JOURNAL OF MAGNETIC RESONANCE **89**, 406–409 (1990)

Elimination of Baseline Artifacts in NMR Spectra by Oversampling

GERHARD WIDER

Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, CH-8093 Zurich, Switzerland

Received April 18, 1990

Protein 3D structure determination by NMR spectroscopy in solution depends primarily on the collection of a large number of uniquely assigned intramolecular distance constraints (*1, 2*). The ability to extract this information from NMR spectra, and in particular from multidimensional spectra, may be limited by artifacts from instrumental instabilities (e.g., t_1 noise) and nonideal electronics (e.g., finite pulse length, filter characteristics). One problem that affects many NMR spectra in one way or another is baseline distortion. In order to evaluate such NMR spectra, automatic baseline correction routines are often applied. Such routines can lead to extensive calculations and they are not always successful in flattening the baseline. This Communication describes an alternative method which corrects the time-domain data to obtain a flat baseline in the resulting spectra.

There are two major sources leading to baseline distortions: phase correction of the resonance lines in the spectra and wrong intensities of the first few data points in the FID. Linear phase correction leads to baseline distortions (*3*) and should therefore be avoided. Depending on the experimental setup, even zeroth-order phase correction may distort the baseline if either single channel detection or quadrature channel detection with sequential sampling is used (*4*). The distorted data points at the beginning of a FID reflect the transient response of the spectrometer which is dominated by the characteristics of the audiofrequency filters used to prevent aliasing noise about the Nyquist frequency (*5*). Correcting the resulting errors by baseline correction in the frequency domain has the disadvantage that the errors in only a few data points in the FID are spread over the whole spectrum of generally thousands of data points. Methods to correct the time-domain data seem therefore to be more suitable. The first few data points could be calculated by some extrapolation method from later, undistorted data points. Thus, for example, linear prediction methods were applied to reconstruct the beginning of the FIDs of 2D NMR measurements (*6*). Another method makes use of the filter characteristics to time properly the data acquisition (*5*). It would also be possible to replace the narrowband analog audio filters by digital filtering, a solution which is not yet available on commercial spectrometers.

Here an alternative method, based on oversampling (*7*), is proposed. Oversampling allows one to open the audiofrequency filters and thereby reduce baseline distortions. Summing blockwise over n data points in the oversampled FID results in a "normal" size FID, where n stands for the ratio of the "normal" dwell time to the oversampling dwell time. If the oversampling ratio n is chosen large enough to guarantee a high-

0022-2364/90 $3.00

Copyright © 1990 by Academic Press, Inc.
All rights of reproduction in any form reserved.

resolution digital representation of the oversampled FID, the distorted first few data points can be reconstructed by simple calculation routines such as a low-degree polynomial extrapolation. Care must be taken to set the delay between the excitation pulse and the first acquired data point correctly to avoid the need for first-order phase correction. The best results were obtained when only those data points in the oversampled FID were distorted which contribute to the first point in the reduced FID. In other words, the filter response function should reach its steady-state value in half a normal dwell time. In the corresponding spectrum this leads to a constant offset of the baseline, which can easily be corrected if the baseline is flat. In these circumstances a correction of the first few data points in the oversampled FID is usually not necessary, but can be replaced by an offset correction after Fourier transformation of the reduced FID. For this procedure a signal acquired with simultaneous sampling of the two quadrature channels is better suited because there is half a normal dwell time available for transient spectrometer responses, rather than only one-quarter with sequential sampling. An additional benefit of oversampling lies in the possibility of increasing the dynamic range of the measurement (*7*), which should, for example, improve the spectra of dilute samples in H_2O.

In order to optimize data storage, oversampled FIDs are usually reduced to normal size FIDs before being stored. An optimized digital filtering procedure could be applied. Due to the necessary substantial calculations a special processor is usually required since there is only limited time available for the conversion during data acquisition. The proposed simplified reduction algorithm based exclusively on additions is fast and allows one to do the reduction on the spectrometer computer during the acquisition of the next evolution time value in multidimensional NMR measurements. This simple reduction scheme leads to an attenuation of individual resonance frequencies, which increases toward the Nyquist frequency. When integrating spectra, this attenuation must be taken into account. For a cosine wave at frequency ω, the attenuation factor is $\sin(\omega \mathrm{DW}/2)/[n \sin(\omega \mathrm{DW}/(2n))]$, where n stands for the oversampling ratio and DW for the normal dwell time. This formula holds strictly only for even numbers of n, but is a very good approximation for odd n larger than 5. For white random noise, successive data points are not correlated and therefore the attenuation described does not occur. This results in a decreasing signal-to-noise (S/N) ratio for resonance frequencies with increasing offsets from the carrier frequency. In a standard experiment with normal dwell time one usually observes a similar frequency-dependent decrease of the S/N, since the audiofrequency filter width is generally chosen larger than the sweep width. The software on our spectrometer automatically sets the four-pole Butterworth filter 25% larger than the sweep width. With oversampling, the filter width can be chosen equal to or smaller than the sweep width, thus avoiding folding of noise. The highest resonance frequencies in a spectrum are usually at about 90% of the Nyquist frequency. At this point the ratio of the S/N values for an oversampled vs a standard spectrum is 0.94, which is a marginal cost to pay for the improved baseline.

During the evolution time of two-dimensional (or n-dimensional) NMR measurements, there are no analog filters which would distort the start of the corresponding FIDs. But usually it is not possible to measure the first data point at time zero because the initial evolution time and the encircling RF pulses have a finite length. By oversampling at the beginning of the FID in t_1, either the first point could be reconstructed

or the first point could not be measured at all with subsequent correction of the resulting offset. With both methods care must be taken that the second and later data points are collected at the correct point in time (*6*) to avoid linear phase corrections.

As an illustration a standard NOESY experiment and one using the oversampling method were recorded with an 8 m*M* sample of basic pancreatic trypsin inhibitor in H_2O at pH 3.5 and $T = 25°C$. The measurements were carried out on a Bruker AM 500 spectrometer. The experimental parameters are given in the legend to Fig. 1. During the two-dimensional Fourier transformation a cosine window was applied in both dimensions. Figure 1 presents cross sections from the standard NOESY (Figs. 1A and 1B) and from the one using the oversampling method (Figs. 1C and 1D). The row at 7.126 ppm (Fig. 1A) demonstrates the curvature and offset of the baseline often seen for rows in NOESY spectra of proteins. The curvature and offsets of the rows lead to severe distortions in the columns (Fig. 1B) and vice versa. Figures 1C and 1D show the corresponding cross sections from the NOESY experiment with oversampling in t_2 and correct timing in t_1 and t_2. No extrapolation was used. The resulting constant offset was corrected after individual Fourier transformations. The

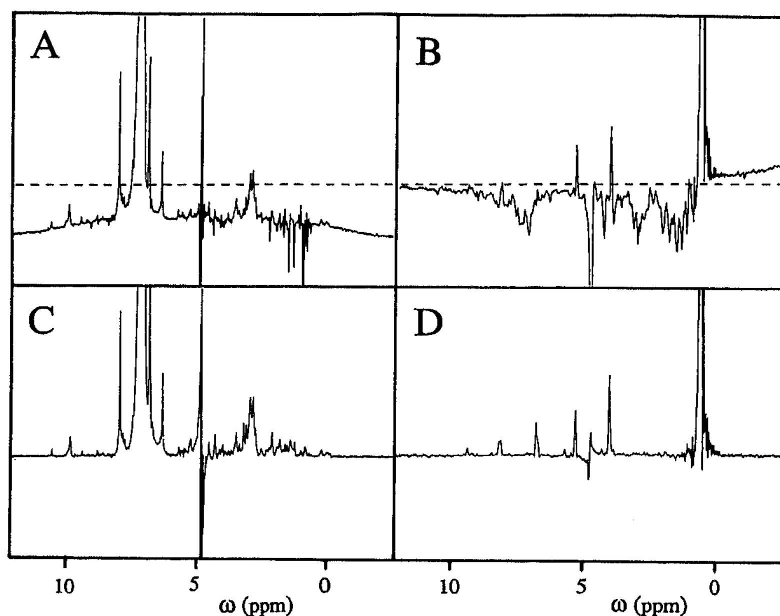

FIG. 1. Cross sections from two NOESY spectra of an 8 m*M* solution of bovine pancreatic trypsin inhibitor in 90% H_2O/10% D_2O at 25°C. The mixing time used was 60 ms. The 2D time-domain data set consisted of 440 and 2048 data points in the t_1 and t_2 dimensions, respectively. Simultaneous sampling was used in the t_2 dimension, and TPPI (*10*) in the t_1 dimension. Before Fourier transformation the t_1 domain was zero filled to 1024 points and cosine windows were applied in both dimensions. A row at 7.13 ppm (A) and a column at 0.62 ppm (B) were taken from a standard NOESY experiment. The cross sections exhibit baseline offsets and curvature as often seen in such spectra of proteins. These baseline distortions lead to severe distortions in the perpendicular cross sections. The broken lines indicate the proper position of the baseline. (C and D) The corresponding cross sections from a NOESY experiment in which the t_2 time domain was oversampled 17-fold. Each FID in t_2 was reduced to 2048 points before storage. The oversampling sweep width was 125,000 Hz, and an audiofrequency filter bandwidth of 100,000 Hz was used. The data were not extrapolated in t_2 nor in t_1 to restore the first points in the FID. In the t_1 dimension the first data point at time zero was set to zero, and the second data point was sampled after one dwell time including the finite length of the initial evolution time. The resulting constant offsets in both dimensions were corrected after the individual Fourier transformations.

same oversampled spectrum was also used to extrapolate the first data points in t_2 and t_1 by quadratic polynomial extrapolation. For this procedure the beginning of the FIDs in t_1 was also oversampled using 10 points with 3 μs increments to reconstruct the first data point at $t_1 = 0$. The same flat baseline as that in Figs. 1C and 1D was obtained.

The oversampling method requires fast sampling. As stated above, the filter response should reach a steady state after half the normal dwell time. With formula [17] of Ref. (*5*) allowing, for example, 1/1000 deviation from steady state, one can estimate a lower limit for the cutoff frequency of the filters to be chosen. For ^{1}H NMR on high-field spectrometers a value of about 40 kHz is obtained using simultaneous acquisition (SIM), or 80 kHz using sequential sampling (SEQ). In the nomenclature of our spectrometer software, the filter bandwidth (FW) equals twice the filter cutoff frequency, and the sweep width (SW) is twice the Nyquist frequency. Therefore, FW and SW must be at least 80 kHz for SIM and 160 kHz for SEQ.

Measurements of ^{1}H NMR spectra of dilute samples in H_2O using low-power presaturation of the H_2O resonance or an alternative method, e.g., with spin-lock purge pulses (*8*), may lead to an intense residual water resonance resulting in a natural baseline curvature. Instead of applying a baseline correction routine, we then prefer to filter out the water signal in the time-domain data. From the original FID with its carrier frequency on resonance with water, a FID is subtracted which is smoothed to contain only low-frequency components, i.e., the residual H_2O resonance. This method is also used by Bax and co-workers (*9*).

In conclusion, this Communication describes a method to remove from NMR spectra baseline curvature which originates from distortions in the first few data points of the FID. The method is based on oversampling and can readily be implemented on most NMR spectrometers to improve the baseline properties of homo- and heteronuclear one- and multidimensional NMR spectra.

ACKNOWLEDGMENT

I thank Professor K. Wüthrich for his critical comments and encouragement.

REFERENCES

1. K. Wüthrich, G. Wider, G. Wagner, and W. Braun, *J. Mol. Biol.* **155,** 311 (1982).
2. K. Wüthrich, "NMR of Proteins and Nucleic Acids," Wiley, New York, 1986.
3. E. O. Stejskal and J. Schaefer, *J. Magn. Reson.* **14,** 160 (1974).
4. D. Marion and A. Bax, *J. Magn. Reson.* **79,** 352 (1988).
5. D. I. Hoult, C.-N. Chen, H. Eden, and M. Eden, *J. Magn. Reson.* **51,** 110 (1983).
6. D. Marion and A. Bax, *J. Magn. Reson.* **83,** 205 (1989).
7. M. A. Delsuc and J. Y. Lallemand, *J. Magn. Reson.* **69,** 504 (1986).
8. B. Messerle, G. Wider, G. Otting, and K. Wüthrich, *J. Magn. Reson.* **85,** 608 (1990).
9. D. Marion, I. Mitsuhiko, and A. Bax, *J. Magn. Reson.* **84,** 425 (1989).
10. D. Marion and K. Wüthrich, *Biochem. Biophys. Res. Commun.* **113,** 967 (1983).