs of complex conjugates, thus representing 57 potential signals, i.e. 57 values of $\nu_j^{(2)}$ and $R_{2j}^{(2)}$. The normal equations corresponding to eq 16 were also formed on the basis of the entire two-dimensional free-induction decay, and consequently 57 times 172 complex values of D_{lj} were evaluated. For each of the 57 values of j, a complete linear-prediction calculation was performed on the 172 complex values of D_{lj} applying 64 complex backward prediction coefficients. Thus parameters corresponding to 206 signals were determined by the calculation. The apparent excess consists mostly of small, insignificant signals that can be classified as artifacts.

To demonstrate the general quality of the estimated parameters a two-dimensional Lorentzian line-shaped spectrum was calculated directly from these values as shown in Figure 11b. For more complex 2D spectra with overlapping signals, "correlation leakage" can give rise to artifactual cross peaks. This was described in detail by Gesmar and Led in ref 39, together with a detailed description of the 2D LP spectrum.

It is important to emphasize that the LP spectrum described above is only a practical way of visualizing a large amount of data. Any specific spectral information must be taken from the parameter tables provided by the LP procedure. In Figures 12 and 13 smaller sections of the DFT spectrum and the LP spectrum are expanded. The corresponding parameters are listed in Table II and III. As can be seen, the

TABLE II. Estimated Parameters[a] Corresponding to the Expanded LP Spectrum in Figure 12b

j	h	$(\nu_j^{(2)}, \nu_h^{(1)})$, Hz	$(\Delta^{(2)}_{(1/2)j}, \Delta^{(1)}_{(1/2)h})$, Hz	I_{hj}[b]	$\sigma(I_{hj})$[c]	φ_{hj},[b] deg	$\sigma(\varphi_{hj})$,[c] deg
5[e]	1	(255.9, 257.8)	(1.0, 0.2)	2.0	0.1	−65	2
5[e]	2	(255.9, 278.5)	(1.0, 3.1)	2.0	0.1	−112	4
6[e]	1	(263.2, 262.2)	(2.0, −0.1)[d]	8.1	0.1	−112	2
6[e]	2	(263.2, 281.1)	(2.0, −3.4)[d]	1.2	0.06	−139	4
7[e]	1	(267.1, 267.0)	(1.2, 5.3)	20.7	0.1	−135	2
7[e]	2	(267.1, 279.0)	(1.2, 4.8)	9.8	0.1	−126	4
7[e]	3	(267.1, 284.2)	(1.2, −35)[d]	6.6×10^{-5}	1.6×10^{-6}	−45	2
8	1	(269.3, 269.4)	(1.0, 0.3)	331.8	0.5	−106	0
8	2	(269.3, 275.9)	(1.0, 1.0)	386.9	0.5	−106	0
8[e]	3	(269.3, 286.7)	(1.0, −31)[d]	4.0×10^{-4}	1.2×10^{-5}	139	2
9[e]	1	(272.3, 272.0)	(1.1, −16)[d]	2.0×10^{-1}	1.6×10^{-3}	−128	0
9[e]	2	(272.3, 272.2)	(1.1, 2.9)	7.6	0.3	−110	2
10	1	(276.0, 269.2)	(1.0, 0.3)	301.5	0.4	−115	0
10	2	(276.0, 275.9)	(1.0, 1.2)	396.7	0.4	−111	0
10[e]	3	(276.0, 284.5)	(1.0, −26)[d]	2.0×10^{-3}	6.7×10^{-5}	155	2

[a] The estimates of frequency pair $(\nu_j^{(2)}, \nu_h^{(1)})$, linewidth pair $[(\Delta\nu^{(2)}_{(1/2)j}, \Delta\nu^{(1)}_{(1/2)h}) = (R^{(2)}_{2j}/\pi, R^{(1)}_{2h}/\pi)]$, intensity (I_{hj}), and phase (φ_{hj}), resulting from an LP calculation applying 800 real LP-coefficients in the t_2 dimension and 64 complex LP coefficients in the t_1 dimension. All 172 × 1024 points of the FID were included, except for the first 16 points in the t_2 dimension. [b] The values of I_{hj} and φ_{hj} refer to the intensity and phase at $t_1 = t_2 = 0$. [c] Standard deviation calculated from the inverse of the normal equation matrix corresponding to eq 1. [d] The significance of negative line width is discussed in ref 7. [e] Instrumental artifacts.

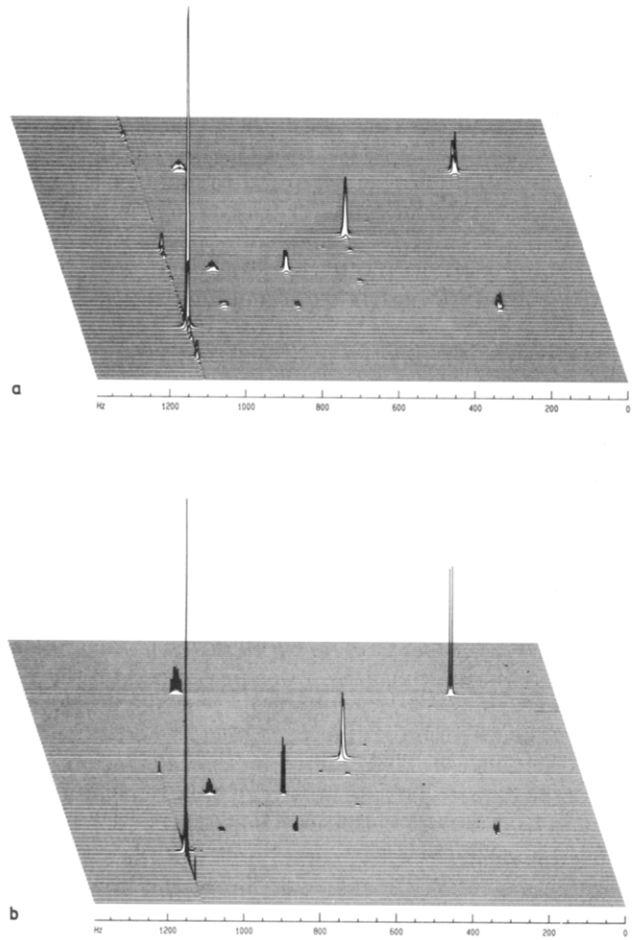


Figure 11. (a) Phase modulated 250.13-MHz ^{1}H two-dimensional FT COSY power spectrum of threonine in D_2O and deuterated DMSO recorded in 172 × 1024 data points applying a spectral width of 1392.76 Hz in both dimensions. The FID was premultiplied by an unshifted two-dimensional sine bell before Fourier transformation in 128 × 1024 points. The resonance at 616.5 Hz is due to DMSO (2.5 ppm). (b) Two-dimensional-LP spectrum (see text) of the FID whose FT spectrum is shown in a (reprinted from ref 39; copyright 1989 Academic Press, Inc.).

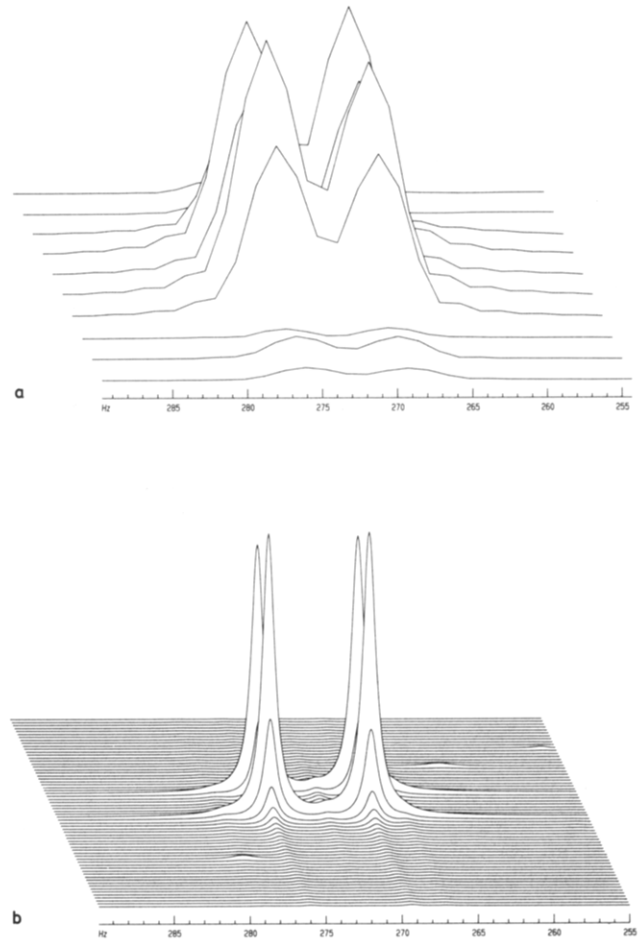


Figure 12. (a) Expansion of the region between 255.0 and 290.0 Hz in both dimension of the FT spectrum in Figure 11a. In this case the FID was "zero filled" until 256 points in the t_1 dimension after the sine-bell filtering. The expansion shows the (CH_3, CH_3) multiplet. (b) LP spectrum of the same region calculated from the estimated parameters in Table II (reprinted from ref 39; copyright 1989 Academic Press).

resolution of the LP method is beyond what can be obtained by inspection of the DFT spectrum, although the latter has been zero filled in both dimensions and premultiplied by a two-dimensional sine bell. It is also worthy of note that the phase estimates are exactly as one should expect: The signals on the diagonal in Table II are in phase while the off-diagonal signals in Table III are pairwise in counterphase.

As mentioned above, a number of small extraneous signals is found by the procedure, particularly in the

TABLE III. Estimated Parameters[a] Corresponding to the Expanded LP Spectrum in Figure 13b

j	h	$(\nu_j^{(2)}, \nu_h^{(1)})$, Hz	$(\Delta\nu_{(1.2)j}^{(2)}, \Delta\nu_{(1/2)h}^{(1)})$, Hz	I_{hj}[b]	$\sigma(I_{hj})$[c]	φ_{hj},[b] deg	$\sigma(\varphi_{hj})$,[c] deg
36	1	(980.2, 269.3)	(1.2, 1.1)	46.5	0.1	73	0
36	2	(980.2, 276.0)	(1.2, 0.6)	45.5	0.1	-109	0
37	1	(985.2, 269.2)	(0.9, 1.1)	35.5	0.1	71	0
37	2	(985.2, 275.6)	(0.9, 0.8)	35.3	0.1	-109	0
38	1	(987.2, 269.2)	(0.7, 0.6)	43.6	0.2	36	0
38	2	(987.2, 275.9)	(0.7, 0.4)	43.9	0.2	-146	0
39	1	(992.1, 269.1)	(1.4, -0.6)[d]	53.4	0.3	58	0
39	2	(992.1, 275.9)	(1.4, -0.4)[d]	54.1	0.3	-126	0
40	1	(993.2, 269.6)	(1.1, 0.0)[d]	63.1	0.3	-103	0
40	2	(993.2, 276.2)	(1.1, 0.2)	63.1	0.3	75	0
41	1	(998.2, 269.3)	(0.7, 0.8)	44.3	0.1	-103	0
41	2	(998.2, 276.0)	(0.7, 0.7)	44.2	0.1	77	0
42	1	(1000.2, 269.7)	(1.0, 1.1)	38.1	0.2	-140	0
42	2	(1000.2, 276.1)	(1.0, 0.8)	37.2	0.2	40	0
43	1	(1005.1, 269.4)	(1.2, 0.8)	44.1	0.1	-137	0
43	2	(1005.1, 276.0)	(1.2, 0.9)	44.2	0.1	43	0

[a] The estimates of frequency pair $(\nu_j^{(2)}, \nu_h^{(1)})$, line width pair $[(\Delta\nu_{(1/2)j}^{(2)}, \Delta\nu_{(1/2)h}^{(1)}) = (R_{2j}^{(2)}/\pi, R_{2h}^{(1)}/\pi)]$, intensity (I_{hj}), and phase (φ_{hj}), resulting from an LP calculation applying 800 real LP-coefficients in the t_2 dimension and 64 complex LP coefficients in t_1 dimension. All 172 × 1024 points of the FID were included, except for the first 16 points in the t_2 dimension. [b] The values of I_{hj} and φ_{hj} refer to the intensity and phase at $t = 0$. [c] Standard deviation calculated from the inverse of the normal equation matrix corresponding to eq 1. [d] The nonpositive line width is due to the uncertainty caused by the short t_1 data record (see ref 7).

vicinity of intense signals. This can also be seen in Figure 13. It should be emphasized that these are not removed by the procedure because they are real signals, in the sense that they conform with eq 12. Only physical knowledge of the system under study can separate them from the relevant signals.

V. Conclusion

The qualitative LP method, i.e. backward and forward LP extrapolation of the time-domain signal before the DFT, is able to improve the quality of NMR spectra considerably. Intensity distortions, nonlinear phase problems, "wiggles", and pseudobaselines are of much less significance in the LP-enhanced spectra than in the normal DFT spectra.

The quantitative LP method can *directly* estimate the frequencies, transverse relaxation rates, intensities, and phases from both one- and two-dimensional NMR data. It does not suffer from the previously described deficiencies of the DFT. In the two-dimensional case there is no need for phase-cycling procedures in order to avoid dispersion-like line shapes. Finally, it produces tables of spectral parameters, making further automated analyses possible without the need for peak finding procedures.

The major drawback of the LP method is the extensive computing time. However, as mentioned above the development of more efficient numerical procedures and faster computers will without any doubt eliminate this disadvantage in the future.

VI. Abbreviations

AD	analogue to digital
COSY	two-dimensional correlation spectroscopy
DANTE	delays alternating with mutation for tailored excitation
DFT	discrete Fourier transform
DMSO	dimethyl sulfoxide
FID	free infuction decay
hGH	human growth hormone
HMBC	two-dimensional heteronuclear multiple-bond correlation spectroscopy
HMQC	two-dimensional heteronuclear multiple quantum correlation spectroscopy
HOHAHA	two-dimensional homonuclear Hartmann–Hahn spectroscopy

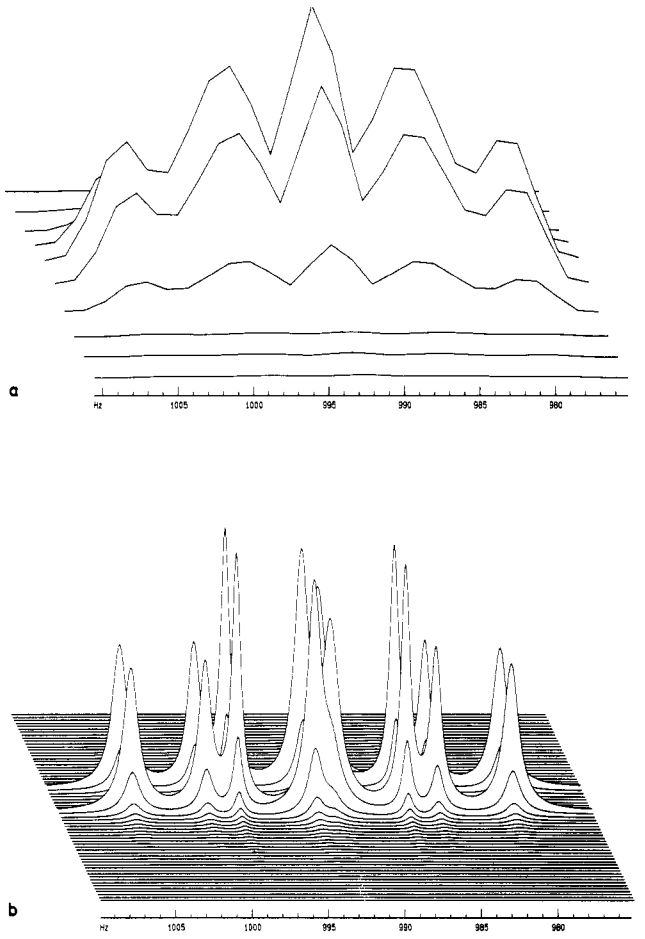


Figure 13. (a) Expansion of the region between 255.0 and 290.0 Hz in the F_1 dimension and 975.0 and 1010.0 Hz in the F_2 dimension of the FT spectrum in Figure 11a. In this case the FID was "zero filled" until 256 points in the t_1 dimension after the sine-bell filtering. The expansion shows the (β-CH, CH_3) multiplet. (b) LP spectrum of the same region, calculated from the estimated parameters in Table III. As seen in Table III the signals are pairwise in counterphase, but for reasons of clearness the phases have all been set to zero in the spectrum (reprinted from reference 39; copyright 1989 Academic Press).

HSQC	two-dimensional heteronuclear single quantum correlation spectroscopy
LP	linear prediction
LPSVD	linear prediction signular value decomposition
LPZ	linear prediction and z transform
MEM	maximum entropy method
MLM	maximum likelihood method
NOESY	two-dimensional nuclear Overhauser enhancement spectroscopy

VII. Acknowledgments

This work was supported by Julie Damm's Studiefond, the Danish Technical Research Council (J.No. 16-3922.H, and 16-4679.H) and the Ministry of Industry (J.No. 85886). We are grateful to Ms. Majbritt Jørgensen for her care and patience when preparing the typewritten manuscript.

VIII. References

(1) Ernst, R. R. *Adv. Magn. Reson.* **1966**, *2*, 1.
(2) Ernst, R. R.; Anderson, W. A. *Rev. Sci. Instrum.* **1966**, *37*, 93.
(3) Sibisi, S.; Skilling, J.; Brereton, R. G.; Laue, E. D.; Staunton, J. *Nature* **1984**, *311*, 446.
(4) de Beer, R.; van Ormondt, D.; Pignappel, W. W. F. *Proc. ICASSP* **1989**, *3*, 1504.
(5) Stephenson, D. S. *Prog. NMR Spectrosc.* **1988**, *20*, 515.
(6) Abildgaard, F.; Gesmar, H.; Led, J. J. *J. Magn. Reson.* **1988**, *79*, 78.
(7) Gesmar, H.; Led, J. J.; Abildgaard, F. *Prog. NMR Spectrosc.* **1990**, *22*, 255.
(8) Otting, G.; Widmer, H.; Wagner, G.; Wüthrich, K. *J. Magn. Reson.* **1986**, *66*, 187.
(9) Plateau, P.; Dumas, C.; Guéron, M. *J. Magn. Reson.* **1983**, *54*, 46.
(10) Hoult, D. I.; Richards, R. E. *Proc. R. Soc. Lond. Ser.* **1975**, *A344*, 311.
(11) Tirendi, C. F.; Martin, J. F. *J. Magn. Reson.* **1989**, *81*, 577.
(12) Marion, D.; Bax, A. *J. Magn. Reson.* **1989**, *83*, 205.
(13) Kay, S. M.; Marple, S. L., Jr. *Proc. IEEE* **1981**, *69*, 1380.
(14) Barkhuijsen, H.; de Beer, R.; Bovée, W. M. M. J.; van Ormondt, D. *J. Magn. Reson.* **1985**, *61*, 465.
(15) Tang, J.; Lin, C. P.; Bowman, M. K.; Norris, J. R. *J. Magn. Reson.* **1985**, *62*, 167.
(16) Barkhuijsen, H.; de Beer, R.; van Ormondt, D. *J. Magn. Reson.* **1985**, *64*, 343.
(17) Barone, P.; Guidoni, L.; Massaro, E.; Viti, V. *J. Magn. Reson.* **1987**, *73*, 23.
(18) Barkhuijsen, H.; de Beer, R.; van Ormondt, D. *J. Magn. Reson.* **1987**, *73*, 553.
(19) Gesmar, H.; Led, J. J. *J. Magn. Reson.* **1988**, *76*, 183.
(20) Tang, J.; Norris, R. *J. Magn. Reson.* **1986**, *69*, 180.
(21) Kumaresan, R. *IEEE Trans. ASSP-31* **1983**, 217.
(22) Brüschweiler, R.; Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *J. Magn. Reson.* **1988**, *78*, 178.
(23) Davis, D. G. *J. Magn. Reson.* **1989**, *83*, 212.
(24) Geen, H.; Freeman, R. *J. Magn. Reson.* **1990**, *87*, 415.
(25) Led, J. J.; Abildgaard, F.; Gesmar, H. *J. Magn. Reson.* **1991**, *93*, 659–665.
(26) Bodenhausen, G.; Freeman, R.; Morris, G. A. *J. Magn. Reson.* **1976**, *23*, 171.
(27) Olejniczak, E. T.; Eaton, H. L. *J. Magn. Reson.* **1990**, *87*, 628.
(28) Led, J. J.; Gesmar, H. *J. Biomol. NMR* **1991**, *1*, 237–246.
(29) Gesmar, H.; Led, J. J. *Computational Aspects of the Study of Biological Macromolecules*; Hoch, J. C., Ed.; Nato Advanced Study Institute Series; Plenum: New York, in press.
(30) Prony, G. R. B. *J. de L'Ecol Polytechnique, Paris* **1795**, *1*, 24.
(31) Kumaresan, R.; Tufts, D. W. *IEEE Trans. ASSP-30* **1982**, 833.
(32) Kumaresan, R.; Ramalingam, C. S.; van Ormondt, D. *J. Magn. Reson.* **1990**, *89*, 562.
(33) Led, J. J.; Gesmar, H.; Hejnaes, K. R.; Hansen, F. B. *J. Am. Chem. Soc.* **1988**, *110*, 4165.
(34) Delsuc, M. A.; Ni, F.; Levy, G. C. *J. Magn. Reson.* **1987**, *73*, 548.
(35) Marion, D.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967.
(36) States, D. J.; Haberkorn, R. A.; Ruben, D. J. *J. Magn. Reson.* **1982**, *48*, 286.
(37) Schussheim, A. E.; Cowburn, D. *J. Magn. Reson.* **1987**, *71*, 371.
(38) Manassen, Y.; Navon, G.; Moonen, C. T. W. *J. Magn. Reson.* **1987**, *72*, 551.
(39) Gesmar, H.; Led, J. J. *J. Magn. Reson.* **1989**, *83*, 53.
(40) Gesmar, H.; Led, J. J. *J. Magn. Reson.* **1988**, *76*, 575.