Journal of Biomolecular NMR, 2 (1992) 485–494
ESCOM

J-Bio NMR 070

Spectral processing methods for the removal of t_1 noise and solvent artifacts from NMR spectra

Nick Manoleras and Raymond S. Norton*

School of Biochemistry and Molecular Genetics, University of New South Wales, Kensington, 2033, Australia

Received 28 March 1992
Accepted 18 May 1992

Keywords: t_1 Noise suppression; Solvent suppression; 2D NMR; Proteins

SUMMARY

A data processing method is described which reduces the effects of t_1 noise artifacts and improves the presentation of 2D NMR spectral data. A t_1 noise profile is produced by measuring the average noise in each column. This profile is then used to determine weighting coefficients for a sliding weighted smoothing filter that is applied to each row, such that the amount of smoothing each point receives is proportional to both its estimated t_1 noise level and the level of t_1 noise of neighbouring points. Thus, points in the worst t_1 noise bands receive the greatest smoothing, whereas points in low-noise regions remain relatively unaffected. In addition, weighted smoothing allows points in low-noise regions to influence neighbouring points in noisy regions. This method is also effective in reducing the noise artifacts associated with the solvent resonance in spectra of biopolymers in aqueous solution. Although developed primarily to improve the quality of 2D NMR spectra of biopolymers prior to automated analysis, this approach should enhance processing of spectra of a wide range of compounds and can be used whenever noise occurs in discrete bands in one dimension of a multi-dimensional spectrum.

INTRODUCTION

Bands of t_1 noise and ridges constitute significant artifacts in frequency-domain 2D NMR spectra of biopolymers in aqueous solution, especially those recorded on older spectrometers, and can be major impediments to automated spectral analysis. The causes of some of the artifacts (Mehlkopf et al., 1984) have been minimised or eliminated by improvements in spectrometer hardware

* To whom correspondence should be addressed at: NMR Laboratory, Biomolecular Research Institute, 381 Royal Parade, Parkville 3052, Australia.

Abbreviations: ATX III, *Anemonia sulcata* neurotoxin III; NOESY, 2D homonuclear dipolar correlated spectroscopy; TOCSY, total correlation spectroscopy; ANI, average noise measurement routine; RT1, t_1 noise reduction routine; BFX, baseline correction routine.

0925-2738/$ 5.00

and software and in techniques for spectral acquisition (Turner and Patt, 1989; Simorre and Marion, 1990). Nevertheless, NMR studies of biopolymers are not only being extended to more demanding applications such as dilute solutions and/or larger molecules, but there is also a trend towards recording spectra without complete phase cycling in order to minimise acquisition time (Marion et al., 1989a). These developments suggest that spectra in which the intensities of the weakest signals are comparable to those of spectral artifacts will continue to be a feature of biopolymer NMR and, therefore, that spectral processing methods to reduce or eliminate artifacts will continue to be of value.

One of the first techniques that provided some improvement in processed 2D spectra was symmetrization (Baumann et al., 1981), but this suffered from drawbacks associated with the possible loss of real peaks or the creation of false ones, and is now rarely used. In recognition of the important link between strong diagonal peaks and t_1 noise and ridges, a spectral acquisition scheme that resulted in diagonal suppression was developed by Denk et al. (1985), but required long spectral acquisition times. Difference methods have also been proposed that involve the subtraction of an average t_1 ridge from each row (Klevit, 1985) or each column (Glaser and Kalbitzer, 1986).

A significant improvement came with the recognition that t_1 ridges arose from incorrect weighting of the first time-domain point in t_1 (Otting et al., 1986). Ridges could be suppressed by recording spectra with sine modulation in t_1 and transforming the data with sine Fourier transform, or by correcting the first time-domain point of data acquired with cosine modulation by dividing by two (Otting et al., 1986) or by means of linear prediction (Marion and Bax, 1989). The sampling delay can also be adjusted to exactly one-half of a dwell time (Bax et al., 1991).

The problem of t_1 noise remains even after implementation of these methods, and is particularly important in automated peak picking (Eccles et al., 1991; Kleywegt et al., 1991; Oschkinat et al., 1991), because a large number of false peaks from regions of t_1 noise are selected. In order to distinguish real peaks from noise, the EASY software package (Eccles et al., 1991) applies a user-defined linewidth minimum. In our experience this approach is not uniformly successful in distinguishing between t_1 noise spikes and real peaks. Using peak picking routines based on this principal and written in this laboratory, we found that there was always a compromise between losing sharp real peaks and allowing too many t_1 noise spikes to be interpreted as real peaks. In the peak picking routines described in the STELLA suite of programs (Kleywegt, 1991), the program LEARN2 is taught to distinguish between real peaks and noise. This method should work well if all the peaks in the spectrum have similar linewidths, but may encounter problems if the linewidths span a wide range or if peaks are significantly distorted by noise. The method for suppression of t_1 noise described in this paper uses an approach based on weighted smoothing of the 2D spectrum.

PRINCIPLE OF OPERATION

The measured intensity at any point of a processed 2D spectrum can be considered to consist of the real value and an error component. The error component is due to many factors, including thermal noise, t_1 noise, and offsets and ridges in both ω_1 and ω_2. Baseplane correction can reduce the effects of ridges and offsets, but a combination of thermal noise and t_1 noise remains. Whereas thermal noise is distributed evenly throughout the spectrum, t_1 noise occurs in discrete bands in ω_1 and it is on this distinction that our method is based. It assumes that each point in any given column has the same error limits, which can be estimated by measuring the average noise for that column. If this is repeated for every column, a t_1 noise profile is obtained that resembles a 1D spectrum. Having this estimate for the average noise in each column enables the application of weighted smoothing to each row, where the smoothing co-efficients are derived from the average noise estimates. This procedure selectively smooths points in areas of high noise, while leaving the points in low-noise areas relatively unaffected. Unlike traditional smoothing techniques, which weight each point equally, weighted smoothing operates on each point relative to its error component *and* the error components of its neighbouring points. This allows points in low-noise regions to influence neighbouring points in noisy regions. The sliding weighted smoothing filter, as used here, is a simple variation on the three-point averaging filter. It is repeated a number of times, allowing additional neighbouring points to influence the value of the current point.

COMPUTATIONAL IMPLEMENTATION

An accurate measurement for the average noise in each column is necessary, and various procedures for achieving this were considered. The first measured the average noise based on a small, operator-defined section of each column, but was limited by its dependence on the operator to define a region that did not contain any real peaks. Another method based the average noise estimate just on the negative peaks, but this relied on all the real peaks in the dataset being purely absorptive. The method finally chosen is based on a statistical analysis of the data. In a simple converging iterative loop, all resolvable real peaks for each column are identified and the average noise is estimated based on all remaining peaks, which are assumed to be due to noise. As there is overlap between the distribution of real peaks and the distribution of noise peaks, it is necessary to decide at what level real peaks can be distinguished from noise. Based on experience with several 2D data sets, an arbitrary value of three times the average noise was chosen. This factor of three is the default value for a user-defined parameter, η, which represents the spread of the noise around the average value.

The software routine ANI (for Average NoIse) operates on every column in the matrix, sequentially. It determines the intensity of every peak, where a peak is defined as the maximum absolute value between two zero crossings, and stores these in an array, A_j. A simple converging iterative loop is then used which progressively removes all resolvable real peaks from the array. In each pass the average noise is calculated by

$$N_i = \frac{1}{n} \sum_{j=1}^{n} A_j \tag{1}$$

where n is the number of peaks. The first pass estimate will be either greater than or equal to the real value for the average noise, depending on the density and amplitude of real peaks. The intensity of every peak, that is the data in A_j, is compared with ηN_i. Any peak having a value greater than ηN_i is identified as a real peak and removed from A_j, and has no further influence on the average noise measurement. N_i is then re-calculated using Eq. 1 based on the remaining peaks, giving a value that is less than the previous estimate. The process is repeated, with each iteration giving a lower value until the process converges on a final value for a given column that represents N_i for that column. In practice, the iterative procedure is stopped when N_i changes by less than 1%

from the value of the previous iteration. In the NOESY and TOCSY spectra described below, fewer than four iterations were required for most columns. Columns that contained no real peaks converged in the first iteration. A plot of N_i for each column in the 2D matrix produces the t_1 noise profile spectrum. The final values for N_i and η are used to derive an important quantity, ε_i, given by

$$\varepsilon_i = \eta N_i \tag{2}$$

This defines the error limits such that the real value for any point i is estimated to be within the region of the measured value plus or minus ε_i.

Once the t_1 noise profile is determined, the software routine RT1 (reduce $\mathbf{t_1}$ noise), which is a weighted smoothing routine, can be applied. It uses the t_1 noise profile in two ways: firstly to determine the smoothing coefficient for a weighted smoothing filter that is applied to each point, allowing the better resolved points to influence neighbouring points in noisy areas, and secondly to limit the amount by which any point can change to ε_i. RT1 operates on each row of a 2D matrix, one at a time. It will only change the value of points in regions where the noise level is higher than an operator-defined threshold level given by

$$L = T\,\varepsilon_{min} \tag{3}$$

where ε_{min} is the lowest ε_i for the matrix, which can be considered to be the thermal noise level, and the parameter, T, is operator defined. Only points that have a value for ε_i greater than L will have their intensities *changed* by weighted smoothing (even though all points are *used* in the smoothing procedure). Thus, if a very high value for T is chosen, only the points in the worst areas of t_1 noise will be affected by smoothing.

Prior to weighted smoothing, the absolute intensity of any point, i, to be smoothed is reduced by an amount equal to ε_i, and any point having an absolute intensity less than its expected ε_i is set to zero. This can be visualised as starting the weighted smoothing process at the minimum intensity estimated to be real (zero for any point whose intensity is less than its ε_i) and allowing the weighted smoothing filter to progressively correct each point to be smoothed under the influence of both the intensities and the ε_i values of neighbouring points. Of course, the smoothing process can only increase the absolute value of any point from this starting value as each point is maintained within a band of its initial measured value plus or minus its ε_i. This is necessary to account for the broad areas of high t_1 noise such as the band around the residual water resonance, where the influence of better resolved neighbouring points is negligible and the noise intensity is high.

The t_1 noise is then reduced along each row using a three-point smoothing filter that is weighted according to the t_1 noise profile, where the corrected value P_{ic} of the i-th point, P_i, is given by

$$P_{ic} = \frac{2\,S_iP_i + (1-S_i)\,S_{i-1}\,P_{i-1} + (1-S_i)\,S_{i+1}\,P_{i+1}}{2\,S_i + (1-S_i)\,S_{i-1} + (1-S_i)\,S_{i+1}} \tag{4}$$

This is a simple variation on the sliding three-point averaging filter, with additional terms added for weighted smoothing. The value for the smoothing coefficient, S_i, at any point, i, is given by

$$S_i = L/\varepsilon_i \tag{5}$$

where S_i is restricted to a maximum of 1.0. From Eq. 5, it is evident that if ε_i for point, P_i, is less than L, then S_i will be set equal to 1.0 and Eq. 4 will reduce to $P_{ic} = P_i$, i.e. such points will not be changed by weighted smoothing. Also from Eq. 4, it is evident that points in high noise areas will be influenced by neighbouring points in lower noise areas. The weighted smoothing process is repeated for a number of user-defined iterations which increases the size of the weighted filter and allows a greater number of neighbouring points to influence each point that is smoothed. Finally, during weighted smoothing, a restriction is placed on each point such that its final intensity will not be varied by more than its ε_i from its initial measured intensity.

The current version of RT1 has three user-defined parameters. Of these, only the threshold parameter, T, needs to be modified in most cases. A low value is best suited to preparation of datasets for peak picking and automated sequential assignment. Varying the second user-defined parameter, η, from the default value of three may be necessary in situations where the range of the t_1 noise in any column is high, leading to inadequate t_1 noise suppression and persistence of the worst t_1 noise spikes. The third parameter is the number of smoothing iterations, which has a default value of 10.

A major requirement for successful t_1 noise suppression using this method is a flat baseplane; ideally one where all noise is distributed evenly above and below the zero level. In the present work this was achieved using BFX (for Baseline FiX), a baseline correction routine developed in our laboratory which is similar in some respects to the baseline correction methods described by Pearson (1977) and Dietrich et al. (1991). This will be described elsewhere.

RESULTS AND DISCUSSION

The t_1 noise suppression routines have been tested using 2D data on ATX III, a 27-residue polypeptide toxin isolated from the sea anemone *Anemonia sulcata* (Beress et al., 1975). The software modules were incorporated in the program FELIX (version 1.1B) from Hare Research Inc. running on Iris 4D/20 and 4D/70 workstations. All code was written in FORTRAN 77.

The t_1 noise profile obtained for a TOCSY spectrum of ATX III in H_2O is shown in Fig. 1B. As expected, the worst regions correspond to the large, sharp diagonal peaks and overall the t_1 noise profile resembles the matrix diagonal (Fig. 1A). RT1 uses the t_1 noise profile to apply weighted smoothing relative to the threshold parameter, T. Figure 2 illustrates the effect of using different values for T on a row (Fig. 2A) from the TOCSY spectrum described in Fig. 1. In Fig. 2B, where a value of 50 was used, only points in the worst t_1 noise regions were affected, most points remaining unchanged. The effect of setting T equal to 10 is shown in Fig. 2C. There is significant improvement without loss of fine structure. The result of taking weighted smoothing to the extreme by setting T to 1 is shown in Fig. 2D.

Figure 3A shows the TOCSY spectrum of ATX III in H_2O, to which BFX has been applied in two steps as described above. Figure 3B is the same spectrum as shown in Fig. 3A, plotted at the same level, but after the application of ANI and RT1. The t_1 noise artifacts, including the noise associated with the water resonance, have been eliminated almost completely. Peak picking from the spectrum in Fig. 3B, whether manual or automated, is significantly simplified. Furthermore, suppression of the residual resonance has been achieved without any significant suppression of

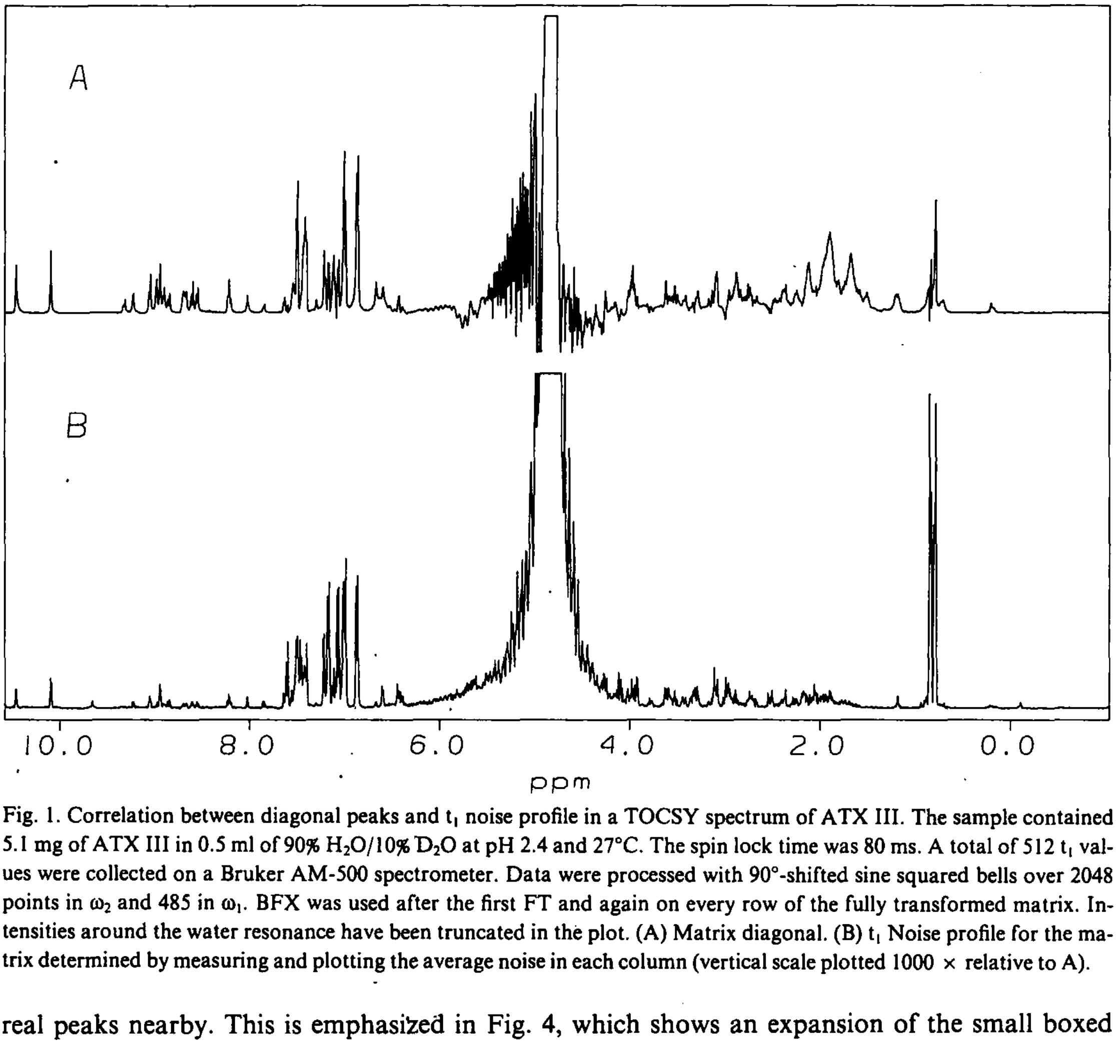

Fig. 1. Correlation between diagonal peaks and t_1 noise profile in a TOCSY spectrum of ATX III. The sample contained 5.1 mg of ATX III in 0.5 ml of 90% H_2O/10% D_2O at pH 2.4 and 27°C. The spin lock time was 80 ms. A total of 512 t_1 values were collected on a Bruker AM-500 spectrometer. Data were processed with 90°-shifted sine squared bells over 2048 points in ω_2 and 485 in ω_1. BFX was used after the first FT and again on every row of the fully transformed matrix. Intensities around the water resonance have been truncated in the plot. (A) Matrix diagonal. (B) t_1 Noise profile for the matrix determined by measuring and plotting the average noise in each column (vertical scale plotted 1000 × relative to A).

real peaks nearby. This is emphasized in Fig. 4, which shows an expansion of the small boxed region close to the water resonance in Fig. 3. The only cross peaks expected in this region are the NH to $C^{\alpha}H$ peaks of Ser^2 and Glu^{20} (Norton, Cross, Braach-Maksvytis and Beress, to be published), and indeed these are the only two peaks remaining after the application of t_1 noise suppression. This is in contrast to some other post-acquisition water suppression methods (Kuroda et al., 1989; Marion et al., 1989b), in which peaks in the vicinity of the water resonance are eliminated by the application of a filter that reduces the intensity of *all* points close to the water resonance.

Other post-acquisition methods improve the region close to the water resonance by reducing baseline distortions in spectra containing strong dispersive water signals (Tsang et al., 1990; Adler and Wagner, 1991). This type of baseline distortion has not been a significant problem in our spectra, but if there were a strong dispersive water signal, it would need to be corrected before the application of ANI and RT1. The method proposed by Adler and Wagner (1991) would be well suited for this purpose, as useful information around the water resonance is not discarded.

Our procedure for measuring the average noise, ANI, has a potential limitation in spectra with very high densities of real peaks. Indeed, any method of noise measurement would be limited in

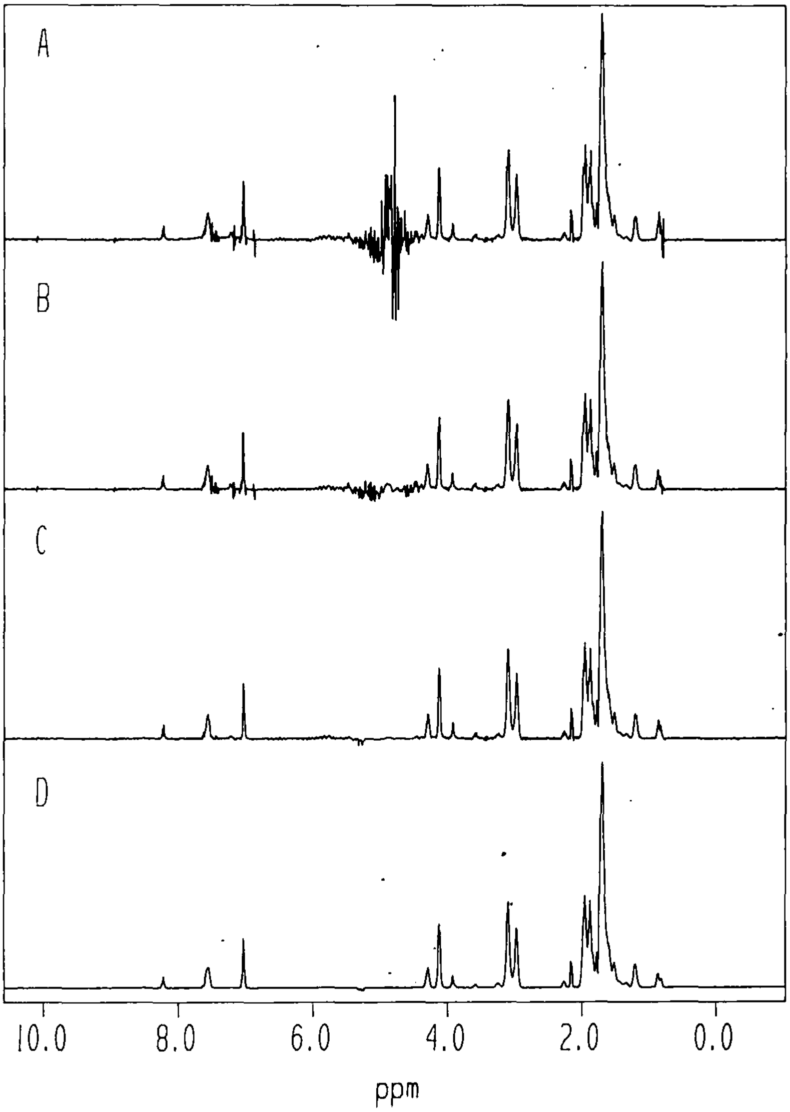

Fig. 2. Effect of using different values for the threshold parameter, T, in the t_1 noise suppression routine, RT1. Default parameters of 3 for η and 10 for the number of smoothing iterations were used. (A) Typical base corrected row (at 1.70 ppm) of the TOCSY spectrum described in Fig. 1. (B) Same as A, but after application of RT1 with T = 50. The weighted smoothing routine selectively smooths points that have a noise component greater than 50 × ε_{min} and the smoothing is weighted depending on that noise component. (C) Same as B, but with T = 10. (D) Same as B, but with T = 1.

this extreme case. However, this problem has not been observed as yet and is much less severe than first envisaged since, even in spectra having a high density of real peaks, the ratio of the number of noise peaks to real peaks is still high.

The potential of weighted smoothing to alter cross-peak volumes has been examined using a sample of cross peaks that included strong peaks and weak peaks in regions of both moderate and extreme t_1 noise of a NOESY spectrum of ATX III in H_2O. Using a high value of 30 for T, which

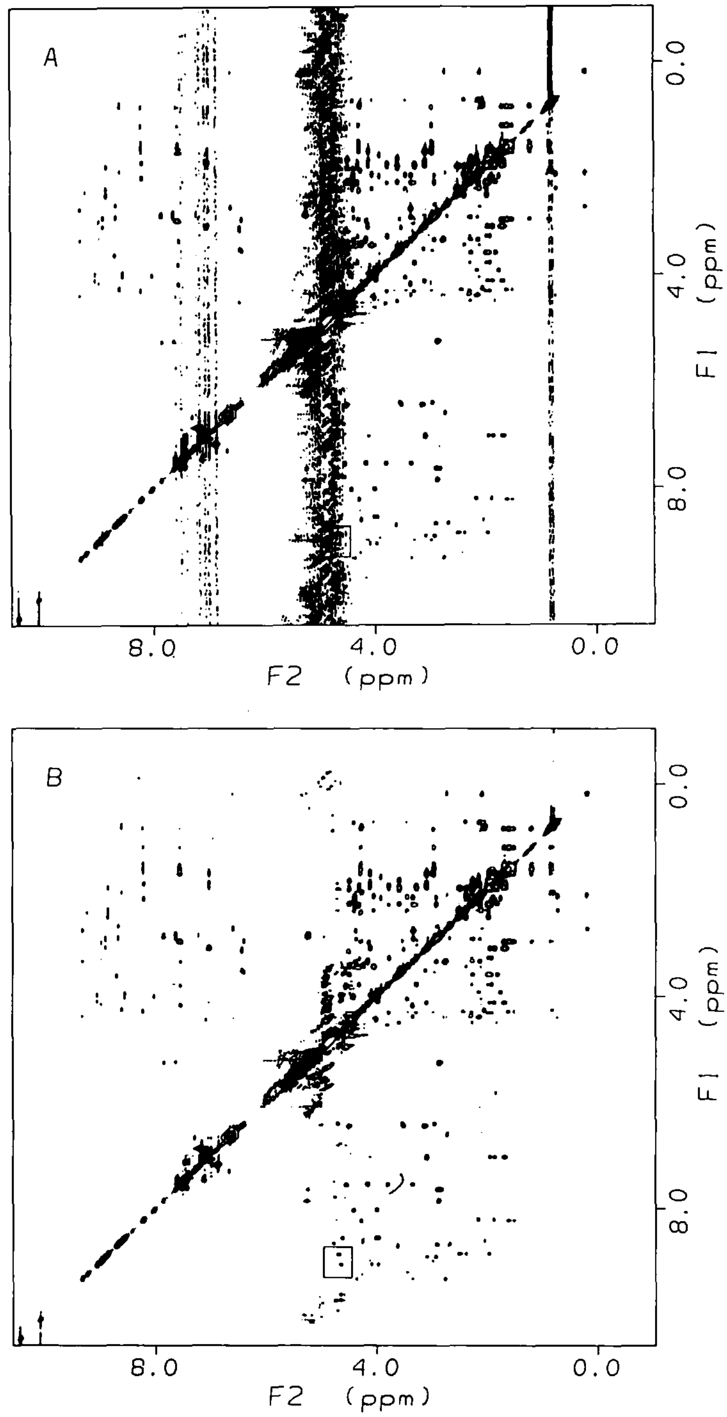

Fig. 3. Results of applying ANI and RT1 to the TOCSY spectrum of ATX III in H_2O described in Fig. 1. (A) Matrix prior to the application of ANI and RT1. (B) Same as A (and plotted at the same level), but after application of the t_1 noise suppression routines with T set at 2.

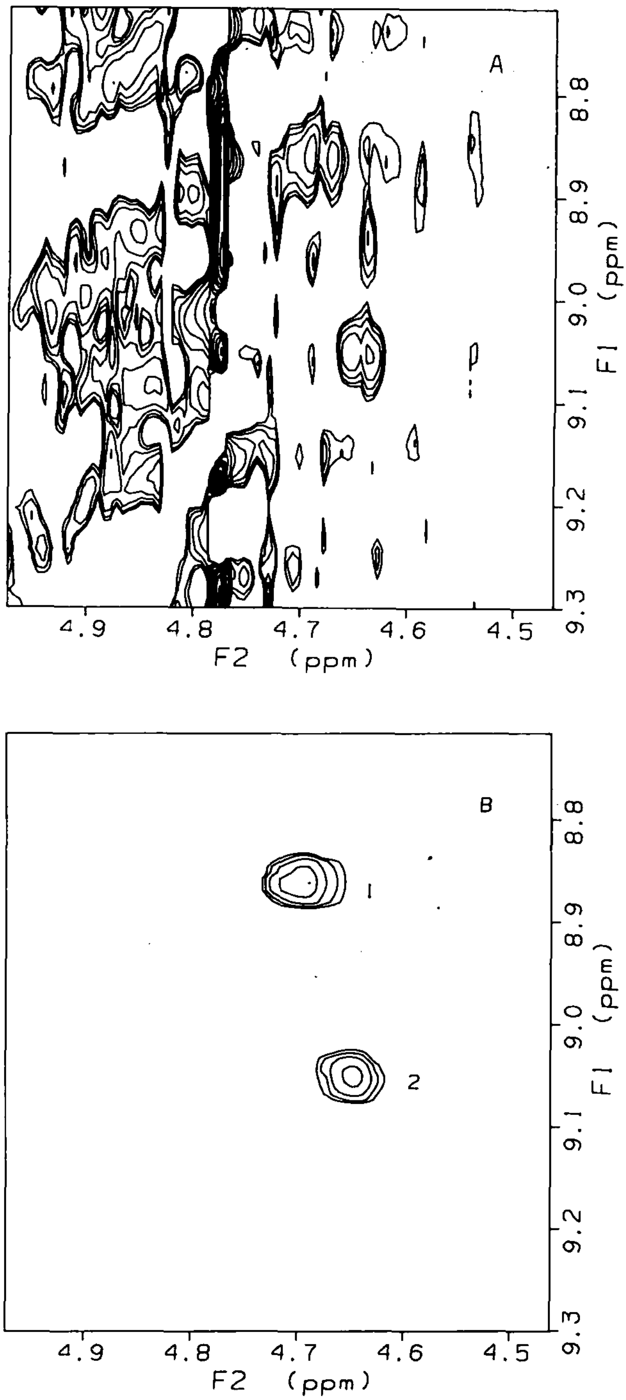

Fig. 4. Expanded views of the boxed regions of Fig. 3, illustrating the improvements in peak recognition. (A) Section of spectrum in Fig. 3A. (B) Corresponding section from Fig. 3B. Based on the level of noise in this region only two peaks could be resolved and the peaks were resolved as single peaks. These peaks were the only peaks expected in this region, with peak 1 corresponding to the NH-C$^{\alpha}$H connectivity of Glu^{20} and peak 2 to NH-C$^{\alpha}$H of Ser^{2}.

is suitable for processing NOESY spectra for volume measurements, only 20% of the cross-peak volumes were changed at all, with the most significant change being 3%. Even in the extreme case, where a value of 1 was used for the threshold parameter, most cross-peak volumes changed by less than 15%, the most significant change being 37%. In fact, the information provided by the t_1 noise profile can also be used in deriving distance constraints from NOESY cross-peak volume measurements. A volume error estimate can be determined for each peak using the t_1 noise profile, and then this can be used to select the more reliable of a symmetric pair of cross peaks by basing the distance calculation on the peak with the lower volume error estimate. In addition, upper and lower distance bounds can be loosened or tightened, based on the volume error estimates.

The execution time of the routines depends on both the quality of the spectral data and the parameters defined by the user. Applying the routines within a FELIX macro to a 2K by 2K matrix on an Iris 4D/20, base correction to every row using BFX requires 5–10 min, while using ANI to determine the t_1 noise profile requires about 2 min, and using RT1 to suppress t_1 noise requires 5–10 min. A copy of the source code for the routines described in this paper is available from the authors.

ACKNOWLEDGEMENTS

We thank Dr. Dennis Hare for providing FELIX and Dr. Laszlo Beress for the sample of ATX III.

REFERENCES

Adler, M. and Wagner, G. (1991) *J. Magn. Reson.*, **91**, 450–454.
Baumann, G., Wider, G., Ernst, R.R. and Wüthrich, K. (1981) *J. Magn. Reson.*, **44**, 402–406.
Bax, A., Ikura, M., Kay, L.E. and Zhu, G. (1991) *J. Magn. Reson.*, **91**, 174–178.
Beress, L., Beress, R. and Wunderer, G. (1975) *FEBS Lett.*, **50**, 311–314.
Denk, W., Wagner, G., Rance, M. and Wüthrich, K. (1985) *J. Magn. Reson.*, **62**, 350–355.
Dietrich, W., Rüdel, C.H. and Neumann, M. (1991) *J. Magn. Reson.*, **91**, 1–11.
Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 111–130.
Glaser, S. and Kalbitzer, H.R. (1986) *J. Magn. Reson.*, **68**, 350–354.
Klevit, R.E. (1985) *J. Magn. Reson.*, **62**, 551–555.
Kleywegt, G.J. (1991) *Computer-Assisted Assignment of 2D and 3D NMR Spectra of Proteins*, PhD thesis, University of Utrecht.
Kleywegt, G.J., Boelens, R., Cox, M., Llinas, M. and Kaptein, R. (1991) *J. Biomol. NMR*, **1**, 23–47.
Kuroda, Y., Wada, A., Yamazaki, T. and Nagayama, K. (1989) *J. Magn. Reson.*, **84**, 604–610.
Marion, D. and Bax, A. (1989) *J. Magn. Reson.*, **83**, 205–211.
Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989a) *J. Magn. Reson.*, **85**, 393–399.
Marion, D., Ikura, M. and Bax, A. (1989b) *J. Magn. Reson.*, **84**, 425–430.
Mehlkopf, A.F., Korbee, D., Tiggelman, T.A. and Freeman, R. (1984) *J. Magn. Reson.*, **58**, 315–323.
Oschkinat, H., Holak, T.A. and Cieslar, C. (1991) *Biopolymers*, **31**, 699–712.
Otting, G., Widmer, H., Wagner, G. and Wüthrich, K. (1986) *J. Magn. Reson.*, **66**, 187–193.
Pearson, G.A. (1977) *J. Magn. Reson.*, **27**, 265–272.
Simorre, J.P. and Marion, D. (1990) *J. Magn. Reson.*, **89**, 191–197.
Tsang, P., Wright, P.E. and Rance, M. (1990) *J. Magn. Reson.*, **88**, 210–215.
Turner, C.J. and Patt, S.L. (1989) *J. Magn. Reson.*, **85**, 492–505.