Journal of Magnetic Resonance **146,** 89–99 (2000)
doi:10.1006/jmre.2000.2102, available online at http://www.idealibrary.com on IDEAL®

A Time-Domain Algorithm for NMR Spectral Normalization

Rocco Romano,*[,1] Maria Teresa Santini,*[,†] and Pietro Luigi Indovina*[,‡]

*Istituto Nazionale per la Fisica della Materia, Unità di Napoli, c/o Dipartimento di Scienze Fisiche, Università degli Studi di Napoli Federico II, Complesso Universitario di Monte S. Angelo, via Cinthia, 80126 Naples, Italy; †Laboratorio di Ultrastrutture, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy; and ‡Dipartimento di Scienze Fisiche, Università degli Studi di Napoli Federico II, Complesso Universitario di Monte S. Angelo, via Cinthia, 80126 Naples, Italy

Received December 20, 1999; revised March 29, 2000

Recently, a new method for quantitatively comparing NMR spectra of control and treated samples, in order to examine the possible occurring variations in cell metabolism and/or structure in response to numerous physical, chemical, and biological agents, was proposed. This method is based upon the utilization of the maximum superposition normalization algorithm (MaSNAl) operative in the frequency domain and based upon maximizing, by an opportune *sign variable measure,* **the spectral region in which control and treated spectra are superimposed. Although the frequency-domain MaSNAl algorithm was very precise in normalizing spectra, it showed some limitations in relation to the signal-to-noise ratio and to the** *degree of diversity* **of the two spectra being analyzed. In particular, it can rarely be applied to spectra with a small number of visible signals not buried in the noise such as generally** *in vivo* **spectra. In this paper, a time-domain normalization algorithm is presented. Specifically, it consists in minimizing the rank of a Hankel matrix constructed with the difference of the two free induction decay signals. The algorithm, denoted MiRaNAl (minimum rank normalization algorithm), was tested by Monte Carlo simulations as well as experimentally by comparing two samples of known contents both with the new algorithms and with an older method using a standard. Finally, the algorithm was applied to real spectra of cell samples showing how it can be used to obtain qualitative and quantitative biological information.**

Key Words: **NMR; algorithm; normalization; NMR of cells; time-domain algorithm.**

INTRODUCTION

NMR spectroscopy is increasingly used, both *in vitro* and *in vivo*, to examine variations induced by physical, chemical, or biological agents which have acted on a sample or on a patient. In these types of studies, in order to obtain relative quantitative information, a comparison between signal intensities of control and treated or exposed samples is often conducted. For instance, in *in vivo* studies, signal intensities of a patient spectrum before and after a therapeutic treatment can be compared to study the effect of the treatment; in *in vitro* studies, signal intensities of control and treated and/or exposed cell samples can be compared to examine variations in cell metabolism and/or structure.

The problem of precisely comparing signal intensities of treated and control samples is actually a very important task. For this purpose, various methods have been developed. For example, in *in vitro* studies, synthetic, such as 3-trimethylsilyl[2,2,3,3-d_4]propionate (*1, 2*), or naturally occurring, such as glucose in ^{13}C spectra (*3*), reference compounds have been used. In addition, in both *in vivo* and *in vitro* studies, enzymatic determination of the concentration of a metabolite present in the samples as well as the assumption that the concentration of a particular metabolite (e.g., the alanine methyl doublet or the creatine singlet) does not change has also been utilized (*4, 5*). In particular, a common method of normalization of spectra is to divide each spectral intensity by either the sum of intensities or the square root of the sum squares of intensities of metabolites whose concentrations are supposed not to change. All these methods are not related directly to the overall intrinsic properties of the samples, but rather to the addition of external substances of known concentration, to the indirect measurement of internal substances, or to an assumption about the concentrations of particular substances.

Recently, a new method for quantitatively comparing the spectra of control and treated samples was proposed (*6*). It was first used in order to study adhesion between cells and biomaterials, in particular in order to study the interactions between K562 leukemic cells and polylysine (*7*).

Since concentration differences result in proportional variations of spectral intensities, nonproportional changes can most likely be attributed to the effects of a particular agent. The proportional variations are described by a normalization factor R, which must be calculated in order to compare control and treated signal intensities. For this purpose, the frequency-domain maximum superposition normalization algorithm (MaSNAl) (*6*) was proposed. It operates in the frequency domain and takes into consideration all the cell metabolites present in the sample. It consists in maximizing, by an opportune *sign variable measure* (*8, 9*), the spectral region in which spectral lines are proportional.

[1] To whom correspondence should be addressed. Fax: (+39) 081676346. E-mail: rocco.romano@na.infn.it.

1090-7807/00 $35.00

It was shown that the MaSNAl algorithm can be applied to spectra with a high signal-to-noise ratio (SNR $>$ 70) and when there are few changes induced by the agent, i.e., when the spectral lines with changes are few with respect to the total number of spectral lines. In particular, the percent ratio of nonsuperimposed points and the total number of spectral points, denoted the *degree of diversity* of two spectra, demonstrated that the MaSNAl algorithm was able to calculate the normalization factor with a bias of less than 1% only for a degree of diversity of less than 50%.

In this paper, a time-domain algorithm, denoted as MiRaNAl (minimum rank normalization algorithm), which is able to calculate the normalization factor R is presented. It is based on the minimization of the rank, obtained by singular value decomposition (SVD) (*10–13*), of an opportune Hankel matrix (*14*). The algorithm was tested by Monte Carlo simulations which demonstrate its ability to determine the normalization factor with very low bias, both in spectra with a low signal-to-noise ratio and in cases in which great changes are induced by the treatment (degree of diversity $>$ 50%). In particular, simulations were conducted for both MaSNAl and MiRaNAl algorithms to compare their ranges of applicability. In order to test the validity of this algorithm experimentally, two samples of known contents were compared. A traditional method based on the use of a standard and the new MiRaNAl algorithm were used.

Finally, the MiRaNAl algorithm was applied to real spectra of cell samples, demonstrating how it can be used to obtain qualitative and quantitative biological information.

RESULTS AND DISCUSSION

The Algorithm

A time-domain NMR free induction decay (FID) experiment can be modeled as a sum of complex exponentially decaying sinusoids,

$$x_n = \sum_{s=1}^{s=S} A_s \exp[i(\phi_0 + \phi_s)] \exp[(-\alpha_s + i2\pi\nu_s)t_n] + e_n, \qquad [1]$$

where S is the number of sinusoids, and A_s, α_s, ν_s, and ϕ_s ($s = 1, 2, \ldots, S$) are the amplitude, damping factor, frequency (in Hz), and phase, respectively, of the sth sinusoid. The value of ϕ_0 is the zero-order phase and e_n is complex white Gaussian noise. The number of complex data points is N, $n = 0, 1, \ldots, N - 1$, and the discretely sampled time steps are $t_n = (n + \eta)\Delta t$,with $t_0 = \eta\Delta t$ the begin time, or dead time of the spectrometer (*15, 16*). The value of ϕ_s, required only under particular experimental conditions, can usually be set equal to zero (*16*) and, in what follows, it will be supposed that this is the situation for the present experimental conditions.

The $L \times M$ Hankel matrix

$$X = \begin{bmatrix} x_0 & x_1 & \cdots & x_{M-1} \\ x_1 & x_2 & \cdots & x_M \\ \vdots & \vdots & \vdots & \vdots \\ x_{L-1} & x_L & \cdots & x_{N-1} \end{bmatrix}, \qquad [2]$$

which consists of N uniformly sampled data points x_n, $n = 0, 1, \ldots, N - 1$, with L and M chosen greater than S and subjected to the constraint $N = L + M - 1$, is considered. If noise is present and if the SNR is not too low, the rank of the Hankel matrix in Eq. [2] can be approximated by S, that is, by the number of harmonic components of the FID, and it can be obtained by SVD (*10–13, 17*). The singular value decomposition theorem states that if X is an arbitrary $L \times M$ complex valued matrix, then there exist unitary matrices $U(L \times L)$, $V(M \times M)$ and p ordered real numbers ($p = \min(L, M)$) $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p > 0$, such that

$$X = U\Sigma V^{\dagger},$$

where $\Sigma(L \times M)$ is such that $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$ and the † denotes Hermitian conjugation. The p numbers are the so-called singular values of the X matrix.

If the X matrix has rank equal to Λ, only its first Λ singular values are greater than zero; that is, for a Λ rank X matrix one has

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_\Lambda > 0$$
$$\sigma_{\Lambda+1} = \sigma_{\Lambda+2} = \ldots = \sigma_p = 0. \qquad [3]$$

An X Hankel data matrix of a noiseless FID comprising Λ complex decaying sinusoids has rank equal to Λ, because its elements are points of the FID and then they are obtained by a linear combination of Λ independent signal components. Each X Hankel matrix row (column) is a linear combination of the same Λ independent signal components and then the X Hankel matrix can have only Λ independent rows (columns). The rank of the X Hankel matrix is Λ. As a consequence, the X Hankel matrix has only Λ singular values different from zero. In Fig. 1a, the singular values of the X Hankel matrix of a simulated noiseless FID, containing 10 exponentially decaying sinusoids, are plotted. As can be seen, only the first 10 singular values are different from zero, while the others are all zero.

If the FID is affected by noise, its X Hankel matrix becomes a full rank matrix because the noise destroys the linear dependence of the rows (columns): noise can be considered the combination of infinite independent signals. However, if the SNR is not too low, that is, if the signal amplitudes can be considered greater than the noise amplitude, the signal-related singular values are very much greater than the noise-related singular values and the rank of the X Hankel matrix still can be

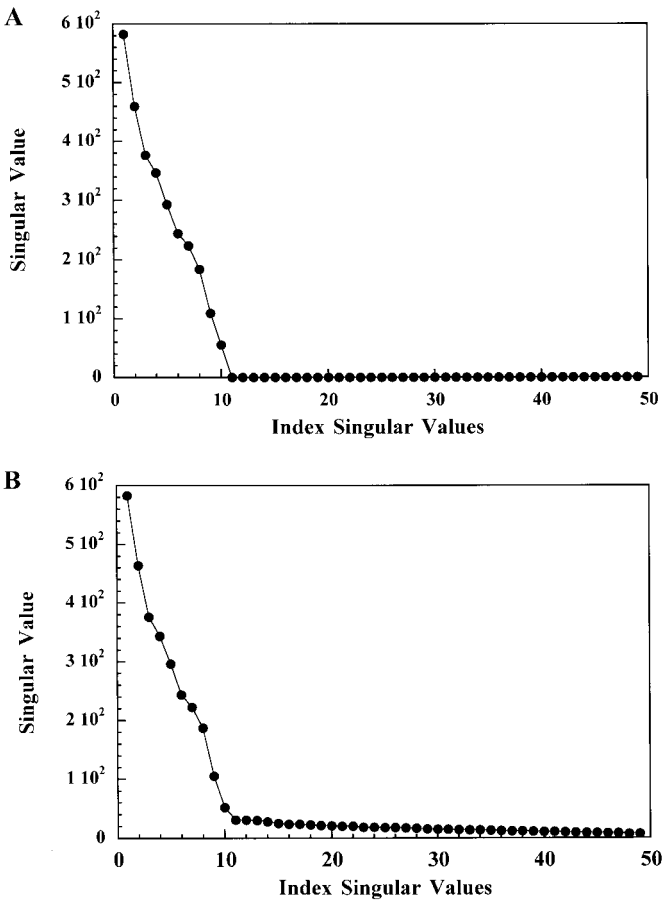

FIG. 1. Plot of the singular values of a simulated FID containing 10 exponentially decaying sinusoids: (a) noiseless FID; (b) SNR = 108.

approximated by Λ. In particular, in the ordered singular values there will be a discontinuity between signal-related and noise-related singular values. In Fig. 1b the singular values of the X Hankel matrix of the above-simulated FID containing 10 exponentially decaying sinusoids, with the addition of Gaussian noise (SNR = 108), are plotted. As can be seen, all singular values are not exactly zero, but at the 10th singular value there is a discontinuity. The rank of the Hankel matrix can be approximated by 10.

At this point, two FIDs $X_1(t)$ and $X_2(t)$ will be considered relative, respectively, to a control sample and to a treated one. It will be supposed that the treatment does not affect some unknown signals, which are proportional due to changes only in concentration. These proportional variations are described by the normalization factor R, and the two FIDs can be written

$$x_{n1} = \exp(i\phi_{01})\Big[\sum_{s=1}^{s=P} P_{s1}\exp[(-\alpha_{s1} + i2\pi\nu_s)t_n] + \sum_{s=P+1}^{s=S1} A_{s1}\exp[(-\beta_{s1} + i2\pi\nu_{s1})t_{n1}]\Big] + e_{n1} \quad [4]$$

and

$$x_{n2} = \exp(i\phi_{02})\Big[\sum_{s=1}^{s=P} P_{s2}\exp[(-\alpha_{s2} + i2\pi\nu_s)t_n] + \sum_{s=P+1}^{s=S2} A_{s2}\exp[(-\beta_{s2} + i2\pi\nu_{s2})t_{n2}]\Big] + e_{n2}, \quad [5]$$

where $n = 0, 1, \ldots, N-1$, ϕ_{01} is the zero-order phase, e_{n1} is the complex Gaussian noise and $t_{n1} = (n + \eta_1)\Delta t$, with $t_{01} = \eta_1\Delta t$ the begin time of FID 1; P_{s1}, α_{s1}, and ν_s ($s = 1, 2, \ldots, P$) are, respectively, the amplitude, damping factor, and frequency (in Hz) of the sth proportional sinusoid of FID 1; and A_{s1}, β_{s1}, and ν_{s1} ($s = P+1, P+2, \ldots, S1$) are, respectively, the amplitude, damping factor, and frequency (in Hz) of the sth nonproportional sinusoid of FID 1. The symbols ϕ_{02}, e_{n2}, $t_{n2} = (n + \eta_2)\Delta t$, $t_{02} = \eta_2\Delta t$, P_{s2}, α_{s2}, ν_s, A_{s2}, β_{s2}, and ν_{s2} have the same meaning for FID 2. In this hypothesis, $P_{s1} = R * P_{s2}$, and $\alpha_{s1} = \alpha_{s2}$ for $s = 1, 2, \ldots, P$, while, in general and due to the effects of the treatment, for $s > P$, $A_{s1} \neq R * A_{s2}$ and β_{s1}, β_{s2}, ν_{s1}, and ν_{s2} could be different.

If one assumes that $\eta_1 \simeq \eta_2 \simeq 0$ (*18*), from Eqs. [4] and [5] it follows that the zero-order phases ϕ_{01} and ϕ_{02} of the two FIDs can be approximated by the phases of their first point x_{01} and x_{02}, respectively; in particular, $\phi_{01} \simeq \arctan(x_{01})$ and $\phi_{02} \simeq \arctan(x_{02})$. To correct the FIDs in their zero-order phase, it is then sufficient to multiply each point of the FIDs by $\exp^{-i\phi_{01}} \simeq \exp^{-i\arctan(x_{01})}$ and $\exp^{-i\phi_{02}} \simeq \exp^{-i\arctan(x_{02})}$, respectively; the FIDs, corrected in their zero-order phase, can then be described as

$$\hat{x}_{n1} = \exp[-i\arctan(x_{01})] * x_{n1} \quad [6]$$

$$\hat{x}_{n2} = \exp[-i\arctan(x_{02})] * x_{n2}, \quad [7]$$

where $n = 0, 1, \ldots, N-1$.

Denoted by

$$\delta\hat{x}_n = \hat{x}_{n1} - K * \hat{x}_{n2}, \quad [8]$$

where K is a test value of the normalization factor R, the Hankel matrix

$$\Delta X(K) = \begin{bmatrix} \delta\hat{x}_0 & \delta\hat{x}_1 & \cdots & \delta\hat{x}_{M-1} \\ \delta\hat{x}_1 & \delta\hat{x}_2 & \cdots & \delta\hat{x}_M \\ \vdots & \vdots & \vdots & \vdots \\ \delta\hat{x}_{L-1} & \delta\hat{x}_L & \cdots & \delta\hat{x}_{N-1} \end{bmatrix} \quad [9]$$

with L and M chosen greater than $S_1 + S_2$ and subjected to the constraint $N = L + M - 1$, will be considered. For $K \neq R$,

TABLE 1
FID 1 and FID 2 Parameters

Amplitude 1	Amplitude 2	Damping 1	Damping 2	Frequency 1	Frequency 2
17.	34.	369.	369.	−3500.	−3500.
10.	20.	327.	327.	−3450.	−3450.
36.	72.	913.	913.	−3150.	−3150.
15.	30.	651.	651.	−3000.	−3000.
20.	40.	236.	236.	−2900.	−2900.
12.	24.	950.	950.	−2750.	−2750.
25.	50.	139.	139.	−2000.	−2000.
33.	66.	210.	210.	−1500.	−1500.
35.	70.	178.	178.	1600.	1600.
15.	30.	555.	555.	1800.	1800.
23.	46.	710.	710.	2100.	2100.
22.	44.	345.	345.	2700.	2700.
35.	70.	455.	455.	2850.	2850.
27.	54.	825.	825.	3020.	3020.
40.	50.	177.	177.	−3335.	−3335.
20.	10.	123.	123.	−3100.	−3100.
15.	40.	275.	275.	2400.	2400.
39.	58.	700.	700.	3000.	3000.
35.	15.	765.	865.	3100.	2900.
14.	50.	455.	555.	3500.	3400.

these matrices will have a rank approximated at the most by $S_1 + S_2 - P$, because in $\Delta X(K)$ there will be a number of independent complex decaying sinusoids equal at the most to the sum of the number of X_1 and X_2 complex decaying sinusoids minus P, being the proportional P decaying sinusoids counted twice in $S_1 + S_2$. However, for $K = R$ and $\eta_1 \approx \eta_2$, the rank of $\Delta X(R)$ will be approximated at the most by $S_1 + S_2 - 2 * P$, because the P proportional complex decaying sinusoids in X_1 and X_2 cancel each other out.

In order to obtain the normalization factor R, the algorithm, which operates in the time domain, consists in calculating the minimum rank of the Hankel matrix in Eq. [9]. If the signal-to-noise ratio is not too low, this calculation can be accomplished by the singular value decomposition looking at an eventually occurring discontinuity in the ordered singular values of the Hankel matrix in Eq. [9]. In fact, due to noise, the singular values of the matrix in Eq. [9] will be all different from zero, but singular values related to noise will be very much smaller than those related to signals. In this manner, in the ordered singular values there will be a discontinuity between signal- and noise-related singular values. The rank of the matrix in Eq. [9] will be approximated by the number of singular values related to signals and when $K = R$, this number will be minimum.

Simulations Testing the Algorithm

In order to test the MiRaNAl algorithm, both simulated FIDs and experimental ^{1}H NMR FIDs of known contents were utilized. The simulated FIDs were generated by complex superposition of exponential decaying sinusoids with additive Gaussian noise. Each point of these Monte Carlo simulations consisted in 30 independent Gaussian noise realizations of couples of FIDs. The program was written in Matlab and in order to generate random numbers, the *randn* Matlab function was used. Each FID consisted of 20 complex decaying sinusoids with a number of sinusoid amplitudes in a ratio of 1:2. The range of amplitudes, in arbitrary units, went from 0 to 100 (the range presented is relative to the FIDs with smaller amplitudes, so that the same range has to be duplicated for the amplitudes of FIDs with greater amplitudes). The linewidth range was 30–300 Hz. The number of nonproportional complex decaying sinusoids, that is, of those sinusoids having amplitude ratios different from 1:2, was varied from 4 to 18, always with 20 as the total number of sinusoids per FID.

Since $R = 2$, the minimization in the MiRaNAl algorithm was carried out in the simplest way, that is, by making K vary from 1.6 to 2.4 with a step of 0.01 and calculating the minimum rank of the Hankel matrix in Eq. [9]. The estimate of R was the K value for which the minimum rank was attained.

Four different sets of Monte Carlo simulations were carried out.

Discontinuity in the ordered singular values. The first simulation was designed to show how, when noise is not zero, there is a discontinuity in the ordered singular values of the Hankel matrix in Eq. [9] and how the minimum rank, and then the normalization factor R, is obtained by looking at this discontinuity. Two FIDs, with the parameter values in Table 1 and $N = 256$, $\Delta t = 0.000125$s, $\delta_1 = 0.00030$, $\delta_2 = 0.00035$, $\phi_{01} = 0.54$ rad, $\phi_{02} = 0.35$ rad, and SNR = 191, were considered. As can be seen in this table, the two FIDs have 14

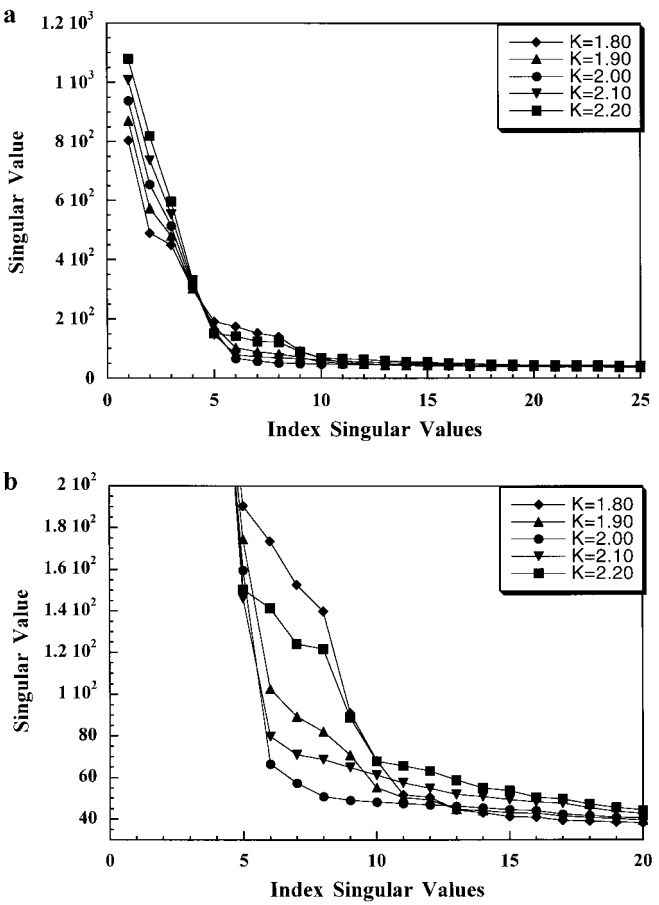

FIG. 2. Ordered singular values of a Hankel matrix obtained by two FIDs for different K values versus their index: (a) the first 25 ordered singular values; (b) the zoom in the first 20 ordered singular values in the range 5–20.

complex proportional decaying sinusoids in a ratio of 1:2, that is, the two FIDs have a normalization factor of $R = 2$, while the other four have nonproportional amplitudes, and the last two have different dampings and different frequencies. When correctly normalized, there will be 6 independent complex decaying sinusoids and so the rank of the Hankel matrix in Eq. [9], for $K = R = 2$, has to be approximated by 6. In Fig. 2a are reported the first 25 ordered singular values of the Hankel matrix in Eq. [9] obtained with $L = 80$ and $K = 1.8, 1.9, 2.0, 2.1, 2.2$. As can be seen, in the range of the first 5–10 singular values there is a discontinuity for all five K values. Figure 2b is a zoom of Fig. 2a in order to better show the region of interest. As this figure shows, for $K = R = 2.0$ there is a discontinuity around the 6th singular value, while for $K \neq 2.0$, the discontinuity appears after the 10th singular value. From the figure, it can be assumed that the minimum approximated rank of the Hankel matrix for the two FIDs considered is approximately 6 and occurs for $K = R = 2.0$.

Dependence on the degree of diversity. In these and in the following Monte Carlo simulations, in order to obtain the index of the singular value which separates signal-related from noise-related singular values, an automatic procedure was used. In particular, if $\sigma_{p-p/4+1}, \sigma_{p-p/4+2}, \ldots, \sigma_p$ was the last quarter of singular values, certainly all noise related, denoted by

$$\tau = \max_i\{(\sigma_{p-i} - \sigma_{p-i+1})\},$$
$$i = p/4 - 1, p/4 - 2, \ldots, 1,$$

the index required was the last j, with $j = 1, 2, \ldots, p - 1$, for which the condition $(\sigma_j - \sigma_{j+1}) \geq 1.5 * \tau$ was satisfied.

The second set of Monte Carlo simulations was designed to study the dependence of the algorithm on the degree of diversity. If two FIDs have all the complex decaying sinusoids that are proportional by a constant factor R, then the two spectra, obtained by Fourier transforming the FIDs, can be superimposed simply by multiplying one of them by R. Normalizing the two spectra, in this case, is very easy: it is sufficient, in fact, to look at the constant factor which makes them superimposed. When the two FIDs have only some complex decaying sinusoids which are proportional by a constant R, the two spectra result superimposed only in some regions, but not totally, when correctly normalized. The degree of diversity of two spectra is defined as the percent ratio between the number of points that are not superimposed and the total number of spectral points, when the two spectra are correctly normalized. On the other hand, when the degree of diversity is small (less than 50%), normalizing a couple of spectra in the frequency domain is easy because, in that case, they result superimposed in the majority of spectral points (*6*). However, the degree of diversity depends on the number of complex decaying sinusoids which are not proportional and on the characteristics of complex decaying sinusoids. Therefore, two spectra can have a great number of proportional complex decaying sinusoids and at the same time a great degree of diversity. However, in this case, in the frequency domain, it is not possible to determine the normalization constant. On the other hand, the MiRaNAl algorithm, which operates in the time domain, should not depend on the degree of diversity since it is not based on the superposition of spectral points.

The results of the MaSNAl algorithm in the frequency domain and those of the MiRaNAl algorithm were compared. Before applying the MaSNAl algorithm the FIDs were zero-filled to 32,768 points and then Fourier transformed. The parameters used to quantitatively compare the behavior of the two above-mentioned algorithms were the absolute percent bias (bias) and the variance (var) defined by the following equations, respectively (*19*),

$$\text{bias}(K) = |E[K] - R|/R * 100$$
$$\text{var}(K) = E[(K - E[K])^2],$$

where $E[\ ]$ denotes the mean operation.

In Fig. 3a, the absolute percent bias is reported as a function

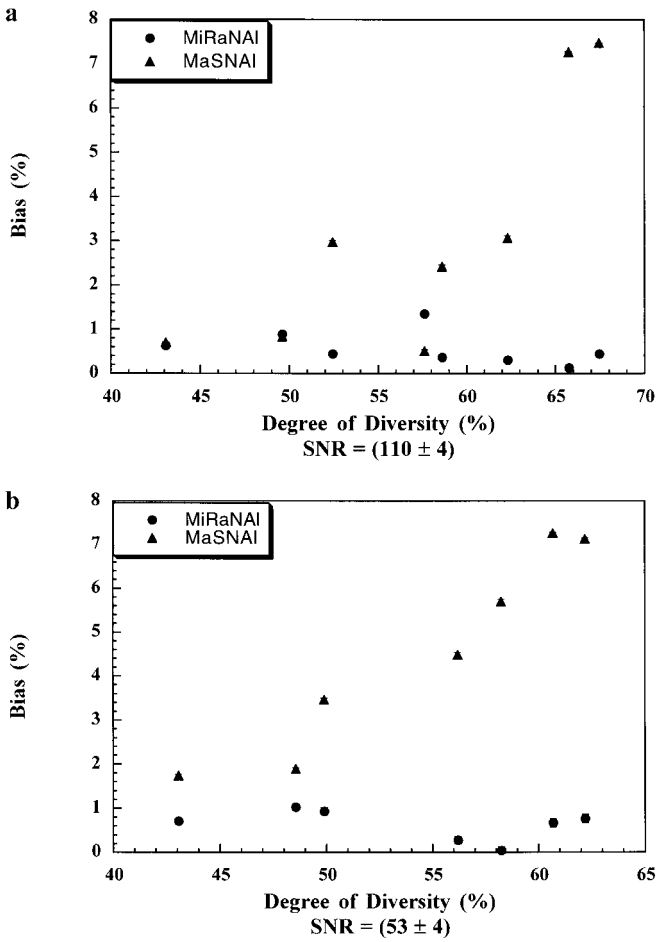

FIG. 3. Percent absolute bias versus degree of diversity: ●, MiRaNAl algorithm; ▲, MaSNAl algorithm: (a) mean SNR = 110 ± 4; (b) mean SNR = 53 ± 4.

of the degree of diversity for simulated spectra having a signal-to-noise ratio of 110 ± 4. As this figure shows, the time-domain MiRaNAl algorithm is, as expected, quite independent of the degree of diversity. Therefore, it is better than the frequency-domain MaSNAl algorithm which cannot be applied for a degree of diversity greater than 50%. Furthermore, the absolute percent bias is reported as a function of the degree of diversity for simulated spectra having a signal-to-noise ratio of 53 ± 4 (Fig. 3b). This figure confirms that the time-domain MiRaNAl algorithm is quite independent of the degree of diversity for noisier spectra as well and again that it is better than the frequency-domain MaSNAl algorithm which does not work well for degree of diversity $>$ 50%. In particular, the MiRaNAl algorithm also has a percent bias of less than 1% for degree of diversity $>$ 50%, when the MaSNAl algorithm gives very biased results.

In conclusion, Figs. 3a and 3b show that the MiRaNAl algorithm is quite independent of the degree of diversity of the spectra and therefore can also be applied in situations in which there are proportional complex decaying sinusoids with a degree of diversity greater than 50%. The MiRaNAl algorithm is also applicable in situations in which it is not possible to obtain the normalization constant in the frequency domain.

In Figs. 4a and 4b the variance is reported as a function of the degree of diversity. In particular, in these figures, it can be observed that the variance is quite independent of the degree of diversity of the spectra and that there are no significant differences in the dispersion of the MaSNAl and MiRaNAl algorithms' estimated values around the true parameter value.

Dependence on the signal-to-noise ratio. The third set of Monte Carlo simulations was designed to examine the dependence of the algorithm on the spectral noise. The signal-to-noise ratio was defined by

$$\mathrm{SNR} = \frac{\psi(\nu_0) - \bar{\psi}}{2 \times \left[\sum_{i=a1}^{a2} \frac{(\psi(\nu_i) - \bar{\psi})^2}{N} \right]^{1/2}}, \qquad [10]$$

where $\bar{\psi} \equiv [\sum_{i=a1}^{a2} \psi(\nu_i)]/(N + 1)$, $\psi(\nu_0)$ is the maximum peak

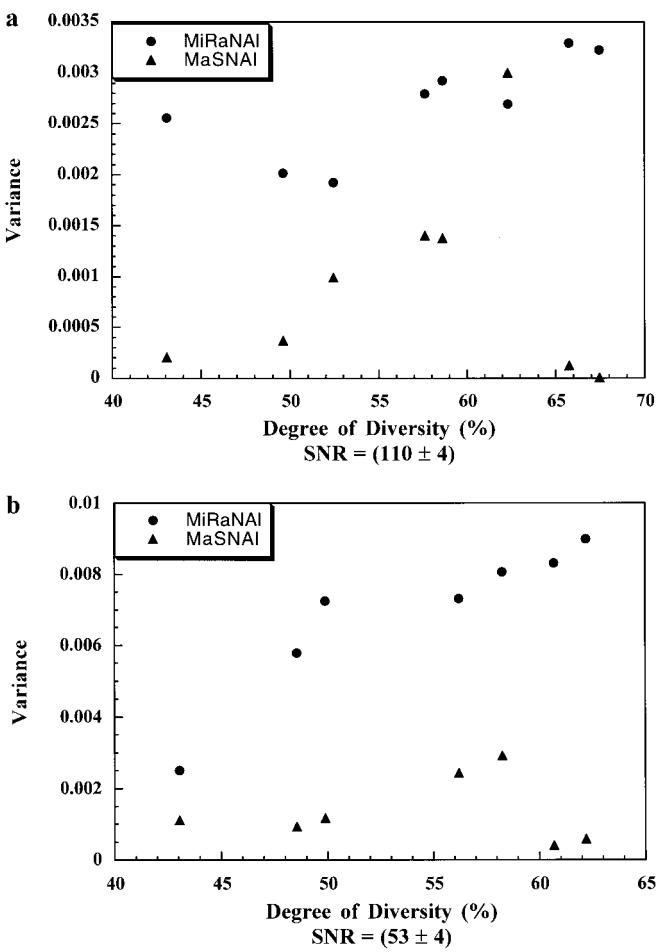

FIG. 4. Variance versus degree of diversity: ●, MiRaNAl algorithm; ▲, MaSNAl algorithm: (a) mean SNR = 110 ± 4; (b) mean SNR = 53 ± 4.

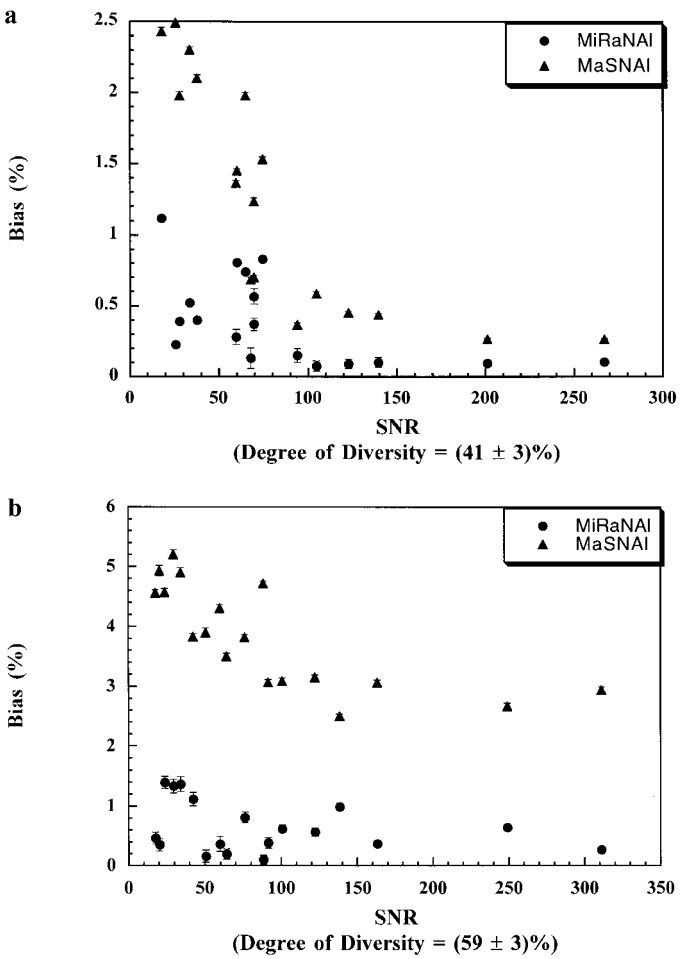

FIG. 5. Percent absolute bias versus SNR: ●, MiRaNAl algorithm; ▲, MaSNAl algorithm: (a) mean degree of diversity = 41 ± 3%; (b) mean degree of diversity = 59 ± 3%.

height, $a1$ and $a2$ are the limits of the noise region with $N + 1 = a1 - a2$, and $\bar{\psi}$ is the DC level of the noise region. In Fig. 5a, the absolute percent bias is reported as a function of the SNR for simulated spectra having a degree of diversity of 41 ± 3%. This figure shows that the MiRaNAl algorithm is quite robust since it has a bias of less than 1% for spectra with a low signal-to-noise ratio (SNR $<$ 18) as well. Furthermore, the MiRaNAl algorithm proved to be more robust than the MaSNAl algorithm which for SNR $<$ 50 is more biased. In Fig. 5b, the absolute percent bias is reported as a function of the SNR for simulated spectra having a degree of diversity of 59 ± 3%. This figure confirms that the MiRaNAl algorithm is very robust (bias less than 1%) and, above all, that it is also applicable in situations in which the MaSNAl frequency-domain algorithm does not work well (degree of diversity $>$ 50%).

Dependence of R values. The last set of Monte Carlo simulations was designed to test the MiRaNAl algorithm for different values of the normalization factor. In particular, each point of this simulation regards simulations with different true R normalization factor values (Fig. 6). In this figure, the estimated value of the normalization factor is reported as a true R value function. As can be seen, there is an optimal correlation between estimated and true values.

From the results presented above, the Monte Carlo simulations demonstrate that, for signal-to-noise ratios greater than 40, the MiRaNAl algorithm is able to determine the normalization factor R of two NMR spectra with a bias of 1% at most.

Quantitative Relationship between Two Spectra Obtained by the MiRaNAl and MaSNAl Algorithms and a More Traditional Added Standard Method

In order to test the validity of the MiRaNAl algorithm experimentally, two samples of known contents were compared. Both samples contained thyrotropin releasing factor (THR, Calbiochem, MW 362.4), deuterated methanol (CD_3OD 99.96%, Cambridge Isotope Laboratories), and sodium trimethylsilyl[2,2,3,3-d_4]propionate (TSP, 10 μmol/ml). The first sample (sample A) consisted of 4.5 mg of THR, 400 μl of CD_3OD, and 10 μl of TSP, while the second sample (sample B) consisted of 4.5 mg of THR, 600 μl of CD_3OD, and 10 μl of TSP.

Five ^{1}H NMR spectra were obtained, for each sample, using a Bruker DPX digital spectrometer operating at 300 MHz. The spectra were accumulated with a 90° flip angle pulse and 64 transients of 8K data points corresponding to a spectral window of ±2097.3 Hz.

Using the traditional method, that is, making a quantitative

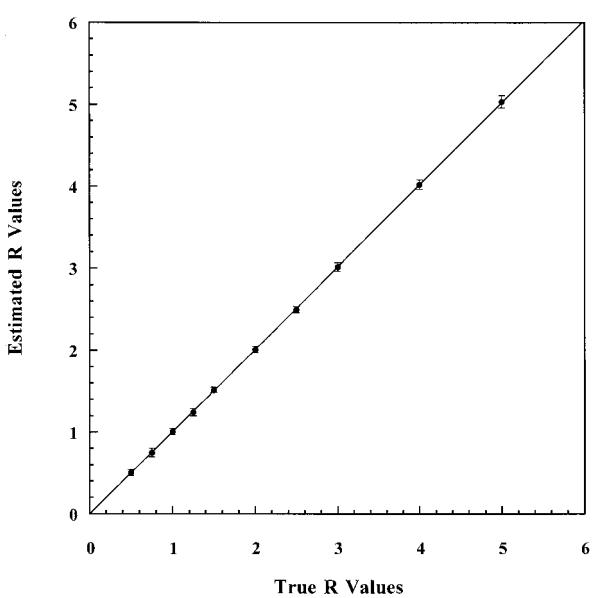

FIG. 6. Estimated R values versus true R values. Couples of FIDs with SNR = 203 ± 14. Curve fit: $y = ax + b$ with $a = 1.006$, $b = -0.002$. Correlation ratio $r = 1.0$.

analysis using the TSP standard, the following results were obtained (*6*):

Sample	THR	CD_3OD residual peak
A	38.7 ± 3.6 μmol/ml	41.9 ± 3.9 μmol/ml
B	20.2 ± 2.5 μmol/ml	36.1 ± 4.6 μmol/ml

[11]

If the percent difference ($\Delta_{CD_3OD}(\%)$) of CD_3OD relative to THR of the sample A spectrum with respect to the sample B spectrum is to be computed, the following equation can be used,

$$\Delta_{CD_3OD}(\%) = \frac{CD_3OD(A) - \dfrac{CD_3OD(B)}{THR(B)}\,THR(A)}{\dfrac{CD_3OD(B)}{THR(B)}\,THR(A)} \times 100, \quad [12]$$

where $CD_3OD(A)$ is the CD_3OD concentration in spectrum A and the other symbols have similar meanings.

The CD_3OD concentrations are proportional to those obtained by considering the CD_3OD residual peaks due to residual protons (*20*); thus, the concentrations obtained by the CD_3OD residual peaks in Eq. [11] can be used directly in Eq. [12]. With this equation, using the concentrations obtained by the TSP standard quantification (traditional method), a $\Delta_{CD_3OD}(\%) = (-39.4 \pm 8.7)\%$ percent difference was obtained (*6*), while, from the known quantities of the added substances and using the same above-cited equation, a $\Delta_{CD_3OD}(\%) = (-33.4 \pm 0.2)\%$ difference was expected.

The same two spectra (A, B) were utilized to obtain the percent difference $\Delta_{CD_3OD}(\%)$ of CD_3OD relative to THR by using the MaSNAl and MiRaNAl algorithms. The spectra were normalized with the algorithms. At this point, the percent difference $\Delta_{CD_3OD}(\%)$ of CD_3OD relative to THR between the spectrum of sample A, normalized with respect to the spectrum of sample B, was obtained directly by comparing the areas of the two spectral CD_3OD residual peaks. The value found with the MaSNAl algorithm was $\Delta_{CD_3OD}(\%) = (-34.2 \pm 1.3)\%$ (*6*), while that found with the MiRaNAl algorithm was $\Delta_{CD_3OD}(\%) = (-32.5 \pm 1.6)\%$.

As can be seen, all three methods yielded percent difference results which were consistent with the expected ones. However, the MiRaNAl and MaSNAl algorithms allowed good results to be obtained without the use of any standard and without quantifying all the spectral lines, but rather by comparing the signals of interest in the normalized spectra.

Biological Application

The MiRaNAl algorithm was applied to the normalization of NMR spectra of cell samples. In particular, the ^1H NMR spectra of control A431 cells (an epidermoid carcinoma) and of these cells exposed to a sublethal dose of 3 Gy of ionizing radiation (Gammacell 220, Atomic Energy of Canada Ltd.) were compared.

The two FIDs, acquired under the same experimental conditions using a Bruker DRX-5.00 spectrometer operating at 500 MHz, were analyzed by the MiRaNAl algorithm and a normalization factor of $R = 1.33$ was found. After normalization, the two normalized FIDs were Fourier transformed: in Fig. 7 are reported the resulting two spectra. As can be seen, the spectra, after the normalization, overlap in a significant number of spectral lines.

In Fig. 8, the difference spectrum obtained from subtraction of the spectrum of controls from that of irradiated cells, after normalization, is reported. As can be seen, the majority of the signals fall around the baseline which appears very flat. In addition, an immediate identification of the spectral components which vary between the two spectra and which are probably the result of the 3 Gy irradiation can be obtained. In particular, important changes are induced in glycerophosphatidylcholine (GPC, 3.24 ppm), phosphatidylcholine (PC, 3.22), choline (Cho, 3.21 ppm), creatine (Cr, 3.03 ppm), and lactate (Lac, 1.33 ppm) and in the lipid region (1.6–0.6 ppm).

The difference spectrum in Fig. 8 shows that there are alterations in creatine (irradiated cells contain a greater amount of Cr) and in lactate (irradiated cells contain a much smaller amount of Lac with respect to control cells) which indicate a perturbation of the energy metabolism of the cells. Furthermore, alterations in phosphatidylcholine (irradiated cells contain a much smaller amount of PC) and in lipid (irradiated cells contain a greater amount of CH_2 and CH_3 lipids than control cells) reveal changes in cell membrane structure.

Thus, from these data, it appears that the MiRaNAl algorithm presented can be adequately utilized to compare NMR spectra of tumor cells.

CONCLUSIONS

In this paper, a time-domain algorithm for the normalization of couples of NMR spectra, in order to compare these spectra and to obtain relative quantitative information without the need of any standards, is presented. It consists in minimizing the rank of the FID difference Hankel matrix and obtaining the normalization constant which is able to reduce the number of linear independent complex decaying sinusoids in the difference of the two FIDs.

Recently an algorithm for NMR spectra normalization was presented (*6*); it was the MaSNAl algorithm operative in the frequency domain. Both the MaSNAl and MiRaNAl algorithms allow normalization by exploiting intrinsic sample properties and thus considerably simplify the measurement procedures. In fact, no addition of substances or particular manipulations of the samples are needed, thus reducing contamination risks. In addition, the algorithms are very easy to

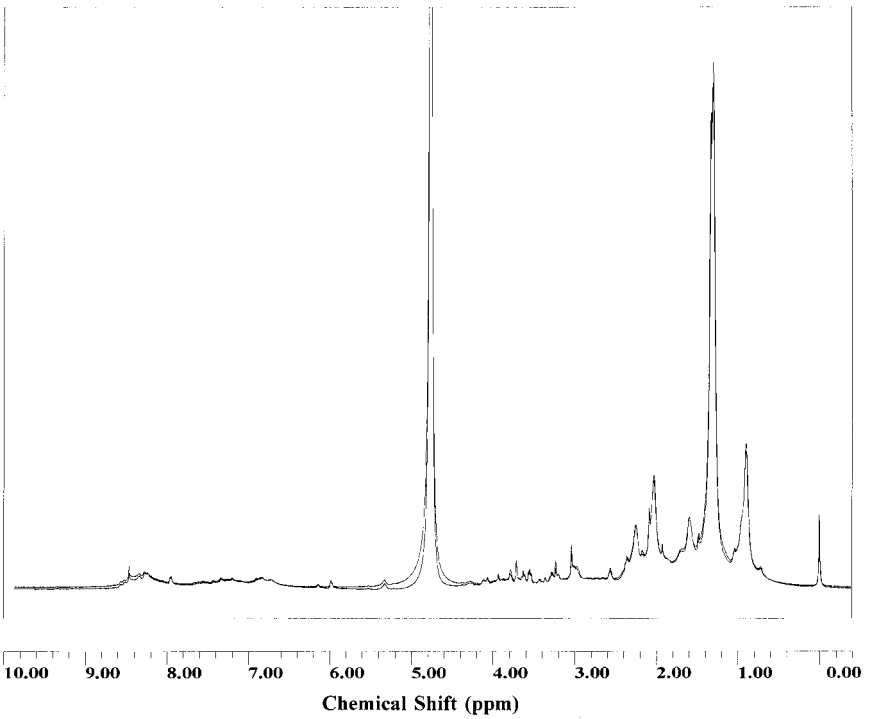

FIG. 7. Superimposed normalized spectra of A431 control cells and of A431 cells irradiated with 3 Gy.

manage and the normalization procedures require only a few minutes. However, MiRaNAl shows some advantages with respect to the MaSNAl algorithm.

First, MiRaNAl operates in the time domain and so does not require FID manipulations (e.g., phase adjustment, baseline correction) reducing, in this manner, the effect of these manual manipulations on the calculation of the normalization factor. In fact, simulations show that the MiRaNAl algorithm is less biased than MaSNAl algorithm.

Furthermore, the MaSNAl algorithm was applicable only when the spectra to be compared had a great number of proportional signals with respect to the total number of signals, that is, when the degree of diversity of the two spectra was less than 50%. This condition is often verified with spectra crowded with signals and when variations, due to the action of an agent, are not so great. In particular, these conditions are often verified with ^{1}H high-resolution NMR spectra of cell samples, but in *in vivo* spectra this is seldom the situation (*21*). Furthermore, in crowded ^{1}H high-resolution NMR spectra of cell samples, changes induced could interest a great number of signals and so the degree of diversity could be greater than 50%.

The MiRaNAl algorithm does not have these kinds of limitations and so has a greater range of applicability than the MaSNAl algorithm. Monte Carlo simulations showed that the MiRaNAl algorithm is consistent and presents very low bias and variance, thus giving an optimal estimation of the normalization factor, also in situations in which the MaSNAl algorithm was not applicable. In particular, the MiRaNAl algorithm could be applied to *in vivo* spectra which are not so crowded and for which the MaSNAl algorithm is not suitable.

In fact, it is an important task in magnetic resonance spectroscopy to monitor metabolic and biochemical changes in numerous disease processes. However, in *in vivo* proton spectroscopy at 1.5 T, the number of proton metabolites that can be observed is very limited and the spectra are not so crowded (*21*). As Monte Carlo simulations demonstrated, the MiRaNAl algorithm is able to correctly normalize spectra with very few signals and, above all, spectra with a few proportional signals, that is, with a degree of diversity greater than 50%. Furthermore, *in vivo* spectra have worse signal-to-noise ratios than *in vitro* spectra and Monte Carlo simulations showed that the MiRaNAl algorithm is less sensitive to SNR ratio than the MaSNAl algorithm.

In conclusion, Monte Carlo simulations showed that the MiRaNAl algorithm extends the range of applicability of the normalization procedure to situations typical of *in vivo* spectroscopy and that it could be a useful tool in monitoring metabolic and biochemical changes in disease processes.

The MiRaNAl algorithm was also tested by comparing two samples of known contents and results were better than those obtained by the traditional method based on the use of a standard, while they were comparable with those obtained by the MaSNAl algorithm. Finally, it was demonstrated that the

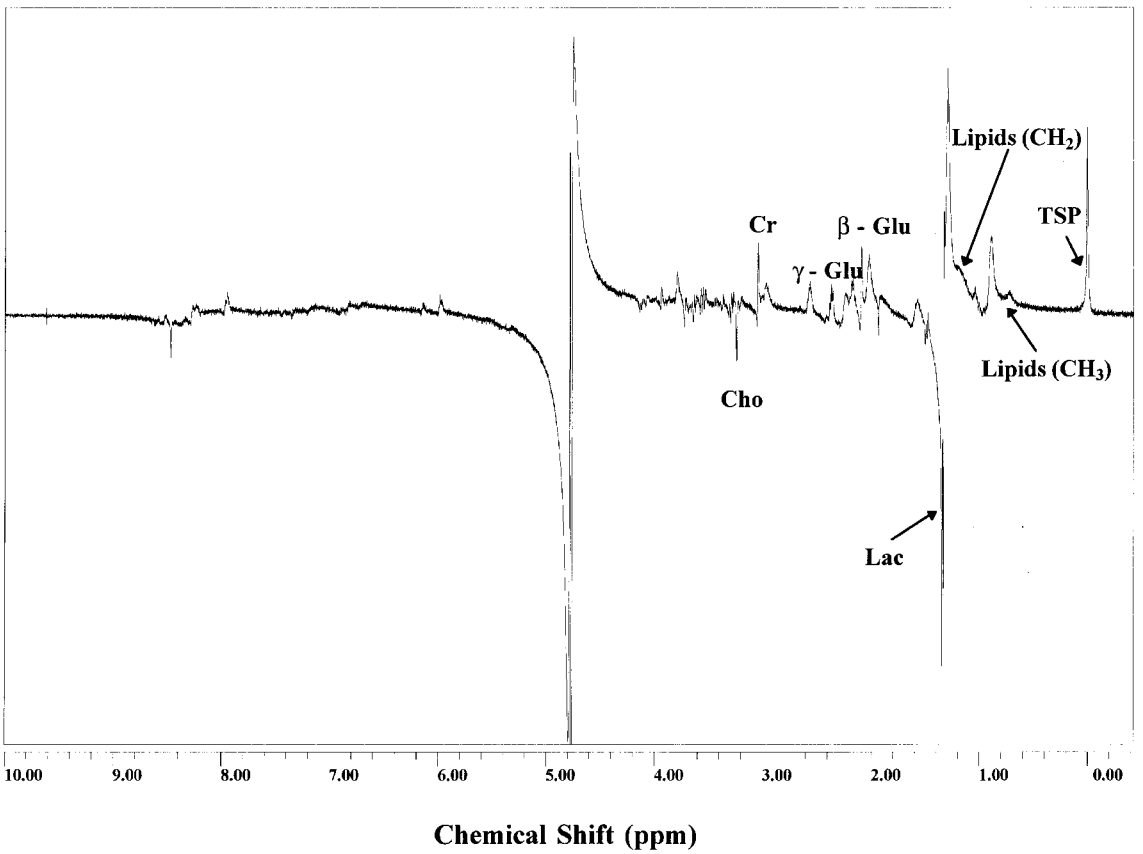

FIG. 8. Difference spectrum obtained from the subtraction of A431 control cells from the spectrum of irradiated cells after normalization.

algorithm can be applied to real spectra of cell samples, allowing the extraction of important biological information.

ACKNOWLEDGMENTS

We thank the Istituto Nazionale per la Fisica della Materia for financial support. We thank Dr. Andrea Motta, Istituto per la Chimica Di Molecole di Interesse Biologico, CNR of Arcofelice, Napoli (Italy), for having acquired the experimental spectra presented in the paper.

REFERENCES

1. C. Rémy, C. Arús, A. Ziegler, E. Sam Lai, A. Moreno, Y. Le Fur, and M. Décorps, In vivo, ex vivo, and in vitro one- and two-dimensional nuclear magnetic resonance spectroscopy of an intrcerebral glioma in rat brain: Assignement of resonances, *J. Neuroschem.* **62,** 166–179 (1994).
2. S. J. Berners-Price, M. E. Sant, R. I. Christopherson, and P. W. Kuchel, ^{1}H and ^{31}P NMR and HPLC studies of mouse L1210 leukemia cell extracts: The effect of Au(I) and Cu(I) diphosphine complexes on the cell metabolism, *Magn. Reson. Med.* **18,** 142–158 (1991).
3. S. M. Ronen, E. Rushkin, and H. Degani, Lipid metabolism in large T47D human breast cancer spheroids: ^{31}P- and ^{13}C-NMR studies of choline and ethanolamine uptake, *Biochim. Biophys. Acta* **1138,** 203–212 (1992).
4. S. Cérdan, R. Parrilla, J. Santoro, and M. Rico, ^{1}H NMR detection of cerebral *myo*-inositol, *FEBS Lett.* **187,** 167–172 (1985).
5. H. Bachelard and R. Badar-Goffer, NMR spectroscopy in neurochemistry, *J. Neurochem.* **61**(2), 412–429 (1993).
6. R. Romano, R. Lamanna, M. T. Santini, and P. L. Indovina, A new algorithm for NMR spectral normalization, *J. Magn. Reson.* **138,** 115–122 (1999).
7. R. Lamanna, A. Motta, R. Romano, G. Rainaldi, F. Flamma, M. Pentimalli, T. Tancredi, P. L. Indovina, and M. T. Santini, Forced adhesive growth of K562 leukemic cells that normally grow in suspension induces variations in membrane lipids and energy metabolism: A proton NMR study, *J. Biomed. Mater. Res.* **46,** 171–178 (1999).
8. M. Reed and B. Simon, "Methods of Modern Mathematical Physics. I. Functional Analysis," Academic Press, New York (1972).
9. F. Riesz and B. Sz. Nagy, "Functional Analysis," Ungar, New York (1965).
10. C. L. Lawson and R. G. Hanson, "Solving Least Squares Problems," Prentice–Hall, Englewood Cliffs, NJ (1974).
11. G. E. Forsythe, M. A. Malcolm, and C. B. Moler, "Computer Methods for Mathematical Computations," Prentice–Hall, Englewood Cliffs, NJ (1977).
12. V. C. Klema and A. J. Laub, The singular value decomposition: Its computation and some applications, *IEEE Trans. Autom. Control* **AC-25,** 164–176 (1980).
13. J. Cadzow, Signal enhancement: A composite property mapping

algorithm, *IEEE Trans. Acoustics Speech Signal Processing* **ASSP-36,** 49–62 (1988).

14. M. K. Maple, "Digital Spectral Analysis with Applications," Prentice–Hall, Englewood Cliffs, NJ (1996).
15. A. van den Boogaart, M. Ala-Korpela, J. Jokisaari, and J. R. Griffiths, Time and frequency domain analysis of NMR data compared: An application to 1D ^{1}H spectra of lipoproteins, *Magn. Reson. Med.* **31,** 347–358 (1994).
16. H. Chen, S. van Huffel, D. van Ormondt, and R. de Beer, Parameter estimation with prior knowledge of known signal poles for the quantification of NMR spectroscopy data in the time domain, *J. Magn. Reson. A* **119,** 225–234 (1996).
17. M. Lupu and D. Todor, Linear prediction and singular value decomposition in NMR signal analysis, *in* "Signal Treatment and Signal Analysis in NMR" (D. N. Rutledge, Ed.), pp. 169–171, Elsevier, Amsterdam (1996).
18. H. Chen, S. Van Huffel, C. Decanniere, and P. Van Hecke, A signal-enhancement algorithm for the quantification of NMR data in the time domain, *J. Magn. Reson. A* **109,** 46–55 (1994).
19. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C: The Art of Scientific Computing," 2nd ed., Cambridge Univ. Press, Cambridge, UK (1986).
20. A. E. Derome, "Modern NMR Tecniques for Chemistry Research," Prentice–Hall, Englewood Cliffs, NJ (1996).
21. L. Kwock, Clinical proton magnetic resonance spectroscopy: Basic principles, *in* "Clinical Application of MR Spectroscopy" (Suresh K. Mukherji, Ed.), Wiley, New York (1998).