Journal of Magnetic Resonance **138,** 115–122 (1999)
Article ID jmre.1999.1708, available online at http://www.idealibrary.com on IDEAL®

A New Algorithm for NMR Spectral Normalization[1]

Rocco Romano,* Raffaele Lamanna,† Maria Teresa Santini,‡ and Pietro Luigi Indovina*

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

Received September 15, 1998; revised December 30, 1998

There is increasing use of high-resolution NMR spectroscopy to examine variations in cell metabolism and/or structure in response to numerous physical, chemical, and biological agents. In these types of studies, in order to obtain relative quantitative information, a comparison between signal intensities of control samples and treated or exposed ones is often conducted. The methods thus far developed for this purpose are not directly related to the overall intrinsic properties of the samples, but rather to the addition of external substances of known concentrations or to indirect measurement of internal substances. In this paper, a new method for quantitatively comparing the spectra of cell samples is presented. It depends on a normalization algorithm which takes into consideration all cell metabolites present in the sample. In particular, the algorithm is based on maximizing, by an opportune sign variable measure, the spectral region in which the two spectra are superimposed. The algorithm was tested by Monte Carlo simulations as well as experimentally by comparing two samples of known contents with the new method and with an older method using a standard. At the end, the algorithm was applied to real spectra of cell samples to show how it could be used to obtain qualitative and quantitative biological information.

Key Words: **NMR; algorithm; normalization; NMR of cells.**

INTRODUCTION

There is increasing use of high-resolution NMR spectroscopy to examine variations in cell metabolism and/or structure in response to numerous physical, chemical, and biological agents. In these types of studies, in order to obtain relative quantitative information, a comparison between signal intensities of control samples and treated or exposed samples is often conducted. For this purpose various methods have been developed. For instance, reference compounds added directly to the cell samples have been used. These compounds may be synthetic, such as sodium 3-trimethylsilyl[2,2,3,3-d_4]propionate (*1*, *2*), or naturally occurring, such as glucose in ^{13}C spectra and inorganic phosphate in ^{31}P spectra (*3*). In addition, enzymatic determination of the concentration of a metabolite present in the samples themselves (e.g., the alanine methyl doublet) and a subsequent comparison of the intensities of resonances with that of the metabolite itself have also been utilized (*4*). All these methods are related not directly to the overall intrinsic properties of the samples, but rather to the addition of external substances of known concentration or to indirect measurement of internal substances.

In this paper, a new method for quantitatively comparing the spectra of cell samples is presented. It is based on a normalization algorithm which takes into consideration all cell metabolites present in the sample. Since concentration differences result in proportional variations of spectral intensities, nonproportional changes can most likely be attributed to the effects of the agent. The proportional variations are described by the normalization factor R, which can be calculated by the algorithm. In particular, it consists in maximizing, by using a sign variable measure, the spectral regions in which spectral lines are proportional. The algorithm was tested by Monte Carlo simulations, which demonstrated its ability to determine the normalization factor with very low bias. In order to test the validity of the algorithm experimentally, two samples of known contents were compared using a traditional method, based on the use of a standard, and the new one, based on the algorithm.

At the end, the algorithm was applied to real spectra of cell samples showing how it could be used to obtain qualitative and quantitative biological information.

RESULTS AND DISCUSSION

The Algorithm

Let us consider two NMR spectra, $\psi_1(\nu)$ and $\psi_2(\nu)$, where ν is the frequency. If Sp is the entire spectrum, let us denote by $Sp - S_R$ (normalizable regions) the spectral regions relative to the baseline and to the spectral lines which have proportional intensities due to concentration differences, while S_R denotes

[1] The MaSNAl program, an Apple Power PC version or the C version, is freely available from Dr. Rocco Romano, Dipartimento di Scienze Fisiche, Università di Napoli Federico II, Istituto Nazionale per la Fisica della Materia, Unità di Napoli, Complesso Universitario Monte S. Angelo, Via Cinthia, 80126 Naples, Italy. E-mail: rocco.romano@na.infn.it. Please send a signed letter confirming that the program will be used for scientific, noncommercial purposes only.

1090-7807/99 $30.00

those regions relative to spectral lines with various ratio intensities due to effects induced by interaction with a particular agent. Two spectral points, $\psi_1(\nu_0)$ and $\psi_2(\nu_0)$, can be considered physically superimposed, for a given estimate of the normalization factor K, if their distance,

$$d(\nu_0; K) \equiv |\psi_2(\nu_0) - K*\psi_1(\nu_0)|, \quad [1]$$

is equal to or less than 2ϵ, where ϵ is the rms noise. If the two above-mentioned spectra are correctly normalized, spectral points belonging to the normalizable regions $Sp - S_R$ are well superimposed, such that the area between the spectra in these regions is at a minimum. An estimate of the normalization factor R can then be obtained by finding the value of K which minimizes the integral

$$\int_{Sp-S_R} d(\nu; K)d\nu. \quad [2]$$

Unfortunately, in the integration domain of the integral in Eq. [2] there is the S_R set, which is not known, being dependent on the unknown R parameter.

Let us define a new effective distance,

$$\delta(\nu_0; K) \equiv \theta(d(\nu_0; K) - 2\epsilon)d(\nu_0; K), \quad [3]$$

where $\theta(x)$ is the step function of Heaviside (*5*, *6*), defined by

$$\theta(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}, \quad [4]$$

which is zero for each pair of physically superimposed points.

Let us consider the S family of S_K sets,

$$S_K = \{\nu \in Sp: \delta(\nu; K) > 0\}. \quad [5]$$

By definition, S_R belongs to this family. Furthermore, let us define the *degree of diversity* between the two spectra as the percentage ratio between S_R and Sp. The degree of diversity can be considered to be a measure of the variations induced by interaction with the agent.

If the degree of diversity of the two spectra is less than 50%, one has to expect that S_R is the minimum size set in S. In fact, for $\epsilon = 0$ (that is, for ideal spectra without noise), if the degree of diversity of the two spectra is less than 50% (i.e., for $K = R$ more than 50% of spectral points are superimposed) and S_R is not the minimum set in S, there should be a K value, $K = K^* \neq R$, such that the number of points in S_{K^*} is less than the number of points in S_R. In such a case, for $K = K^* \neq R$ more than 50% of the spectral points are well superimposed. Thus, for $K = R$, more than 50% of spectral points should not be superimposed, because points which are superimposed for $K = K^* \neq R$ cannot be superimposed for $K = R$. But that is in contrast with the hypothesis that the degree of diversity is less than 50% and thus, for $\epsilon = 0$, S_R must be the minimum set in S.

For this reason, one may ask that the size of the normalizable region,

$$I(K) = \int_{Sp-S_K} d\nu, \quad [6]$$

should be a maximum when $K = R$. This condition is sufficient by itself to give an estimate for the normalization factor R provided that a degree of diversity of less than 50% and spectra not to noisy are considered.

For the S family, a set of measures $\mu_K(\{\nu\})$ for which $\mu_K(S_K) = 0$ for all K can be easily constructed. Thus, the above-mentioned integral can be rewritten as

$$I(K) = \int_{Sp} d\mu_K(\nu). \quad [7]$$

The simplest family of measures having a null value on S_K is the Dirac one (*6*),

$$\mu_{K_D}(\{\nu\}) = \chi_{(-\infty,0]}(\delta(\nu; K)), \quad [8]$$

with $\chi_{(-\infty,0]}(\delta(\nu; K))$ being the characteristic function of $(-\infty, 0]$, defined by the relation (*6*)

$$\chi_{(-\infty,0]}(x) = \begin{cases} 1 & \text{if } x \in (-\infty, 0] \\ 0 & \text{if } x \notin (-\infty, 0] \end{cases}. \quad [9]$$

With this measure, maximizing the integral in Eq. [7] is equivalent to maximizing the number of superimposed spectral points. In the above description, the only effect of noise considered thus far is defining as superimposed those points for which $d(\nu; K) <= 2\epsilon$. However, noise also produces casual superpositions, that is, superpositions that cannot be considered to belong to the normalizable regions, and prevents some superpositions that ought to be present. In addition, it should be considered that isolated crosslinks (points of intersection between two spectral lines) also satisfy the condition $\delta(\nu; K) \leq 0$ and thus may be improperly included in the integral of Eq. [7]. In fact, crosslinks cannot belong to the normalizable regions where spectral lines have all the same proportionality and thus cannot intersect between themselves, but from the Dirac measure's point of view they are superimposed points and must be included in the integral of Eq. [7]. In order to consider these effects, a measure which assigns a weight to points based on their neighbors must be introduced. It can be assumed that a random noise superposition event or a crosslink is isolated and thus the neighboring points are not superimposed, while ν

values from normalizable regions have neighboring points which are also superimposed.

In order to take these considerations in account, the maximum superposition normalization algorithm, which will be denoted as MaSNAl, is presented. In particular this new algorithm is based on the sign variable measure (*5–8*)

$$\mu_{\mathrm{MaSNAl}}(\{\nu\}) = \chi_{(-\infty,0]}(\delta(\nu; K)) \int_{\nu-3\beta}^{\nu+3\beta} (1 - 2\theta(\delta(\nu'; K))) \times e^{-(\nu-\nu')^2/\beta^2} d\nu', \quad [10]$$

where β is an opportune constant and the other quantities have already been defined.

The first factor in Eq. [10] is the Dirac term, which allows $\mu_{\mathrm{MaSNAl}}(\{\nu\})$ to satisfy the condition $\mu_{\mathrm{MaSNAl}}(S_K) = 0$ for all K. The second factor assigns to the ν values a weight which depends upon the neighboring points in the interval $(\nu - 3\beta, \nu + 3\beta)$. In fact, due to the Gaussian weighting, if the limits on the integral were changed to $(-\infty, +\infty)$, the measure would not be affected very much. To each neighboring superimposed ν' value is assigned a weight of $e^{-(\nu-\nu')^2/\beta^2}$ which depends on the distance of ν' from ν, while to the nonsuperimposed ν' values, a weight of $-e^{-(\nu-\nu')^2/\beta^2}$ is assigned. In this manner, the intrinsic symmetry between superimposed and nonsuperimposed neighboring points is respected and ν values which have the majority of neighboring points superimposed give a greater contribution to the integral

$$I(K) = \int_{Sp} \chi_{(-\infty,0]}(\delta(\nu; K)) \int_{\nu-3\beta}^{\nu+3\beta} (1 - 2\theta(\delta(\nu'; K))) \times e^{-(\nu-\nu')^2/\beta^2} d\nu' d\nu, \quad [11]$$

which must be maximized. On the other hand, ν values which have many nonsuperimposed neighboring points can actually give a negative contribution.

It should be noted that the measure presented above depends on the parameter β. Nevertheless, it can be shown that this dependence is very weak and that, in particular, there is a threshold such that for β values greater than this threshold, the results are practically independent of the parameter itself.

Simulations Testing the Algorithm

Both simulated spectra and experimental ^{1}H NMR spectra of known contents were utilized to test the MaSNAl algorithm. The simulated spectra were generated by Fourier transformation of complex superposition of exponential decaying sinusoids with additive Gaussian noise. Each point of these Monte Carlo simulations consisted of 50 independent Gaussian noise realizations of couples of spectra (each FID consisting of 2048 points, zero-filled to 32,768 points and then Fourier transformed). The program was written in C[1] and to generate random numbers the Minimal random number generator by Park and Miller with Bays–Durham shuffle and added safeguards was used (*9*). Each spectrum was made of 81 spectral lines with the majority of spectral lines having amplitudes in a ratio of 1:2. The range of amplitudes were, in arbitrary units, from 35,000 to 60,000 (the range presented is relative to the spectra with smaller amplitudes, so that the same range must be duplicated for the amplitudes of spectra with greater amplitudes). The line width range was 20–96 Hz. The number of nonproportional spectral lines, that is, of those spectral lines having ratios different from 1:2, was varied from 2 to 20, always with 81 as the total number of spectral lines per spectrum. The amplitudes of these spectral lines were distributed between those having a greater intensity in spectrum 1 and those having a greater intensity in spectrum 2 in such a manner that the diversity degree level increased. Since $R = 2$, the maximization in the MaSNAl algorithm was carried out in the simplest way, that is, making K vary from 1.6 to 2.4 with a step of 0.01 and calculating the real value of the integral Eq. [11] that was to be maximized. The estimate of R was the K value for which the maximum was attained. We used this strategy so as not to introduce bias due to the procedure of maximization.

Four different sets of Monte Carlo simulations were carried out.

Dependence on β. The first simulation was designed to show that the algorithm is weakly dependent on β and to determine an optimum value for the parameter itself. In particular, three simulations at different signal-to-noise ratios (SNR) and degree of diversity values were performed. For each simulation, a normalization factor of $R = 2$ was chosen and estimated K values were found by the MaSNAl algorithm (● in Fig. 1) versus β values. As can be seen in this figure, in all three simulations mentioned above, for β values greater than 10, K estimations practically no longer depended on β. In particular, a good value for this parameter was found to be $\beta = 15$. In fact, greater β values not only did not give better results, but also lengthened computation time. In any case, K estimations better than those given by the Dirac measure in Eq. [8] (○ in Fig. 1) were obtained, even for $\beta = 1$.

Dependence on the degree of diversity. The second set of Monte Carlo simulations was designed for studying the dependence of the algorithm on the degree of diversity. In particular, results of the MaSNAl algorithm were compared with both data obtained by maximizing the integral in Eq. [7], using the Dirac measure of Eq. [8] (Dirac algorithm), and results obtained by minimizing the integral in Eq. [2], where the integration domain was arbitrarily extended over the entire spectrum (Lebesgue algorithm). The parameters used to quantitatively compare the behavior of the three abovementioned algorithms were the bias, the variance (var), and the

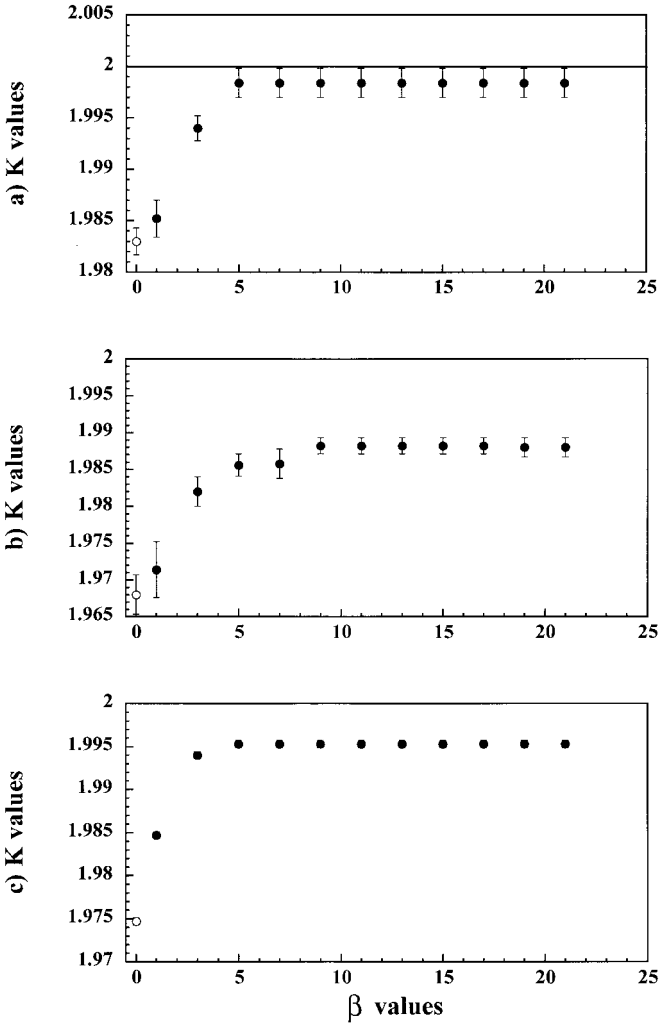

FIG. 1. Estimated K versus β parameter values for three different simulations. (a) Signal-to-noise ratio (SNR) = 103.4 ± 1.0, degree of diversity = (41.81 ± 0.42)%, normalization factor $R = 2$. (b) SNR = 103.7 ± 1.1, degree of diversity = (44.64 ± 0.36)%, $R = 2$. (c) SNR = 105.6 ± 1.6, degree of diversity = (48.89 ± 0.34)%, $R = 2$.

mean squared error (MSE) defined by the following equations, respectively (*9*):

$$\text{bias}(K) = E[K] - R$$
$$\text{var}(K) = E[(K - E[K])^2]$$
$$\text{MSE}(K) = E[(K - R)^2].$$

In Fig. 2, the bias is reported as a function of the degree of diversity for simulated spectra having a signal-to-noise ratio of 104 ± 1. As expected, the Lebesgue algorithm is the worst R estimator (K estimates out of range are not reported in Fig. 2) and can provide reasonable estimations only for very low degrees of diversity. In fact, only in this case does the arbitrary extension of the integration domain in Eq. [2] to the overall

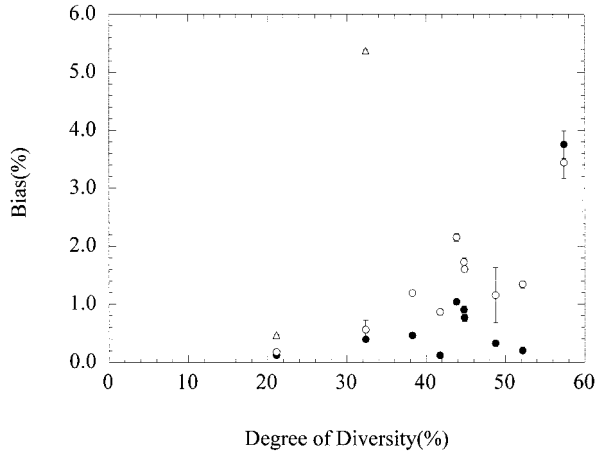

FIG. 2. Percentage absolute bias versus degree of diversity: (●) MaSNAl algorithm; (○) Dirac algorithm; (△) Lebesgue algorithm. For degree of diversity values greater than 30%, the Lebesgue values are out of the range of the figure and are not reported. Simulated spectra have mean SNR = 104.5 ± 0.6.

spectrum have negligible effects. On the other hand, both Dirac and MaSNAl algorithms display a low bias up to degrees of diversity of 50%. Nevertheless, MaSNAl presents a much slower variation and consistently lower values than the Dirac algorithm. The increment of the bias values for degree of diversity values greater than 50% is consistent with the validity limit of the MaSNAl algorithm.

In Fig. 3, the variance is reported as a function of the degree of diversity. In particular, in this figure it can be seen 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 estimated values around the true parameter value of the unbiased Dirac and MaSNAl algorithms.

Finally, the MaSNAl algorithm also yields better results than the Dirac algorithm for the mean squared error, which describes the interplay between the bias and the dispersion of estimated values around the true parameter value, reported as a function of the degree of diversity (results not shown).

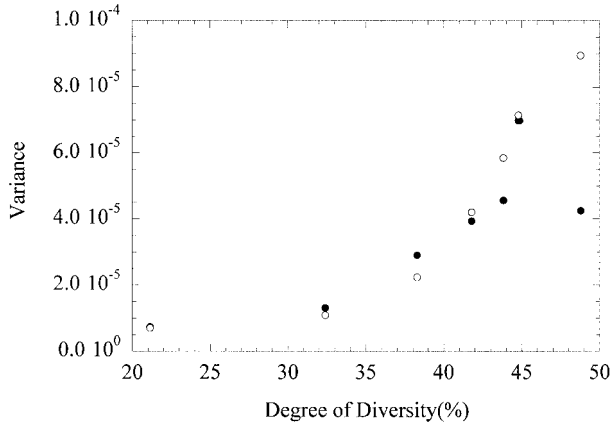

FIG. 3. Variance versus degree of diversity: (●) MaSNAl algorithm; (○) Dirac algorithm. Simulated spectra have mean SNR = 104.5 ± 0.6.

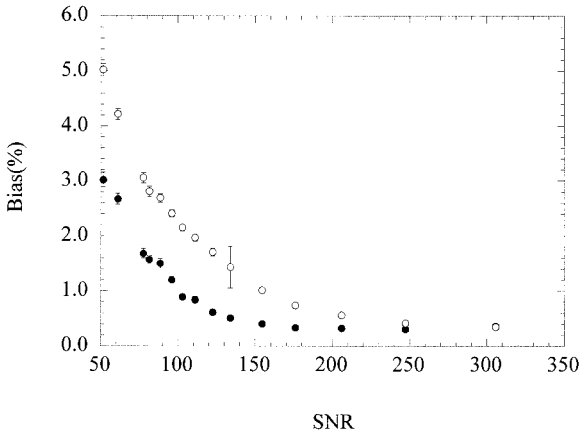

FIG. 4. Percentage absolute bias versus SNR: (●) MaSNAl algorithm; (○) Dirac algorithm. Degree of diversity = (43 ± 2)%.

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 noise of the spectra. The signal-to-noise ratio was defined by the equation

$$\mathrm{SNR} = \frac{\psi(\nu_0) - \bar{\psi}}{2 \times [\sum_{i=a1}^{a2} (\psi(\nu_i) - \bar{\psi})^2/N]^{1/2}}, \qquad [12]$$

where $\bar{\psi} \equiv [\sum_{i=a1}^{a2} \psi(\nu_i)]/(N + 1)$, $\psi(\nu_0)$ is the maximum peak height, $a1$ and $a2$ are the limits of the noise region with $N = a1 - a2$, and $\bar{\psi}$ is the DC level of the noise region. Figure 4 shows the bias as a function of the signal-to-noise ratio for the Dirac and MaSNAl algorithms. The two algorithms have comparable bias for high SNR values, while for lower values, the MaSNAl algorithm is less biased. In Fig. 5, the variances of the two algorithms are shown. Both the Dirac and the MaSNAl algorithms yield comparable results. In particular, the variance is constant for SNR value SNR ≥ 200. For lower SNR values, an exponential increase is observed due to the increased level of noise. Finally, the mean squared error versus SNR (results not shown) clearly indicates that dispersion of estimated values around the true parameter value is lower for the MaSNAl algorithm, especially at low SNR values.

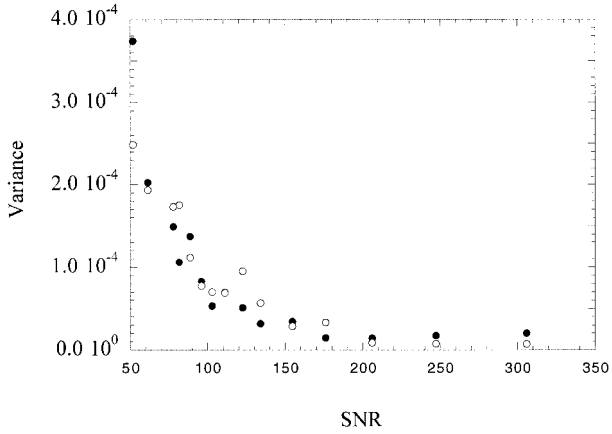

FIG. 5. Variance versus SNR: (●) MaSNAl algorithm; (○) Dirac algorithm. Degree of diversity = (43 ± 2)%.

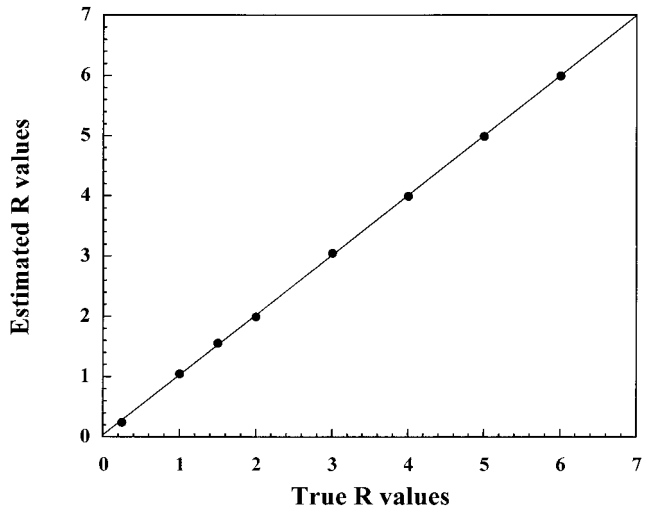

FIG. 6. Estimated R values versus true R values. Couples of spectra with SNR = 144 ± 7 and degree of diversity = (41 ± 5)%. Curve fit: $y = ax + b$ with $a = 0.99$, $b = 0.03$. Correlation ratio $r = 1$.

Dependence on R values. The last set of Monte Carlo simulations was designed to test the MaSNAl 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, in the limit of the algorithm validity (degree of diversity < 50% and SNR > 70), the MaSNAl algorithm is able to determine the normalization factor R of two NMR spectra with a bias of 2% at most.

QUANTITATIVE RELATIONSHIP BETWEEN TWO SPECTRA CONDUCTED BY THE MaSNAl ALGORITHM AND A MORE TRADITIONAL METHOD

In order to test the validity of the MaSNAl algorithm experimentally, two samples of known contents were compared. Both samples contained thyrotropin releasing factor (THR, Calbiochem, MW = 362.4), deutered 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 for 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 ^1H NMR spectra were obtained for each sample using a Bruker DPX digital spectrometer operating at 300 MHz. The experiments were carried out with a 90° flip angle pulse and 64 transients of 8-K data points corresponding to a $\pm$2097.3 Hz spectral window were accumulated.

Using the traditional method, that is, making a quantitative analysis using the TSP standard, the following results were obtained:

Sample	THR	CD_3OD residual peak
A	(38.7 $\pm$ 3.6) μmol/ml	(41.9 $\pm$ 3.9) μmol/ml
B	(20.2 $\pm$ 2.5) μmol/ml	(36.1 $\pm$ 4.6) μmol/ml.

[13]

If we wish to compute the percentage difference ($\Delta_{CD_3OD}(\%)$) of CD_3OD relative to THR for the sample A spectrum with respect to the sample B spectrum, the following equation can be used,

$$\Delta_{CD_3OD}(\%) = \frac{CD_3OD(A) - \frac{CD_3OD(B)}{THR(B)}THR(A)}{\frac{CD_3OD(B)}{THR(B)}THR(A)} \times 100, \quad [14]$$

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 (*11*); thus, the concentrations obtained by the CD_3OD residual peaks in Eq. [12] can be used directly in Eq. [13]. With this equation, using the concentrations obtained by TSP standard quantification (traditional method), a $\Delta_{CD_3OD}(\%) = (-39.4 \pm 8.7)\%$ difference was obtained, while from the known quantities of the added substances and using the same equation cited above, a $\Delta_{CD_3OD}(\%) = (-33.4 \pm 0.2)\%$ difference was expected.

The same pairs of spectra (A, B) were utilized to obtain the percentage difference $\Delta_{CD_3OD}(\%)$ of CD_3OD relative to THR by using the MaSNAl algorithm. With the algorithm, the spectra were normalized (i.e., the maximum numbers of spectral lines were made to superimpose; since THR contained the majority of spectral lines, it was superimposed in the normalized spectra). At this point, the percentage difference $\Delta_{CD_3OD}(\%)$ of CD_3OD relative to THR between the spectra of sample A, normalized with respect to the spectra of sample B, was obtained directly by comparing the areas of the spectral CD_3OD residual peaks. The value found was $\Delta_{CD_3OD}(\%) = (-34.2 \pm 1.3)\%$. As can be seen, both methods yelded percentage difference results which were consistent with the expected ones. However, the MaSNAl algorithm allowed us to obtain good results without the use of any standard and without quantifyng all the spectral lines, but rather by comparing the signals of interest in the normalized spectra.

BIOLOGICAL APPLICATION

The MaSNAl algorithm was applied to the normalization of NMR spectra of cell samples. In particular, the NMR spectra of control human K562 erythroleukemic tumor cells and of these same cells grown for 48 h on polylysine were compared.

The two spectra, acquired in the same experimental conditions, were analyzed by the MaSNAl algorithm and a normalization factor of $R = 0.91$ was found. The two normalized spectra are reported in Fig. 7. As can be seen, the spectra, after normalization, overlap in a significant number of spectral lines. This is confirmed by the self-consistency calculation, after which the degree of diversity was found to be 29%. In simulations, in fact, the exact value of R is known and thus the exact degree of diversity can be computed in order to decide the applicability of the method. In experimental cases, this test of applicability must be made by self-consistency; that is, it should be founded on an estimated value for R. The degree of diversity is then computed and checked to be certain that it is less than 50%. If this is the case, the estimated R value can be accepted for self-consistency. If the degree of diversity is greater than 50%, the estimated R value must be rejected.

In Fig. 8, the difference spectrum obtained from subtraction of the spectrum of controls from that of cells grown on polylysine 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 interaction with polylysine, can be obtained. In particular, intense signals are present at about 3.4 to 2.9 ppm (particularly at 3.24, 3.22, 3.21, and 3.03 ppm corresponding to glycerophosphatidylcholine, GPC; phosphatidylcholine, PC; choline; and creatine, respectively), at about 2.1 to 2.4 ppm (particularly at 2.34 and 2.09 ppm assigned to γ-glutamate and β-glutamate, respectively), at about 1.33 ppm corresponding to lactate, and at about 1.6 to 0.6 ppm (the lipid region). The resonances in this region can be assigned to the CH_2 and CH_3 groups of lipids. As is visible from the spectrum, control K562 cells contain a much smaller amount of PC with respect to polylysine-exposed cells and a larger amount of GPC, choline, and creatine. In addition, controls contain much less γ-glutamate and β-glutamate than treated cells, more lactate, fewer CH_2 lipids, and more CH_3 lipids. Thus, from these data, it appears that the MaSNAl algorithm presented can be adequately utilized for the comparison of NMR spectra of tumor cells.

CONCLUSIONS

In this paper, a new algorithm for the normalization of couples of NMR spectra, used to compare these spectra and to

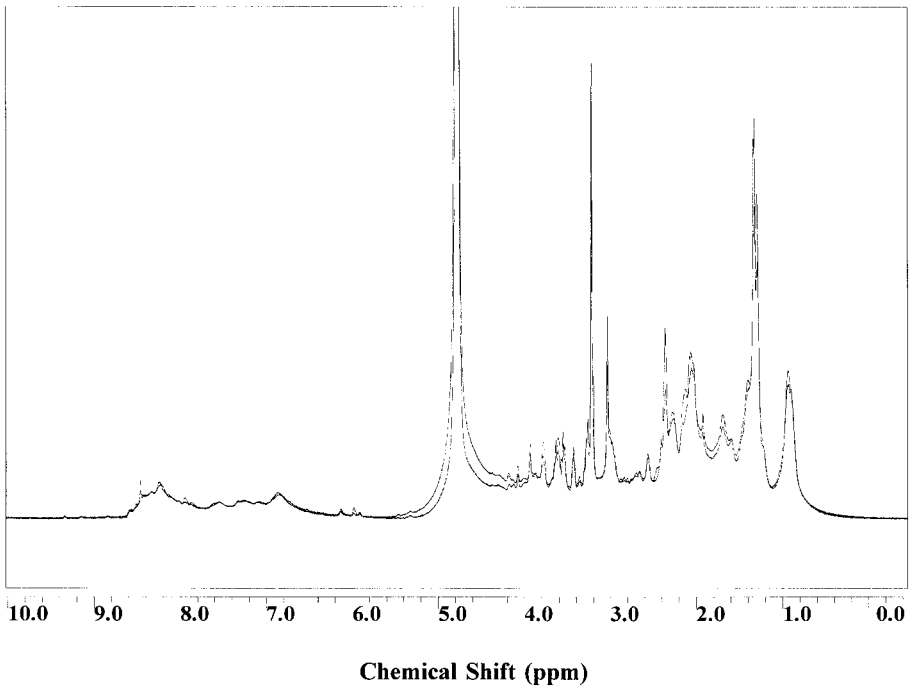

FIG. 7. Superimposed normalized spectra of K562 controls and of K562 cells grown on polylysine.

obtain relative quantitative information without the need of any standards, is presented. It consists of maximizing, by using a sign variable measure, the spectral regions in which spectral lines are proportional and evaluating the relative normalization factor. In this manner, normalization is accomplished by exploiting intrinsic sample properties and thus it considerably simplifies the measurement procedures. In fact, no addition of substances or particular manipulations of the samples are needed, thus reducing contamination risks. In addition, the method is very easy to manage and the normalization procedure requires only a few minutes. Furthermore, the method could be used in all situations in which the addition of a

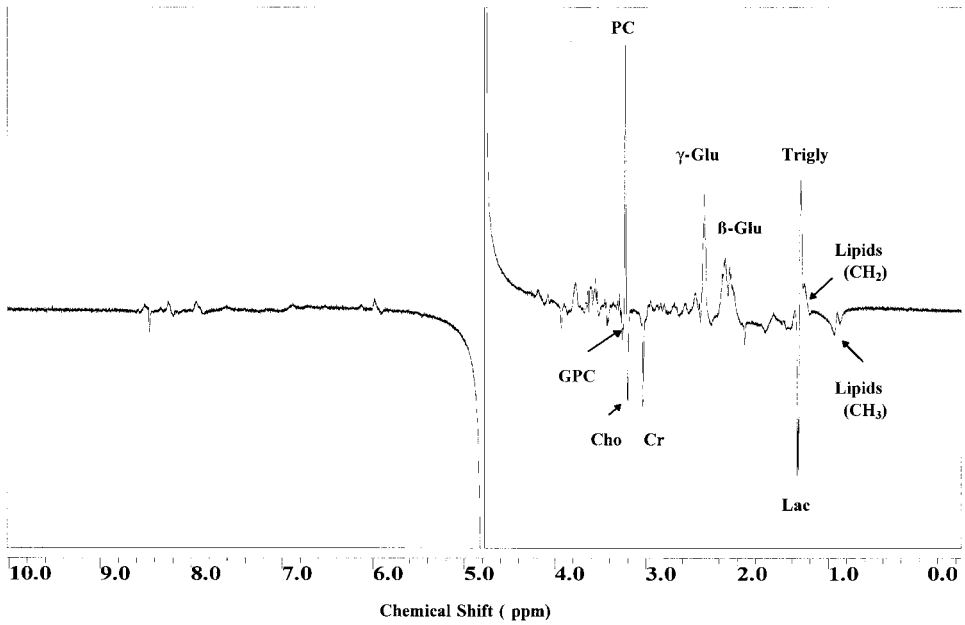

FIG. 8. Difference spectrum obtained from the subtraction of K562 control cells from the spectrum of cells grown on polylysine.

standard is not easy to manage, for example, in *in vivo* spectroscopy. Monte Carlo simulations show that, in the applicability limit of a degree of diversity less than 50%, and a signal-to-noise ratio greater than 70 the algorithm is consistent and presents very low bias and variance, thus giving an optimal estimation of the normalization factor. The algorithm was also tested by comparing two samples of known contents and results were better then those obtained by the traditional method based on the use of a standard. Finally, it is demonstrated that the algorithm can be applied to real spectra of cell samples, allowing the extraction of important biological information.

ACKNOWLEDGMENT

We thank Dr. Andrea Motta, Istituto per la Chimica Di Molecole di Interesse Biologico, CNR of Arcofelice, Naples, for acquiring the experimental spectra presented in the paper and for helpful discussion.

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 intercerebral glioma in rat brain: Assignment of resonances, *J. Neurochem.* **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. A. N. K. Kolmogorov and S. V. Fomin, "Elementi di teoria delle funzioni e di analisi funzionale" [Elementy teorii funktsij i funktsional'nogo analiza], Edizioni Mir, Moscow (1980).
6. M. Reed and B. Simon, "Methods of Modern Mathematical Physics. I. Functional Analysis," Academic Press, New York (1972).
7. F. Riesz and B. Sz. Nagy, "Functional Analysis," Ungar, New York (1965).
8. V. S. Vladimirov, "Le distribuzioni nella fisica matematica" (Obobščennye funktsii v matematičeskoj fisike), Edizioni MIR, Moscow (1981).
9. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Numerical Recipes in C: The Art of Scientific Computing," second ed., Cambridge Univ. Press, Cambridge, UK (1986).
10. M. K. Maple, "Digital Spectral Analysis with Applications," Prentice–Hall, New Jersey (1996).
11. A. E. Derome, "Modern NMR Tecniques for Chemistry Research," Pergamon, Elmsford, NY (1997).