Journal of the Korean Magnetic Resonance Society **2011**, *15*, 157-174

DOI : 10.6564/JKMRS.2011.15.2.157

Simultaneous Spectral Resolution and Sensitivity Enhancement in MR spectrum: Maximum Likelihood Deconvolution Reconstruction

Gwang-Woo Jeong[1*], Jenny Eunice Jeong[2], and Heoung-Keun Kang[1]

[1]*Department of Radiology, Chonnam National University Hospital, Chonnam National University Medical School, Gwang-Ju 501-757, Republic of Korea*
[2]*Department of Electronics Engineering, Ewha Womans University, Seoul 120-750, Republic of Korea*
(Received November 30, 2011; revised December 07, 2011; accepted December 13, 2011)

Abstract: Although the use of apodization functions in connection with postprocessing of a 2D NMR spectrum proves improved spectral quality, there is usually a trade-off between resolution enhancement and noise suppression due to a classical "uncertainty principle." In this study, therefore, a mathematical deconvolution technique called "Maximum Likelihood Deconvolution (MLD)" was adopted to achieve the spectral resolution and sensitivity enhancement simultaneously.
The MLD technique greatly facilitates visualization and restoration of the genuine spectral information from complex 2D NMR spectra that would be problematic with the conventional apodization/FT processing. In particular, application of the MLD to the 2D-NOE spectrum would be very useful to derive the important proton connectivities, which are essential to achieve elucidating the 3D molecular structure.

keywords: apodization; deconvolution; Maximum Likelihood Deconvolution(MLD)

INTRODUCTION

Multidimensional(mD) Nuclear Magnetic Resonance(NMR) methods have revolutionized the practice of determining biomolecular structures[1-6]. Their power derives from their ability to disperse

* To whom correspondence should be addressed. E-mail : gwjeong@jnu.ac.kr

the signals from one another, especially in spectra that rely on correlations arising from through space interaction via the nuclear Overhauser effect(NOE). In other words, NOE spectra provide a powerful source for the structure determination of biological molecules in solution, because the NOE in an NMR spectrum is affected by the distances and motional properties of the nuclei in a molecule; therefore the NOE carries important information on the average molecular structure[1,2,7]. However, direct application of the two-dimensional(2D) NOE spectroscopy to macromolecular biological samples is often restricted by a lack of spectral resolution, which is due to line broadening and extensive peak overlaps accompanied by low signal-to-noise(S/N) ratio. These spectral imperfections results from "non-ideal" experimental conditions such as imperfect relaxation of sample, including truncation of the free induction decay(FID), or from poor magnetic field homogeneity, or dynamics. Such limitations are frequently serious for accurate qualitaitive and quantitative analysis of the complex and noisy spectra of proteins or nucleic acids[7-9].

Generally, there are two commonly used ways to diminish such spectral problems: (a) utilization of higher-dimensions, 3D, 4D, and 5D NMR spectroscopy, especially for ^{13}C and ^{15}N-labeled molecules[10-12] and (b) mathematical manipulation of acquired time-domain or/and frequency-domain data, *e.g.*, application of apodization functions[9,13] or spectral deconvolution techniques[14-20]. Although mD NMR spectroscopy allows an increase in resolution by spreading spectral information into the additional dimensions, there are still limitations, especially for isotopically unlabeled molecules and those which cannot be dissolved at concentrations >1mM. Also, the amount of information available

from an mD matrix does not proportionally increase with the increase of data acquisition and processing cost, because routine NMR experimental problems such as intrinsic spectroscopic inhomogeneities and non-linear instrumental response still remain. In routine Fourier transform(FT) processing of NMR signals, the FID in the time-domain is often apodized by appropriate mathematical functions such as Gaussian, exponential, or sine-bell windows in order to enhance resolution and/or reduce noise. Although the use of apodization functions provides improved spectral quality, there is usually a trade-off between resolution enhancement and noise suppression. Therefore, alternative methods are required to achieve spectral resolution and sensitivity enhancement simultaneously. Here, a spectral deconvolution technique called "maximum likelihood deconvolution (MLD)", which is a mathematical process to find the best estimate of the true value of the FID signal from the observed spectral information[19-23], is introduced to achieve both aspects. This technique has some advantages[21,22] over the other methods mentioned above: more effective spectral reconstruction than in the use of apodization functions by simultaneous resolution and sensitivity enhancement, and furthermore, simple data processing which may give enough access to spectral information to avoid isotope labeling and/or mD NMR spectroscopy.

In this paper, we utilized synthetic 2D spectra as known spectral information not only to compare qualitative and quantitative aspects between conventional FT processing and MLD, but also to evaluate quantitative strength and potential limitations of MLD, together with analysis of systematic error sources given by MLD. Subsequently, the MLD and its related protocols are applied

to a 2D NMR experiment on DNA decamer $[d(CGTACGTACG)]_2$ to aid assignment and increase the number of NOE constraints over those acquired by conventional apodization/FT methods.

EXPERIMENTAL

Simulation of 2D FID signals

Synthetic data sets used in this study were generated by a spectral simulation program[13] that produces "FlDs" offering pure Lorentzian lineshapes with various root-mean-square noise levels (RMS noise=average peak-to-peak noise/2.5). All the NMR parameters, chemical shift, linewidth, and intensity, are used as input values with phase=0. Use of synthetic data sets provides two main advantages over the experimental data sets alone. First, the actual values for all spectral information, before adding Gaussian random noise, are known. Second, it is easy to create test data sets having peculiar spectral features such as extremely poor spectral resolution with low S/N ratio. Therefore, it is very useful to use synthetic data sets as the "standard" for investigating the potential limitation of quantitative and qualitative analysis of 2D NMR spectra reconstructed by MLD reconstruction.

The synthetic dataset shown in Fig. 1a was constructed to evaluate the ability of MLD to distinguish peak information from heavily noisy conventional spectrum. This synthetic spectrum contains a single peak with 6 Hz line width and Gaussian random noise added, giving nearly ~1:1 SNR. As shown in Fig. 2a, another test spectrum was synthesized for the evaluation of accuate

representation of peak location from a noisy and complex conventional spectrum with the help of MLD reconstruction. This synthetic 2D spectrum consists of heavily overlapped 18 peaks with digital resolution of 0.195 Hz/point. The symbols designated with "+" represent the actual locations of synthetic peaks before adding pseudo-random Gaussian noise.

DNA decamer sample preparation

The test system was a complementary DNA decamer $[d(CGTACGTACG)]_2$, which was supplied by the Midland Certified Reagent Company(Midland, Texas 79701, U.S.A). A 3 mg sample was dissolved in 0.4 ml of buffer containing 0.1 M NaCI, 0.01 M PO_4^{3-}, 0.1 mM EDTA and 0.01% NaN_3 at pH 7. The exchangeable protons were replaced by drying the sample three times with glass-distilled 99. 8% D_2O (Merck), and the final sample was prepared by dissolving it in 0.4ml 99.96% D_2O (Aldrich). The sample was then transferred to a 5 mm NMR tube under argon; any dissolved O_2 was removed by blowing argon into the sample for 10 minutes and the tube was sealed with a cap and paraffin film under argon.

2D-NMR experiments

The NMR experiments were recorded on a GN-500 NMR spectrometer with a proton resonance frequency of 500 MHz at 30 ^oC. The 2D-NOE proton spectrum was collected continuously without taking the sample out of the magnetic field. The 2D spectra were recorded in the phase-sensitive

mode with quadrature detection in both dimensions. The residual water signal was suppressed by a 0.40 s DANTE pulse train, followed by a 0.05 s delay before the first 90^{o} pulse.

The 2D-NOE spectrum was recorded in the following form : 2048 real data points in t2, 400 real in t1, 16 transients for each FID and with spectral widths in both dimensions of 4405 Hz. An echo-read pulse was used to produce flat baseline. Two "dummy" cycles were used.

2D NMR data processing

The NMR data were transferred to a SUN workstation and processed using the NMRZ program (Tripos Associates Inc., St. Louis, MO). The same window functions and phase parameters were used. The t2 nd t1 ridges were corrected by scaling the intensity of the first data points on both dimensions (0.1 and 0.5, respectively) prior to FT. The cosine-bell cubic and cosine-bell squared window function were used for first (t2) and second (t1) dimension, respectively before FT, and all 2D spectra were Fourier transformed into a 4096 x 4096 matrix by zero-filling. The peak volumes were measured by the "box integration" method(Tripos Associates Inc., St. Louis, MO), where the intensities of all points in a boxed region were summed. A box is located on the peak of interest, and then the box size is adjusted to cover all significant peak intensities. The integral boxes for heavily overlapped peak regions, in particular, were determined with the assistance of MLD-derived box information, because with the conventional FT spectrum it would be difficult to decide the optimal box size and location due to the ambiguity from spectral overlaps. In each case the same box sizes

were used in conventional- and MLD-processed spectra. The volume deviation for each peak was calculated as the difference of the measured integral from the true volume. Note that "true volume" is defined as the "input volume" before adding RMS noise.

When using MLD, there are two principal parameters: a spectral line narrowing factor and the iteation number which are adjusted to find the best spectral guality, without line distortion and suppression of meaningful small signals along with the noise. In this study, the optimal parameters were determined with respect to the original line widths, degree of line overlaps and S/N ratios.

RESULTS AND DISCUSSION

Figure 1 demonstrates the superiority of MLD technique in spectral interpretation, compared with conventional FT methods, with and wiout apodization. Top views in Fig. 1 compare four sets of stacked plots, while bottom views show corresponding one dimensional projections. Figure 1a is the synthetic spectrum of a 6 Hz linewidth peak with Gaussian random noise added, which were processed by conventional FT alone without any apodization. Figure 1b is the apodized spectrum with Lorentzian-to-Gaussian weighting function having Lorentzian line sharpening(-5 Hz) and Gaussian broadening (+5 Hz) which has the center at the first point in the FID. Figure 1c isanother the apodized spectrum with Gaussian function with 5Hz line broadening, which has the center at the first point in the FID. Altough the use of Gaussian apodization provided improved spectral quality to a certain extent, the MLD reconstruction(Fig. 1d) is definitely superior in its abilities of resolution

and sensitivity enhancement by supressing the rolling baseplane and spectral noise. The most import aspect to note for this demonstration is that the conventional spectrum contains significant spectral noise even after apodization. The MLD reconstruction bestows much better visual clarity by effective noise supression and flattening of the baseplane. This evidence is clearly supported by comparing the conventional FT and its MLD one dimensional projections located in the bottom. MLD shows a very dramatic and consistent reconstruction of the single peak from the original, nearly ~1:1 S/R spectrum. Also, note how MLD processing significantly facilitates spectral interpretation. The synthesized peak position was 3966 Hz in F2 direction which is well reproduced in the MLD reconstructions.

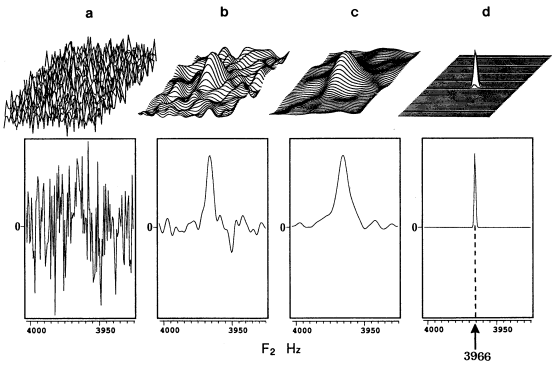

Figure 1. Synthetic spectra of a 6 Hz linewidth peak at 3966 Hz with Gaussian random noise added, which were processed by conventional FT or maximum likelihood deconvolution. The bottom views are one dimensional slices extracted from the center of the corresponding stacked plots along the F2 direction. (a) No apodization, (b) Lorentzian-to-Gaussian transformation consisting of 5 Hz Lorentzian line narrowing and 5 Hz Gaussian smoothing, (c) Gaussian transformation consisting of 5 Hz line broadening, and (d) MLD processed spectrum

The synthetic dataset shown in Fig. 2a was constructed to evaluate the ability of MLD to distingush peak information from noisy and complex conventional spectra. This data set consists of 18 resonance peaks having low S/N, where line widths ranged from 3 Hz to 14 Hz. The symbols which are designated with "+" represent the actual locations of 18 synthetic peaks and also the corresponding peak numbers are shown in the resulting MLD deconvoluted spectrum with contour mode(Fig. 2d), which was reconstructed with 6 Hz line-narrowing and 100 iterations. Peaks 1 and 2 are very broad with medium intensities, and peak 5 is a small peak located on the shoulder of the large peak group. Without prior information on the peaks under investigation it would be almost impossible to assign all of the peak locations correctly from Fig. 2a or 2b alone. Although a variety of apodization function were utilized to resolve the peak ambiguity, their application usually increases resolution at the expense of S/N. However, MLD processing provided a dramatically improved spectral quality and allowed us to find the accurate frequency information. Note that peaks 4, 5, and 9-12 were well resolved from the large peak group. It is clear that the MLD processing provides a dramatically improved spectral quality and restores the original frequencies and coupling constant information with high fidelity. This evidence is very clear by comparing 1D projections shown in Fig. 3. These projections were extracted from the three different slices along F2 direction, which were outlined by "a", "b", and "c" in Fig. 2d, respectively. The dotted lines show the 1D projections processed by conventional FT, while the solid lines represent the corresponding projections of MLD spectra. It is easy to contrast two views. It is difficult to recognize the peak centers in the

conventional FT spectra. However, MLD processing clearly restores the peak information from the conventional FT spectrum.

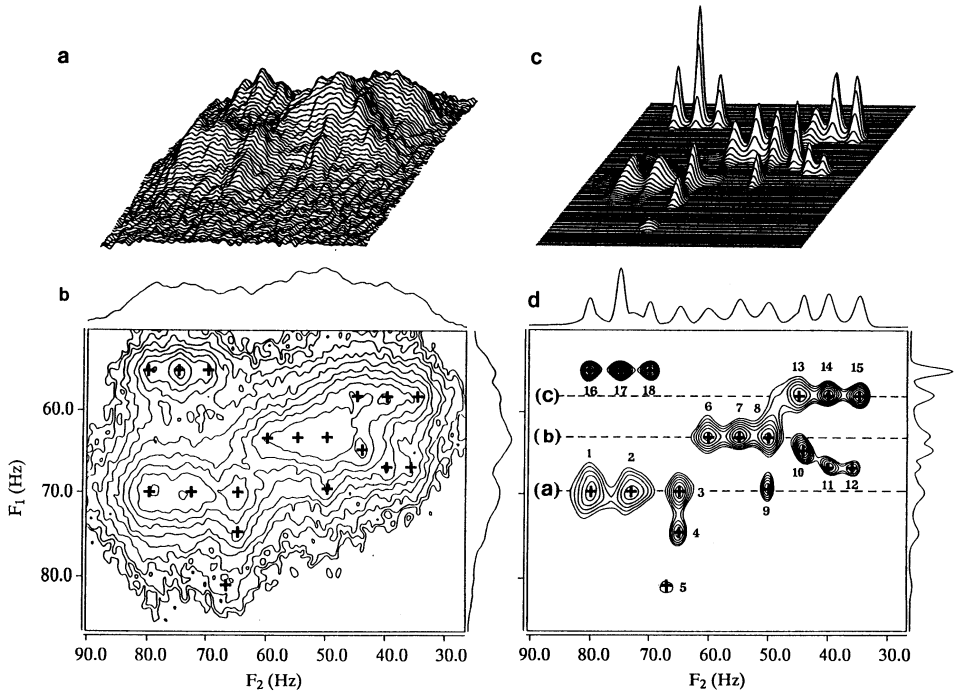

Figure 2. (a and b) Stacked and contour plots of the synthetic 2D spectra with digital resolution of 0.195 Hz/point, which was processed by conventional FT method. The symbols designated with "+" represent the actual locations of the heavily overlapping 18 synthetic peaks before adding pseudo-random Gaussian noise. (c and d) MLD reconstruction with 6Hz line-narrowing and 100 iterations. The plots are roughly 65 Hz on F2 by 40 Hz in F1, giving the peaks a somewhat broader appearance in the F1 dimension.

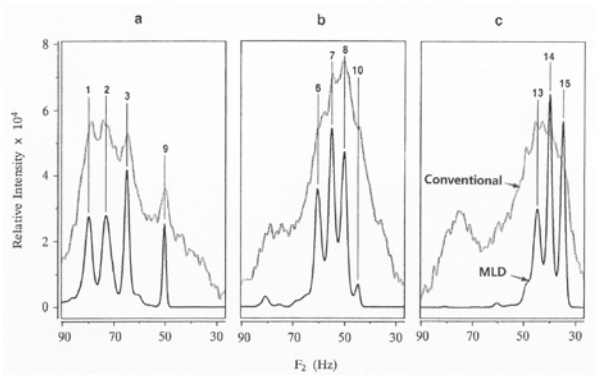

Figure 3. Comparison of the 1D projections of the spectra processed by conventional FT(dotted line) and MLD reconstruction(solid line), respectively. The projections were extracted from the three different slices along F2 direction, which were outlined by "a", "b", and "c" in Fig. 2d.

Figure 4 compares the fidelity of overall peak location in F_2 dimension of both conventional and MLD spectra shown in Figs. 2b and 2d. The deviations in peak location along y-axis were scaled with absolute values. It is very clear, at a glance, that MLD processing provides more accurate chemical shift information than does the conventional FT method. The average deviation from the actual frequencies for overall peaks was ±0.15 Hz for MLD and ±0.67 Hz for conventional processing, that is, MLD yields ~5 times more accurate chemical shift information than does conventional peak-maximum-picking for this crowded region. Note that many of the peak locations in the conventional spectrum could not have been identified without assistance from the MLD-derived peak information. The next important issue concerns peak integrals. Given the overlap in Fig. 2a and 2b, most spectroscopists would not attempt to measure integrals, abandoning information in this region for producing distance constraints, for example, from 2D NOE spectra. The point is emphasized by using

box integration where the box sizes and centers are chosen from the MLD spectrum. Although the MLD result is far from perfect, given the r^{-6} (r: a distance between two protons of interest) distance dependence on NOE integrals, the resulting distance errors would not be large[7,22,23]. This example approaches a worst-case scenario, so in most real examples the peak integrals will be defined more accurately.

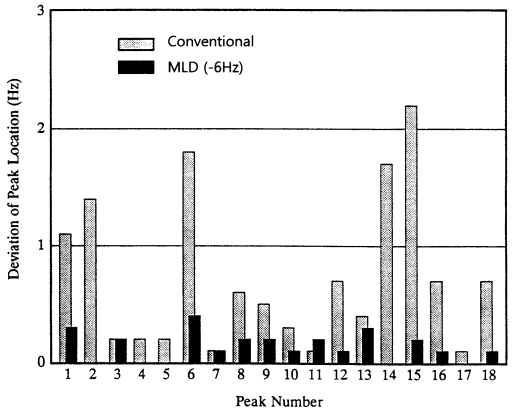

Figure 4. Bar graph comparing the fidelities in peak location between conventional and MLD spectra. The bars show deviations of peak locations from the actual input frequencies corresponding to 18 resonance peaks shown in Fig. 2. The shaded bars represent the errors for conventional spectrum and the filled bars are for MLD spectrum. Several peaks in the MLD spectra reproduce the frequencies within digital resolution(peak number 2, 4, 5, 7, 10, 12, 14, 16-18).

Figure 5 shows the 500MHz 2D-NOE spectrum(4K×4K) of the DNA decamer molecule(～3500M.W.), $[d(CGTACGTACG)]_2$, at 250ms mixing time, where diagonal peaks were suppressed. The spectrum was acquired at 30 ℃ and the HDO frequency was calibrated at 4.874 ppm.

Figure 5 is the symmetrized 2D-NOE spectrum. The diagonal symmetrization was performed with the point having the minimum absolute value but retaining the original sign. This spectrum illustrates the typical overlap problem in an NOE spectrum. Most regions of this spectrum are heavily overlapped, although the "base-to-H3'" region outlined by the zoom box are relatively simple. Before MLD processing, the spectrum was diagonally symmetrized with the function of minimum of absolute-value retaining sign. The spectral symmetrization gives a better spectral quality by partially recovering restricted F_1 resolution and spectral averaging between two portions diagonally positioned. MLD deconvolution was performed with 6 Hz line narrowing(100 iterations) on the extracted region in Fig. 5.

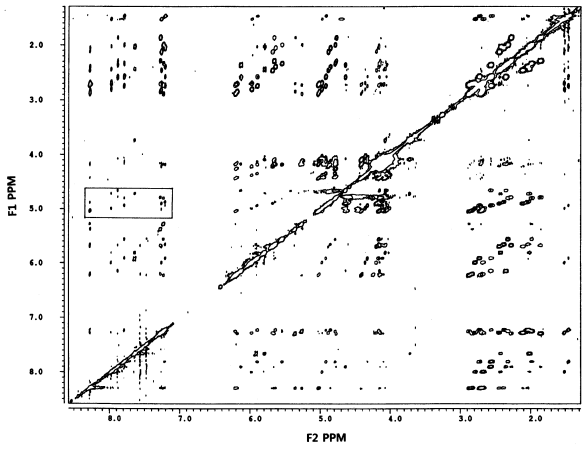

Figure 5. A symmetrized 2D-NOE spectrum(4k x 4k data size) of the DNA 10-mer molecule(MW～3500), $[d(CGTACGTACG)]_2$, with suppression of diagonal peaks, followed by diagonal symmetrization. The diagonal symmetrization was performed with the point having the minimum absolute value but retaining the original sign at 250 ms mixing time. The spectrum was acquired at 30 ^{o}C and the HDO frequency was calibrated at 4.874 ppm.

The conventionally processed and symmetrized MLD spectra are shown in Figs. 6a and 6b, respectively. The NOE walk for base-to-H3' is clearly demonstrated in the MLD version. Figure 7 shows an expansion of the circled regions of Fig. 6. Figures 6b and 7b demonstrate the resolution enhancement aspects of the MLD protocol. Note that two different NOE walks have been clearly assigned from the MLD spectrum in Fig. 7b, while the assignment of NOE walks in the conventional FT spectrum is almost impossible because of its lack of peak information with coupling patterns.

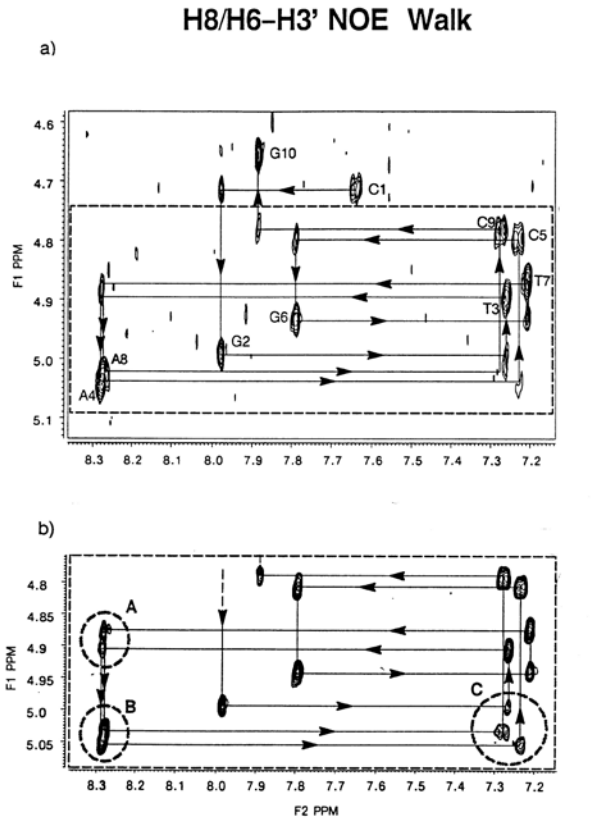

Figure 6. (a) The conventional FFT spectrum of a part of the "base-to-H3'" region that was expanded from the box in Fig. 5. The abbreviations in the figure stand for purine and pyrimidine bases: C, Cytosine; A, Adenine; G, Guanine; and T, Thymine. (b) The MLD spectrum from the same segment outlined with the dotted lines as in Fig. 6a. In the MLD version, the NOE walk for base-to-H3' is clearly demonstrated.

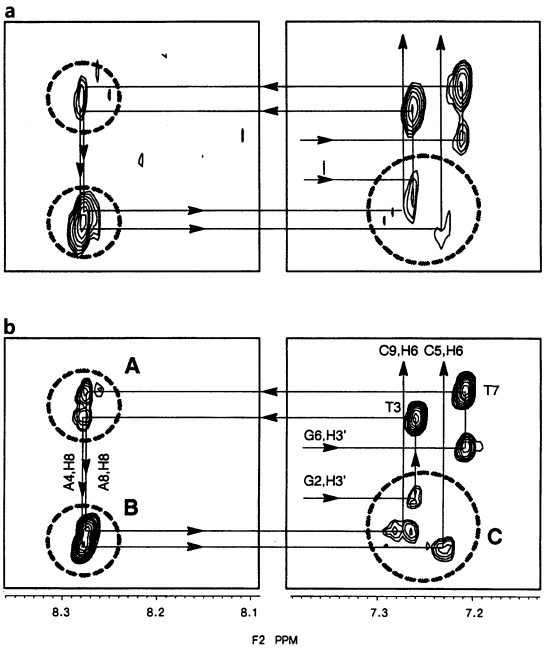

Figure 7. Comparison of the conventional FT spectrum(a) and the MLD spectrum(b). The spectral regions shown in both cases were extracted from the circled regions indicated with "A", "B", and "C" in Fig 6.

In general, quantitative errors for 2D NOE cross-peaks increase the uncertainty in distance constraints; in heavily overlapping regions these errors are often so large as to preclude use of many constraints. Now our advanced MLD-related data processing facilitates the extraction of accurate spectral information for cross-peaks from a complex, experimental 2D NOE spectrum associated with heavy overlap and low S/N. The use of MLD protocols is remarkably promising for gaining a larger

number of NOE-based distance constraints that can be used in accurate three-dimensional molecular structure determination.

However, there are drawbacks[9] due to the constrained nonlinear deconvolution nature of MLD. Below S/N of about 4, for example, MLD provides a poorer quantification of peak intensity, although it gives a much better estimate of frequency and coupling constant. Another disadvantage of the MLD is that the user is required to input two types of parameters: the spectral line width to be removed(in each dimension) and the number of iteration. The selection of the two independent values is operator-dependent. It is also desirable to extend 2D MLD to higher dimensions for effective and accurate spectral interpretation, together with inherent advantages of mD NMR experiments.

CONCLUSION

The MLD technique leads to substantial improvements in spectral resolution and S/N ratio, and thus it allows us to assign the proton resonance peaks accurately, especially from heavily overlapped multiplets. The advantages allow outstanding accuracy for retrieval of chemical shifts and coupling constants from in-phase cross-peaks, together with accurately partitioning overlapped peaks and, often, individual components from complex multiplets even in the face of a low S/N ratio. In addition, 2D-NMR methods (possibly 3D and 4D NMR) combined with deconvolution technique would be a powerful tool for further research works for accurate elucidating the 3D molecular structures

associate with developing effective medicines for chemotherapy of cancer or AIDS, finding how carcinogens cause the development of malignant cells, and exploring how gene expression is regulated.

Acknowledgment

This study was supported by a grant (CRI 09051-1) from Chonnam National University Hospital Research Institute of Clinical Medicine.

REFERENCES

1. K. Wüthrich, NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York (1986)
2. R. R. Ernst, *CHIMIA* **41**, 323, (1987)
3. A. Bax, *Annu. Rev. Biochem.* **58**, 223, (1989)
4. T. James and V. J. Basus, *Annu. Rev. Phys. Chem.* **42**, 501, (1991)
5. G. M. Clore , A. M. Gronenborn, *Science* **252(5011)**, 1390, (1991)
6. G. Wagner, V. Thanabal, B. J. Stockman, J. W. Peng, N. R. Nirmala, S. G. Hyberts, M. S. Goldberg, D. J. Detlefsen, R. T. Clubb, M. Adler, *Biopolymers* **32(4)**, 381, (1992)
7. G.W. Jeong, H. K. Kang, *Concepts Magn. Reson.* 1**8A**, 146, (2003)
8. G. C. Levy, P. N. Borer, G. W. Jeong, "Deconvolution and Quantitative Enhancement of Multidimensional Spectra," *Computer-Enhanced Analytical Spectroscopy*, edited by C. L. Wilkins, Plenum Press, New York(1993)
9. G. W. Jeong, S. Wang, P. N. Borer, G. C. Levy, *J. Magn. Reson.* **A103**, 123, (1993)
10. C. Griesinger, O. W. S∅rensen, R. R. Ernst, *J. Mag. Reson.* **84**, 14, (1989)
11. L. E. Kay, D. Marion, A. Bax, *J. Mag. Reson.* **84**, 72, (1989)
12. G. M. Clore, A. M. Gronenborn, *Progress in NMR Spectroscopy* **23**, 43, (1991)
13. R. E. Hoffman, G. C. Levy, *Progress in NMR spectroscopy* **23**, 211, (1991)

14. Y. L. Shen, H. K. Kang, T. H. Kim, T. Sundaram, H. J. Kim, G. W. Jeong, *J. Kor. Magn. Reson. Soc.* **13**, 64, (2009)

15. A. R. Mazzeo, M. A. Delsuc, A. Kumar, G. C. Levy, *J. Mag. Reson.* **81**, 512, (1989)

16. G. W. Halsey, W. E. Blass, Deconvolution with Applications in Spectroscopy, P. A. Jansson, Ed., Academic Press, New York/Orlando(1984)

17. M. A. Delsuc, F. Ni and G. C. Levy, *J. Magn. Reson.* **73**, 548, (1987)

18. D. S. Stephenson, *Progress in NMR spectroscopy* **20**, 515, (1988)

19. J. N. Kapus, Maximum Entropy Models in Science and Engineering, Wiley Eastern Limited, New Delhi, India(1989)

20. F. Ni, H. A. Scheraga, *J. Magn. Reson.* **82**, 413, (1989)

21. R. E. Hoffman, A. Kumar, K. D. Bishop, P. N. Borer and G. C. Levy, *J. Magn. Reson.* **83**, 586,(1989)

22. G. W. Jeong, H. K. Kang, *Concepts Magn. Reson.* **14**, 402, (2002)

23. G. C. Levy, S. Wang, G. W. Jeong, P. N. Borer, "Software Approaches for Determination of 3-dimensional Molecular Structures from Multidimensional NMR," in Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy, Edited by J. C. Hoch, F. M. Poulsen, and C. Redfield, Plenum Press, New York(1991)