Journal of Biomolecular NMR, 1 (1991) 237–246
ESCOM

J-Bio NMR 022

Linear prediction enhancement of 2D heteronuclear correlated spectra of proteins

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 14 February 1991
Accepted 16 May 1991

Keywords: Linear prediction; Heteronuclear correlation; 2D NMR spectra; Protein; Insulin

SUMMARY

Linear prediction has been used to extrapolate the t_1 domain of natural abundance ^1H–^{13}C correlated two-dimensional (2D) FIDs of insulin. The FIDs were obtained by two different heteronuclear correlation experiments, one that utilizes heteronuclear multiple-quantum coherence during t_1, and one that utilizes ^{13}C single-quantum coherence. It is shown that the enhancement of the resolution and sensitivity in the F_1 dimension of the Fourier transform spectrum that results from the linear prediction extrapolation allows the t_1 domain to be confined to a relatively short time period where the signal intensity is at maximum. In particular, it is found that the enhancement thus obtained is sufficiently good to allow an observation of the difference between the F_1 line widths in the single-quantum and double-quantum coherence spectra.

INTRODUCTION

Two-dimensional (2D) heteronuclear correlated spectra, correlating either ^{13}C or ^{15}N nuclei with protons, are of increasing interest in studies of the structure and dynamics of proteins (Bax et al., 1989; Wagner, 1989). This is due partly to the larger chemical-shift dispersion of the heteronuclei compared to ^1H that facilitates the assignment both of the heteronuclear spectrum and the proton spectrum, and partly to the complementary information, for example the reorientation rate of individual C–H bonds, that can be obtained from such spectra. With the proton detection of the ^1H-heteronuclear correlated experiments, nowadays done routinely on commercial spectrometers, the sensitivity of these experiments, compared with the corresponding ^1H–^1H correlated 2D experiments, is decreased only by the abundance of the heteronucleus.

In the case of proteins, the sensitivity of the ^1H-detected heteronuclear correlated experiments is still too low for routine use of these experiments, unless isotopic enrichment is applied. Consequently, even in the case of the relatively sensitive ^{13}C nucleus, only a few extensive studies have

been reported (Wagner and Brühwiler, 1986; Gao et al., 1990; Kessler et al., 1990) in which the assignment of the heteronuclear spectrum of a protein has been made on the basis of a 2D ^{1}H-detected heteronuclear correlated experiment with the heteronucleus in natural abundance.

This sensitivity problem can, in principle, be alleviated by confining the data acquisition in the t_1 domain to a sufficiently small part of the FIDs where the signal intensity is maximum. Although this inevitably reduces the spectral resolution in the F_1 dimension correspondingly, signal overlap might still be avoided due to the large chemical-shift dispersion of the heteronucleus. More seriously, this procedure would necessitate a substantial digital filtering in order to avoid truncation errors in the t_1 domain (Gesmar et al., 1990), which would further deteriorate the resolution and reduce the achieved sensitivity. However, if the FIDs in the t_1 domain are extrapolated by linear prediction (LP), the truncation error can be reduced considerably, as demonstrated by Tirendi and Martin (1989), or in practice eliminated if the applied algorithm allows a sufficiently long extrapolation of the FIDs. Although this extension does not add new information to the data, it prevents a distortion of the information already contained in the experimental data and, thus, allows a better exploitation of this information. It is also noteworthy that the LP extrapolation does not rely on the assumption of single exponential decay. As the LP coefficients depend not only on the frequencies but also on the decay rates, differences in decay rates for sinusoids with identical frequencies are also transferred to the extrapolation, if they are significant. Obviously, a similar spectrum, free of truncation artifacts, could be obtained by an *experimental* extension of the FID. However, aside from the use of an unreasonable amount of spectrometer time, an acquisition toward the end of the decay of the FID would increase the noise level in the spectrum. Therefore, assuming that the resolution given by the truncated t_1 domain is sufficiently high, the LP extension is preferable to an experimental extension.

Here we demonstrate the advantage of applying forward LP extrapolation to the t_1 domain of one-bond heteronuclear correlated 2D NMR spectra of proteins. In particular, it is shown that LP extrapolation provides a unique approach for the exploitation of the inherently higher resolution of the heteronuclear single-quantum coherence (HSQC) experiment compared with the heteronuclear multiple-quantum (HMQC) experiment.

MATERIALS AND METHODS

The protein that has been used here as an example is an insulin analogue, B9(Asp) insulin, in which the serine residue in the 9th position of the B-chain of native human insulin has been substituted by an aspartic acid residue. The insulin analogue was supplied by Novo Nordisk. As found recently (Kristensen et al., 1991) this analogue shows less line-broadening than the native insulin under the applied experimental conditions. Still, it exists primarily as dimers under these conditions, resulting in relatively broad resonances in both the proton and the carbon spectra.

Two different C–H correlated experiments were carried out. The pulse sequences of these experiments are presented in Fig. 1. The first pulse sequence (Müller, 1979) utilizes heteronuclear multiple-quantum coherence during t_1, while the second pulse sequence (Bodenhausen and Reuben, 1980) utilizes heteronuclear single-quantum coherence during this period. The spectra were recorded in the antiphase mode in both cases; that is, without refocusing before the acquisition, and without ^{13}C decoupling during the acquisition. As pointed out by Bax et al. (1990), the dephasing caused by the passive ^{1}H–^{1}H couplings is not refocused during t_1 in sequence (a) in Fig. 1, since

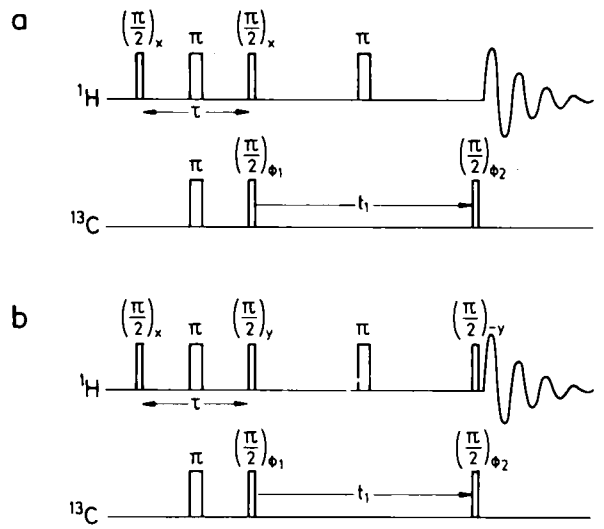


Fig. 1. Pulse sequences for ^{1}H-detected one-bond heteronuclear chemical-shift correlation; (a) HMQC scheme utilizing heteronuclear multiple-quantum coherence during the evolution period, according to Müller (1979). (b) HSQC scheme utilizing ^{13}C single-quantum coherence during the evolution period, according to Bodenhausen and Reuben (1980). In both schemes the basic phase cycle was Φ_1: +x,−x,+x,−x; Φ_2: +x,+x,−x,−x; receiver: +,−,−,+. In addition, in scheme (b) the phases of the two $(\pi/2)(^1H)$ and the $\pi(^{13}C)$ pulse of the INEPT preparation pulse sequence, and the $\pi(^1H)$ pulse in the middle of t_1 were cycled independently in steps of 180°, resulting in a 32-step phase cycle. The sign of the receiver was changed together with each change of the phases of the $(\pi/2)(^1H)$ pulses. Quadrature detection along t_1 was achieved by incrementation of Φ_1 by 90° for subsequent t_1 increments (TPPI, Drobney et al., 1979; Bodenhausen et al., 1980; Marion and Wüthrich, 1983). A delay time $\tau = J_{HC}/2$, corresponding to a ^{1}H–^{13}C one-bond coupling constant $J_{HC} = 140$ Hz, was applied.

all protons experience the effect of the nonselective 180° pulse in the middle of the evolution period. This dephasing will result in a homonuclear J coupling structure in the F_1 dimension (Kay et al., 1989). For proteins, these coupling constants are often smaller than the F_1 line widths, and the couplings will therefore only show up as an increase of the line widths.

In sequence b in Fig. 1, the active proton magnetization lies along the z-axis during the evolution period, and, therefore, no dephasing of the proton magnetization due to passive ^{1}H–^{1}H couplings takes place during this period. Consequently, no splitting or line-broadening due to coupling appear in the F_1 dimension. Also, the difference in the transverse relaxation rate during t_1 in the two experiments may give rise to different line widths in the two cases, as discussed in detail by Bax et al. (1990). Thus, in the HMQC experiment (Fig. 1a) the relaxation during t_1 is given by the relaxation of the double-quantum coherence, I_xS_y, while in the HSQC experiment (Fig. 1b) it is determined by the average relaxation of the two single-quantum coherences, S_x and I_zS_y.

The two experiments were carried out on the same sample under identical experimental conditions; i.e. at 305 K using a 7 mM D_2O solution of B9 (Asp) insulin at pD 1.82 contained in a 5-mm sample tube. The FIDs were recorded with nonspinning sample on a Bruker AM 500 NMR spectrometer equipped with an inverse probe and a BSV-8 amplifier. Both 2D FIDs were represented by 512 data points in t_1 and 8192 data points in t_2. The ^{1}H and ^{13}C sweep widths were 20 000 Hz and 25 000 Hz, respectively, resulting in t_1 and t_2 acquisition times of 10.2 ms and 205

ms, respectively. In both experiments, each FID consisted of 384 scans with a total delay of 1.63 s between scans. The recording time was 89 h.

The experimental data was processed on a VAXstation 3100/38 using locally developed software. The experimental FIDs were extended to a complete decay by an LP calculation. In the

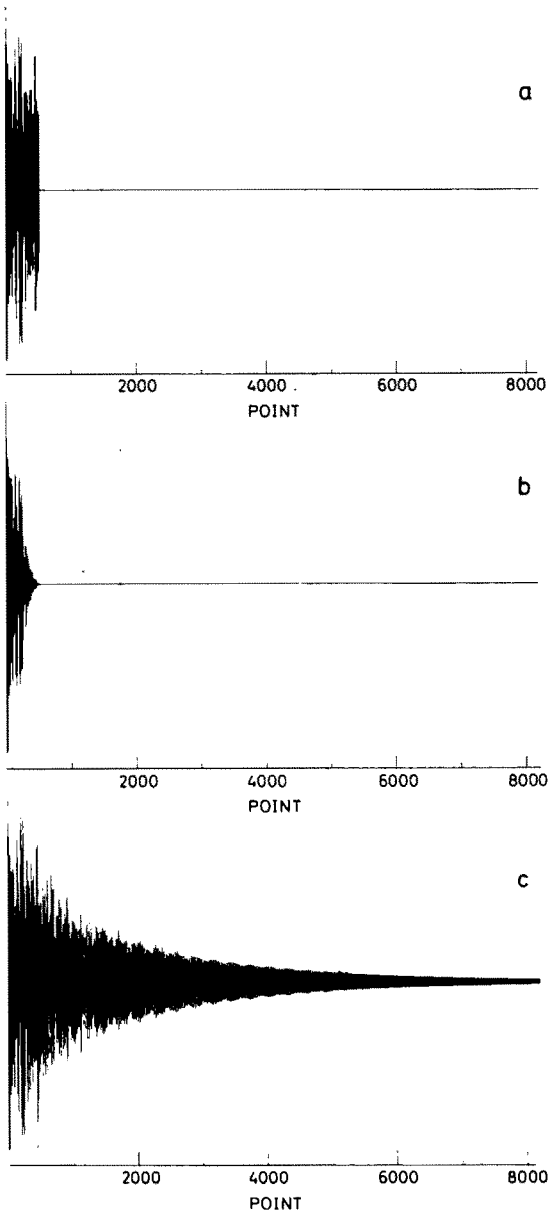


Fig. 2. An experimental FID along t_1 (512 data points, t_1 acquisition time 10.2 ms), corresponding to the trace indicated by an arrow in the HMQC C–H correlated 2D spectrum in Fig. 4; (a) zero-filled to 8192 data points; (b) zero-filled to 8192 data points and digitally filtered using a $\cos^2$ window function; (c) LP extrapolated to 8192 data points. For details of the LP extrapolation, see text.

present examples, 150 forward coefficients were derived from the 512 experimental data points using 354 equations. The 150 roots of the characteristic polynomial (Kay and Marple, 1981) were calculated from the LP coefficients. Because of the noise, a number of roots fell outside the unit circle (Gesmar and Led, 1991), corresponding to non-decaying frequency components of the extrapolated signal. In order to assure the decay of the extrapolation, these roots were reflected to fall inside the unit circle (Zhu and Bax, 1990), and a modified set of prediction coefficients was calculated. It should be emphasized that it is essential to apply this stability correction in cases where the FID is prolonged by more than just a small fraction. Here, the FIDs were extrapolated from

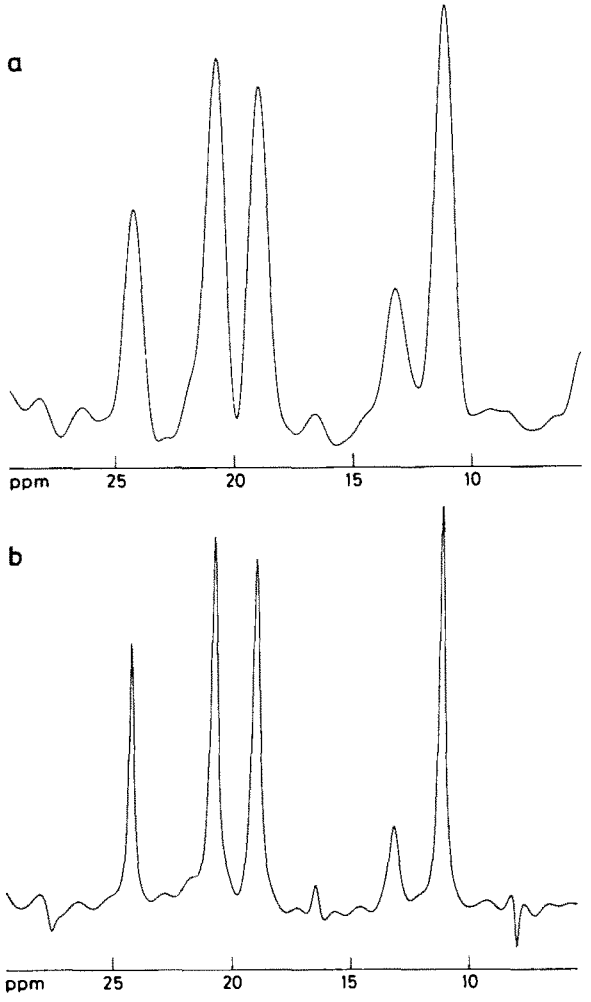
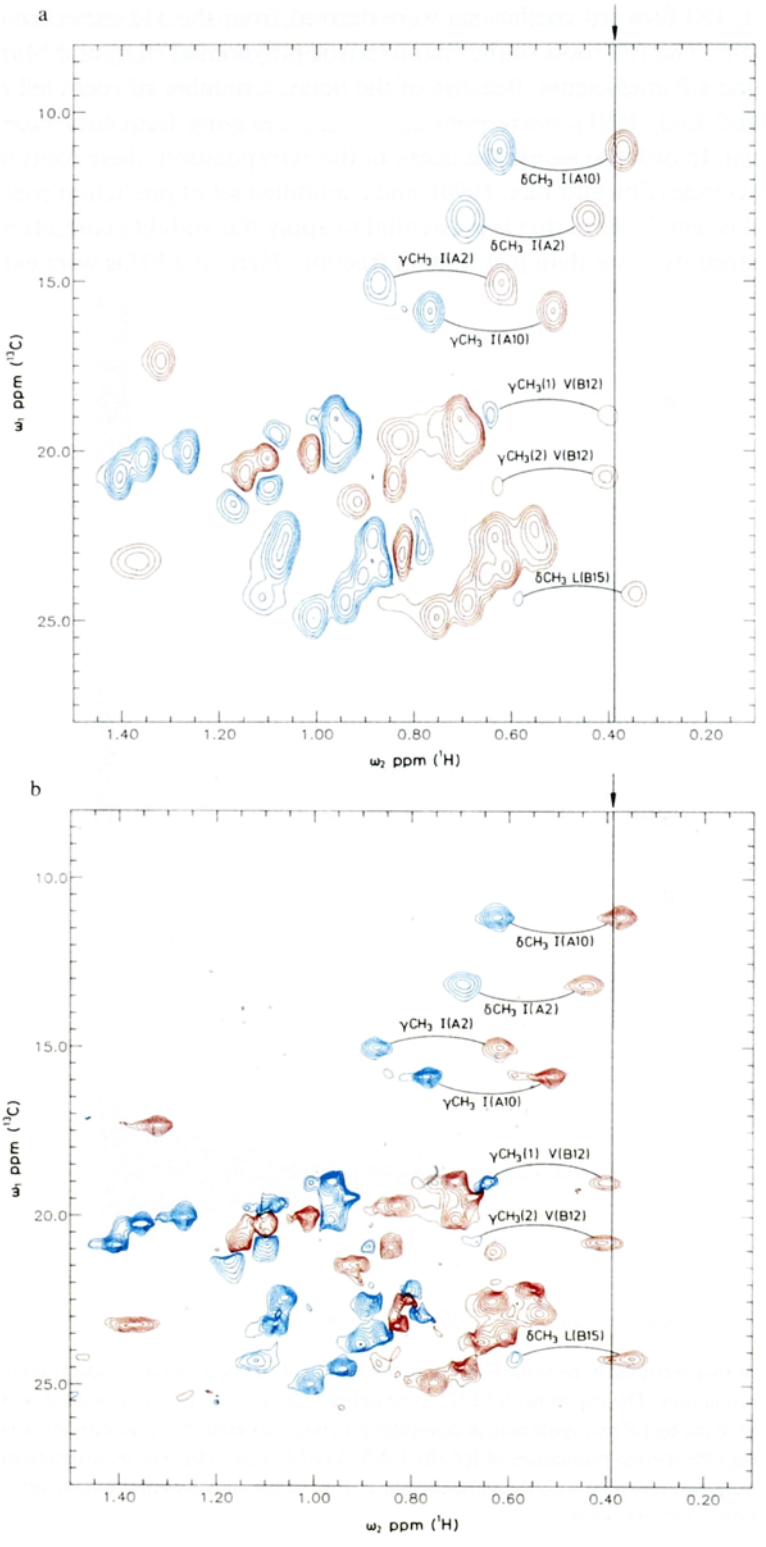


Fig. 3. Comparison of two versions of the same F_1 trace in the HMQC C–H correlated 2D spectrum in Fig. 4; (a) without and (b) with LP extrapolation. The experimental FID, containing 512 data points, was extended to 8192 data points, in (a) by zero-filling, and in (b) by LP extrapolation. A $\cos^2$-filter function was used in (a); no filtering was applied in (b). The trace was taken through the methyl resonances of I(A10), I(A2), V(B12), and L(B15) as indicated in the spectrum in Figs. 4a and 4b. The spectrum was recorded with the pulse scheme in Fig. 1a that involves multiple-quantum coherence during t_1. The digital resolution is 3.05 Hz/point.


a
10.0
15.0
20.0
25.0
ω₁ ppm (¹³C)
δCH₃ I(A10)
δCH₃ I(A2)
γCH₃ I(A2)
γCH₃ I(A10)
γCH₃(1) V(B12)
γCH₃(2) V(B12)
δCH₃ L(B15)
1.40
1.20
1.00
0.80
0.60
0.40
0.20
ω₂ ppm (¹H)
b
10.0
15.0
20.0
25.0
ω₁ ppm (¹³C)
δCH₃ I(A10)
δCH₃ I(A2)
γCH₃ I(A2)
γCH₃ I(A10)
γCH₃(1) V(B12)
γCH₃(2) V(B12)
δCH₃ L(B15)
1.40
1.20
1.00
0.80
0.60
0.40
0.20
ω₂ ppm (¹H)

512 points to 8192 points. The extrapolation, including the rooting of the characteristic polynomial, took 50 CPUs per FID. Also, the first seven t_2 values were corrected by backward LP estimation for each value of t_1 (Marion and Bax, 1989). All LP calculations were made using the program LPEXTRAPOL.

RESULTS AND DISCUSSION

The effect of applying normal zero-filling and digital filtering versus LP extrapolation to an experimental FID along t_1 is illustrated in Fig. 2 (the FID corresponds to the trace in the HMQC experiment indicated in the contour plot in Fig. 4).

Figure 3 shows the Fourier transform of the $\cos^2$-filtered and the LP-extrapolated FID in Fig. 2. No filtering was applied in the latter case. A comparison of the two spectra shows that the LP extrapolation of the FIDs results in a substantial improvement of the spectrum, both with respect to the width and the shape of the resonances. The enhanced resolution that results from these improvements is further displayed in the contour plot of the methyl region in Fig. 4. Thus, signals that are indistinguishable in the spectrum (Fig. 4a), obtained from the zero-filled and $\cos^2$-filtered t_1 domain, are resolved in the spectrum (Fig. 4b), obtained from the LP-extrapolated t_1 domain. The assignments indicated in Fig. 4 were made on the basis of the proton frequencies and the assignment of the proton spectrum (Kristensen et al., 1991).

The LP extrapolation technique was also applied to the HSQC data of the B9(Asp) insulin. Figures 5 and 6 show the results, corresponding to the results of the HMQC experiment in Figs. 3 and 4. From a comparison of Figs. 3b and 5b, it is immediately apparent that the line widths are significantly smaller in the HSQC spectrum than in the HMQC spectrum. It is also apparent from a comparison of Figs. 3a and 5a that the difference in line width in the two cases would not have been appreciated had LP extrapolation not been applied, since the difference is completely overshadowed by the extra line width imposed by the filter function. The distortions of the line shape that are seen in the contour plots in Figs. 4b and 6 are due to the uncertainty of the determination of the LP coefficients that transfer the relaxation rates to the extrapolation. As the LP extrapolations of the individual slices are done independently, the variation of the width along F_2 of a given 2D resonance can be taken as a qualitative measure of the uncertainty. The differences between the line widths that appear in the LP-extrapolated HSQC and HMQC spectra in Figs. 4b and 6 are significantly larger than these variations in line width, and are, therefore, real. Bax et al. (1990) observed a similar difference in line width in a corresponding set of HMQC and HSQC C–H correlated spectra of staphylococcal nuclease. However, these authors used a t_1 acquisition time considerably longer than here (128 ms versus 10.2 ms), in order to allow an almost complete decay of the FID, thereby avoiding a digital filtering in the t_1 dimension that would have increased the line

←

Fig. 4. 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 Fig. 3. The spectrum was recorded with the pulse scheme in Fig. 1a that involves multiple-quantum coherence during t_1, and detection of antiphase signals. Positive signals are indicated by red contours, and negative signals by blue contours. 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 in either t_1 or t_2. For further details see the legend of Fig. 3.

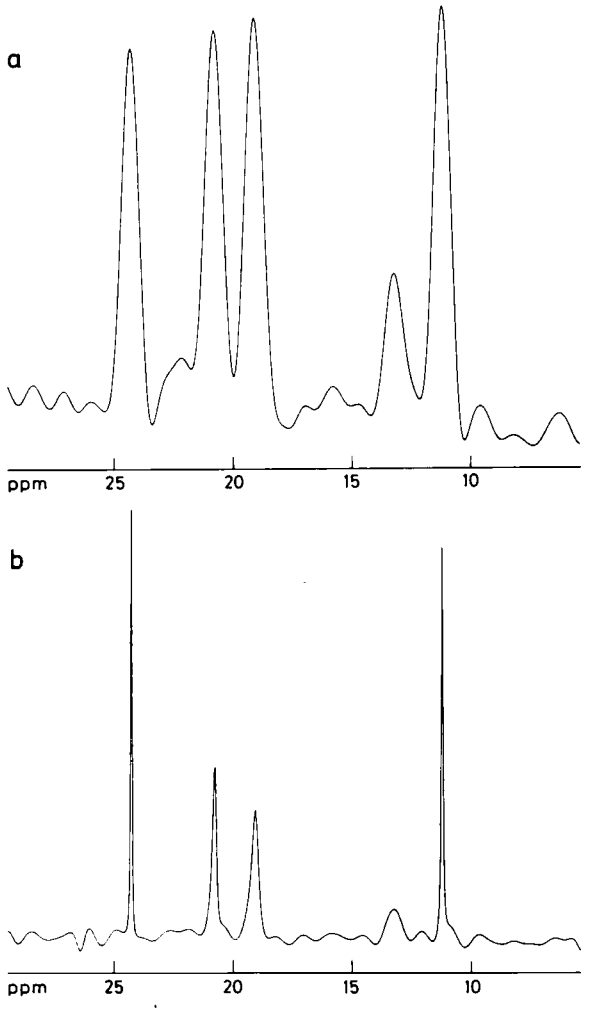


Fig. 5. Comparison of two versions of the same F_1 trace in the HSQC C–H correlated 2D spectrum in Fig. 6; (a) without and (b) with linear prediction extrapolation. The trace is indicated in the contour plot in Fig. 6, and corresponds to the trace in Fig. 3 with the same digital resolution. The FID was recorded with the pulse scheme in Fig. 1b that involves ^{13}C single-quantum coherence during t_1. The data extension and the filtering were as described for the HMQC experiment.

width and obscured the difference in line width in the two experiments (cf. Figs. 3 and 5). In the present study, the application of t_1 filtering was made unnecessary by the LP extrapolation. Also, as pointed out previously, an experimental extension of the t_1 time domain corresponding to that applied by Bax et al. (1990) would make the study here unfeasible, due to the considerable experimental time. Thus, it seems compulsory in cases like the one investigated here to apply LP extrapolation, or some other spectral estimation technique that does not suffer from the truncation artifacts that characterize the discrete Fourier transformation.

Finally, the results shown in Figs. 5b and 6 suggest that information about the dynamics and flexibility of specific sites in a protein can be obtained from the F_1 line widths of the HSQC spectrum, since these line-widths are given entirely by the transverse relaxation of the heteronuclei and

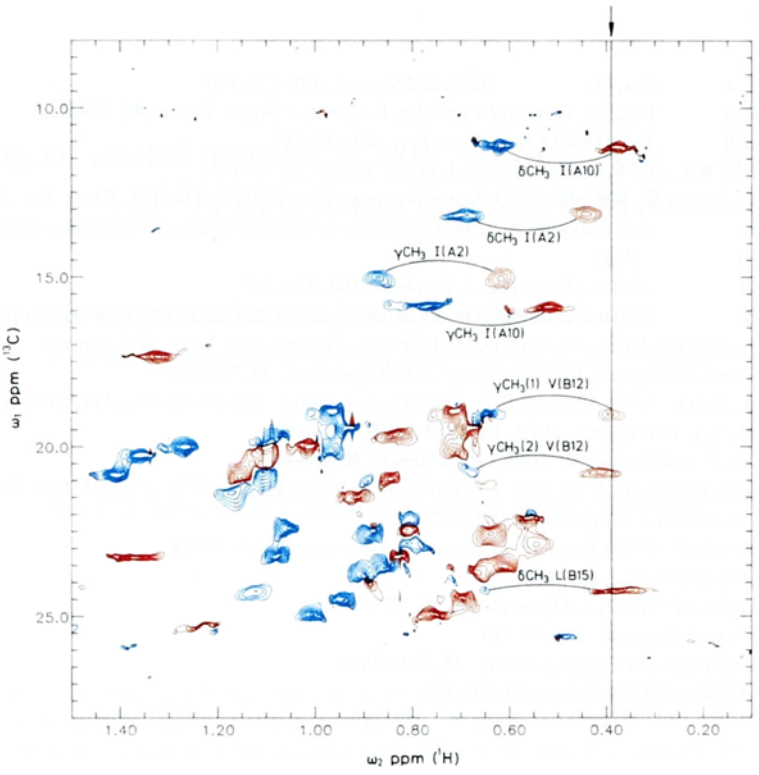


Fig. 6. Contour plot of the methyl region in the one-bond HSQC C–H correlated spectrum of B9(Asp) insulin. The vertical line at 0.39 ppm (ω_2) indicates the trace shown in Fig. 5. The FID was recorded with the pulse scheme in Fig. 1b that involves ^{13}C single-quantum coherence during t_1. Positive signals are indicated by red contours, and negative signals by blue contours. No digital filtering was used in either t_1 or t_2. For further details see the legends of Figs. 3 and 5.

the longitudinal relaxation of the proton (Bax et al., 1990). A more detailed analysis of the relaxation rates, and an assignment of the resonances of the protonated carbons in B9(Asp) insulin are in progress, and will be published elsewhere. Suffice it to note here that a significantly smaller F_1 line width is observed for the γCH_3 and the δCH_3 groups of I(A10) than for the γCH_3 and the δCH_3 groups of I(A2), which strongly suggests a relatively high flexibility of the I(A10) side chain.

In conclusion, the results presented here show that LP enhancements of one-bond C–H correlated spectra of proteins, and in particular of HSQC spectra, make detailed heteronuclear studies of proteins feasible, even with the heteronucleus in natural abundance.

The applied linear prediction program, LPEXTRAPOL, written in VAX/VMS FORTRAN and capable of handling TPPI data, can be obtained from the authors by request.

ACKNOWLEDGEMENTS

This work was supported by the Danish Technical Research Council, J. Nos. 16-3922.H and 16-4679.H, the Ministry of Industry, J. No. 85886, and Julie Damm's Studiefond. We also thank Novo Nordisk for providing the B9(Asp) insulin.

REFERENCES

Bax, A., Sparks, S.W. and Torchia, D.A. (1989) *Methods Enzymol.*, **176**, 134-150.

Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. (1990) *J. Magn. Reson.*, **86**, 304-318.

Bodenhausen, G. and Reuben, D.J. (1980) *Chem. Phys. Lett.*, **69**, 185-189.

Bodenhausen, G., Vold, R.L. and Vold, R.R. (1980) *J. Magn. Reson.*, **37**, 93-106.

Drobny, G., Pines, A., Sinton, S., Weitekamp, D.P. and Wemmer, D. (1979) *Faraday Div. Chem. Soc. Symp.*, **13**, 49-55.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon, Oxford, pp. 25-32.

Gao, Y., Boyd, J. and Williams, R.J.P. (1990) *Eur. J. Biochem.*, **194**, 355–365.

Gesmar, H. and Led, J.J. (1991) *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy* (Ed. Hoch, J.C.) NATO ASI series A, Plenum, New York, NY, in press.

Gesmar, H., Led, J.J. and Abildgaard, F. (1990) *Progr. NMR Spectrosc.*, **22**, 255-288.

Kay, L.E., Brooks, B., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *J. Am. Chem. Soc.*, **111**, 5488-5490.

Kay, S.M. and Marple, S.L. (1981) *Proc. IEEE*, **69**, 1380-1419.

Kessler, H., Schmieder, P. and Bermel, W. (1990) *Biopolymers*, **30**, 465-475.

Kristensen, S.M., Munk Jørgensen, A.M., Led, J.J., Balschmidt, P. and Hansen, F.B. (1991) *J. Mol. Biol.*, **218**, 221–231.

Marion, D. and Bax, A. (1989) *J. Magn. Reson.*, **83**, 205-211.

Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967-974.

Müller, L. (1979) *J. Am. Chem. Soc.*, **101**, 4481-4484.

Tirendi, C.F. and Martin, J.F. (1989) *J. Magn. Reson.*, **81**, 577-585.

Wagner, G. (1989) *Methods Enzymol.*, **176**, 93-113.

Wagner, G. and Brühwiler, D. (1986) *Biochemistry*, **25**, 5839-5843.

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