Journal of Mathematical Chemistry, Vol. 40, No. 1, July 2006 (© 2006)
DOI: 10.1007/s10910-006-9123-8

Mathematical optimization of *in vivo* NMR chemistry through the fast Padé transform: potential relevance for early breast cancer detection by magnetic resonance spectroscopy

Dževad Belkić*

Department of Oncology and Pathology, Karolinska Institute, Building P-9, P.O. Box 260, Stockholm SE-171776, Sweden
E-mail: Dzevad.Belkic@ki.se

Karen Belkić

Department of Oncology and Pathology, Karolinska Institute, Building P-9, P.O. Box 260, Stockholm SE-171776, Sweden and Institute for Prevention Research, The University of Southern California School of Medicine, Los Angeles, CA, USA

Mathematical advances in signal processing through the fast Padé transform (FPT) can greatly improve the information extracted via *in vivo* nuclear magnetic resonance (NMR) chemistry. The FPT is a frequency-dependent, non-linear rational polynomial approximation of the exact Maclaurin series, which dramatically improves resolution and signal-to-noise ratio in a stable manner with robust error analysis and provides precise numerical data for all the peak parameters (position, height, width and phase) for every true resonance including those that are weak and/or overlapping. The concentrations of many of the chemical constituents of tissues can thereby be accurately determined. These advantages of the FPT are particularly germane for *in vivo* NMR detection and quantification of a number of molecular markers of breast cancer, such as phosphocholine, as well as lactate, which cannot be assessed using standard Fourier data analytical techniques applied to *in vivo* NMR in the clinical setting.

KEY WORDS: fast Padé transform, NMR chemistry, early cancer diagnostics, breast cancer

Abbreviations: CDP – cytosine diphosphate; CHESS – **CHE**mical **S**hift **S**elective; CI – confidence interval; FFT – fast Fourier transform; FPT – fast Padé transform; GPC – glycerophosphocholine; PA – Padé approximant; PC – phosphocholine; MR – magnetic resonance; MRI – magnetic resonance imaging; MRS – magnetic resonance spectroscopy; MRSI – magnetic resonance spectroscopic imaging; NMR – nuclear magnetic resonance; SD – standard deviation; SNR – signal to noise ratio; TE – echo time.

*Corresponding author.

1. The fast Padé transform for generic time signals: particular appropriateness for *in vivo* NMR chemistry

The fast Padé transform (FPT) is a frequency-dependent, non-linear rational polynomial approximation P_K/Q_K (diagonal) or P_{K-1}/Q_K (non-diagonal) of the exact Maclaurin series. The expansion coefficients in the latter series are provided by the raw time signal points $\{c_n\}$. In the literature this polynomial quotient, as a rational function of angular frequency ω, is known as the Padé approximant (PA). The PA has been alternatively termed the fast Padé transform for signal processing [1]. This is done to point out a special feature of the PA in signal processing, i.e. the possibility of obtaining a shape spectrum from a time signal via a non-parametric or shape estimation similar to the fast Fourier transform (FFT) [2]. This type of estimation is performed by evaluating the Padé spectrum P_{K-1}/Q_K or P_K/Q_K without searching for the spectral parameters at all, such as the complex-valued nodal frequencies $\{\omega_k\}$ and the corresponding amplitudes $\{d_k\}$. This is in contradistinction, e.g. to Hankel–Lanczos Singular Value Decomposition [3], which computes the shape spectrum exclusively through obtaining the peak parameters $\{\omega_k, d_k\}$ first. The FPT computes the ansatz spectrum $F(z^{-1}) = (1/N)\sum_{n=0}^{N-1} c_n z^{-n}$, which is a truncated version of the mentioned Maclaurin series, via the *unique* ratio of two polynomials, e.g. in the non-diagonal form $F(z^{-1}) = P_{K-1}(z^{-1})/Q_K(z^{-1})$ at any real or complex frequency ω as follows:

$$F(z^{-1}) = \frac{1}{N}\sum_{n=0}^{N-1} c_n z^{-n}, \qquad F(z^{-1}) \approx \frac{P_{K-1}(z^{-1})}{Q_K(z^{-1})}. \tag{1}$$

Here, z is a complex harmonic variable, $z = \mathrm{e}^{i\omega\tau}$, where τ is the sampling time. The real and imaginary part of the complex number u will be denoted by $\mathrm{Re}(u)$ and $\mathrm{Im}(u)$, respectively.

1.1. Resolution and signal to noise ratio enhancements by the FPT compared to the FFT

The FFT is a low-resolution spectral/image estimator, which provides a shape spectrum from pre-assigned frequencies whose minimal separation is determined solely by the acquisition time, $T = N\tau$. In other words, the FFT spectrum is defined only at the Fourier grid points, $2\pi k/T\,(0 \le k \le N-1)$ [1]. The strategy applied in attempts to improve resolution has been to increase the acquisition time and thereby decrease the distance $1/T$ between the grid points. This, however, is often not an adequate solution, because, e.g., *in vivo* nuclear magnetic resonance (NMR) signals become heavily corrupted with background noise at longer acquisition times. Since these time signals decay exponentially, the larger

signal intensities are observed early in the recording. It is therefore advantageous to encode the time signal as rapidly as possible, i.e. to avoid long acquisition times at which mainly noise will be recorded. Thus, there are two mutually exclusive requirements, and as a result, within the FFT, attempts to improve resolution lead to a worsening of the signal-to-noise ratio (SNR). The FFT is a linear transform, and, as such, imports noise as intact from the measured time domain data to the theoretically analyzed frequency domain, further contributing to poor SNR [1]. In essence, the FFT lacks an extrapolation property. With the FFT it is a usual practice to add extra data points that are zeros, in order to lengthen the signal. Namely, instead of recording a signal with nothing in it but noise, the signal is artificially lengthened by adding zeros. The first doubling ($N \to 2N$) leads to a sinc-interpolation of the original FFT [4]. Specifically, with one zero filling the FFT of the doubled exponentially decaying time signal yields twice as many Fourier coefficients as the FFT of the original time signal, with the new values located between the original Fourier coefficients [4]. Further addition of zeros to the exponentially decaying time signal beyond $2N$ does not lead to any more interpolation relative to the first doubling of N.

In contrast, the FPT is a truly powerful interpolator/extrapolator. Due to the presence of the implicit polynomial inversion via Q_K^{-1} in $P_{K-1}(z^{-1})/Q_K(z^{-1})$, inference is gained from a non-measured infinite number of signal points by using only the available finite set $\{c_n\}(0 \leq n \leq N-1)$. The FPT does not use the fixed Fourier mesh $\tilde{\omega}_k = 2\pi k/T (k = 0, \ldots, N-1)$ in the frequency domain, and therefore, can be computed at any frequency ω. Resolution in the FPT is not pre-determined by T. Rather, the resolution in the FPT is the average distance $\Delta\omega_{\text{ave}}$ between adjacent frequencies in a selected range. In a given frequency window, one usually has $\Delta\omega_{\text{ave}} < \Delta\tilde{\omega}_{\min}$, so that the FPT has a better resolving power, relative to the resolution $\Delta\tilde{\omega}_{\min} = 2\pi/T$ in the FFT. Thus, a major advantage for analysis of MR signals is that the Padé theory performs both extrapolation and interpolation, which translate directly into enhanced resolution.

The FFT is linear because the coefficients $\{a_{n,k}^{\text{FFT}}\}$ of the transformation from the time to the frequency domain are *independent* of the signal points:

$$F_k = \frac{1}{N} \sum_{n=0}^{N-1} a_{n,k}^{\text{FFT}} c_n, \qquad a_{n,k}^{\text{FFT}} = \text{constants} = \mathrm{e}^{2i\pi nk/N}. \tag{2}$$

As opposed to the FFT whose linearity preserves noise fully from the time signal, the FPT is a non-linear mapping, such that its coefficients $a_{n,k}^{\text{FPT}}$ are *dependent* upon the time signal points $\{c_n\}$. The non-linearity of the FPT effectively permits noise suppression. Furthermore, as is clear from equation (2), the FFT has a linear convergence ($1/N$) with increased signal length N. In contrast, the convergence of FPT is quadratic ($\sim 1/N^2$) or better [1].

Rapid convergence together with the enhanced resolution effectively result in markedly improved information content extracted by the FPT from *in vivo* NMR signals encoded at short acquisition times. Convergence with the FPT is not only rapid, but also extremely stable with robust error analysis. This has been confirmed in Refs. [1,5,6] by detailed comparisons of the FPT and FFT from high-field clinical *in vivo* ^{1}H NMR data. The fidelity of the FPT has also been internally cross-validated via two equivalent variants, the FPT$^{(+)}$ and FPT$^{(-)}$ with the initial convergence-regions inside and outside the unit-circle, respectively. Due to the uniqueness of the FPT, all the physical resonances from the two variants have been demonstrated to coincide after convergence has been reached in both variants [7].

1.2. Quantification: the FPT computes all peak parameters and entails no fitting

The FFT is exclusively a non-parametric estimator, providing only the shape of spectral structures, but not their quantifications. The peak parameters are subsequently extracted in post-processing by fitting the obtained structures to a sum of Lorentzians and/or Gaussians, or the so-called Voigt profiles. Besides the fact that fitting is non-unique, much information contained in the signal is not obtained accurately, such as the actual position, width, height, and phase of resonances [1]. In Refs. [1,2,8] elaboration and validation is provided for the computational methods by which the FPT yields quantitative spectral parameters. This is done without fitting and, furthermore, the solution is unique. To extract the peak parameters, one solves the equation,

$$Q_K(z^{-1}) = 0, \tag{3}$$

which is known as the characteristic equation or the secular equation of the data (Hankel) matrix. Every polynomial of degree n has n roots. Thus, equation (3) has K roots z_k^{-1} $(1 \leq k \leq K)$, where,

$$z_k = \mathrm{e}^{i\omega_k \tau}, \qquad \omega_k = -\frac{i}{\tau} \ln(z_k). \tag{4}$$

Once the K roots $\{z_k^{-1}\}$ of $Q_K(z^{-1})$ are found, the corresponding amplitudes $\{d_k\}$ are computed from the following explicit expression:

$$d_k = \frac{P_{K-1}(z_k^{-1})}{Q'_K(z_k^{-1})}, \tag{5}$$

where $Q'_K(z^{-1})$ is the first derivative of the denominator polynomial $Q_K(z^{-1})$,

$$Q'_K(z^{-1}) = \frac{\mathrm{d}Q_K(z^{-1})}{\mathrm{d}z^{-1}}. \tag{6}$$

The parametric complex Lorentzian spectrum is obtained from the Heaviside partial fraction expansion:

$$\frac{P_{K-1}(z^{-1})}{Q_K(z^{-1})} = \sum_{k=1}^{K} \frac{d_k}{z^{-1} - z^{-1}}. \tag{7}$$

Thus, the FPT arrives at the spectral parameters $\{\omega_k, d_k\}$ with minimal computational effort and maximal accuracy. Here, the real, $\mathrm{Re}(\omega_k)$, and the imaginary, $\mathrm{Im}(\omega_k)$, part of ω_k are the position and the width of the kth peak, whereas $|d_k|/\mathrm{Im}(\omega_k)$ and $\arg(d_k)$ are the corresponding height and phase, respectively. Specifically, the FPT can find all the peak parameters of every physical metabolite *without ever using the Fourier spectrum*, or any spectrum at all, for that matter. The Padé spectrum $\sum_{k=1}^{K} d_k \big/ (z^{-1} - z_k^{-1})$ can be subsequently constructed for visualization purposes in any of the desired modes (absorption, magnitude, etc.). We have developed both MATLAB as well as C++ user-friendly software based upon the FPT. This software automatically and efficiently performs the entire quantification by using only raw time signals [1,2,6–8].

1.2.1. Overlapping resonances (degenerate roots) can be quantified by the FPT

The above formulae are valid for non-degenerate spectra, i.e. for all distinct roots $\{z_k^{-1}\}$ of equation (3). When some of these roots coincide with each other (the so-called degenerate roots – leading to overlapping resonances), the above formulae should be modified. For example, if the kth root z_k^{-1} of $Q_K(z^{-1})$ has $M_k \leq K$ multiplicity (i.e. it is repeated M_k times), then the Padé spectrum from equation (7) will acquire the form:

$$\frac{P_{K-1}(z^{-1})}{Q_K(z^{-1})} = \sum_{k=1}^{J} \sum_{m_k=1}^{M_k} \frac{d_{k,m_k}}{(z^{-1} - z_k^{-1})^{m_k}}, \quad d_{k,1} = d_k \tag{8}$$

where $M_1 + M_2 + \cdots + M_J = K$. Here d_{k,m_k} are the new amplitudes that generalize equation (5) according to:

$$d_{k,m_k} = \frac{P_{K-1}(z_k^{-1})}{Q_K^{(m_k)}(z_k^{-1})}, \tag{9}$$

where $Q_K^{(m)}(z^{-1})$ is the mth derivative of the denominator polynomial $Q_K(z^{-1})$,

$$Q_K^{(m)}(z^{-1}) = \left(\frac{\mathrm{d}}{\mathrm{d}z^{-1}}\right)^m Q_K(z^{-1}). \tag{10}$$

Also for the degenerate roots of equation (2), for the given N values of the c_n's, one always obtains the unique quotient P_{K-1}/Q_K. The time signal points

$\{c_n\}$ associated with the *physical* Lorentzian spectrum from equation (6) with non-degenerate (distinct) roots $\{z_k^{-1}\}$ are given by:

$$c_n = \sum_{k=1}^{K} d_k \mathrm{e}^{in\tau\omega_k}, \qquad \mathrm{Im}(\omega_k) > 0. \tag{11}$$

On the other hand, the c_n's corresponding to the *physical* non-Lorentzian spectrum with degenerate (multiple) roots $\{z_k^{-1}\}$ read as:

$$c_n = \sum_{k=1}^{J} \sum_{m_k=1}^{M_k} d_{k,m_k} (n\tau)^{m_k-1} \mathrm{e}^{in\tau\omega_k}, \qquad \mathrm{Im}(\omega_k) > 0. \tag{12}$$

All fitting techniques in the frequency domain use the given Fourier spectrum for extraction of peak parameters for the metabolites of interest. As mentioned, the FPT avoids fitting altogether, and accomplishes accurate quantification by extracting the parameters of all the physical metabolites directly from the encoded raw signal $\{c_n\}$ given by equations (11) and (12) that correspond to the Lorentzian and non-Lorentzian spectra (7) and (8), respectively.

Every parametric estimator of, e.g. exponentially damped time signals yields, by definition, the Lorentzian spectrum of the type $\sum_{k=1}^{K} d_k/(z^{-1} - z_k^{-1})$. However, this latter sum is nothing but the Padé polynomial quotient, as indicated in equation (7). Hence, the FPT is optimally suited to spectrally analyze exponentially decaying time signals. These signals are attenuated complex exponentials (harmonics) with either constant or varying amplitudes according to (11) or (12) that yield the Lorentzian (7) or non-Lorentzian (8) spectrum, respectively. Thus, the FPT retrieves with fidelity overlapping or hidden resonances, including those that may be disguised as noise [1]. In realistic synthesized models, it has been illustrated that the FPT successfully identifies overlapping peaks that are entirely missed by the FFT [8].

1.3. The FPT applied to a wide variety of time signals

Regardless of the specific discipline, spectra are generally of a Lorentzian or a non-Lorentzian type. The former non-degenerate spectra are engendered from time signals with harmonic components that contain only distinct frequencies with constant amplitudes. The latter, degenerate spectra are generated from time signals that have distinct as well as confluent frequencies with stationary and time-varying amplitudes. The FPT can handle both types of time signals, which, by definition, yield the described Lorentzian and non-Lorentzian spectra. This has been clearly demonstrated in numerous applications in chemistry and physics [9–14] as well as in medicine [1, 5, 15–17]. The above-described time

signals have been encoded in a large number of chemistry and physics experiments using measuring techniques such as Infrared Spectroscopy, Ion Cyclotron Resonance Mass Spectroscopy, as well as NMR, etc. The FPT has been effectively applied to many such signals, with quantification of a large number of resonances via extraction of peak parameters (position, height, width, and phase) [9–13]. Prior to these applications on encoded data, extensive testing was performed on simulated signals of the same types as those measured. This is a vital step, since for synthesized signals it is possible to select the number of harmonics and to prescribe the frequencies and amplitudes to any level of accuracy. Using this strategy, the FPT has been shown to retrieve all four parameters for each resonance within machine accuracy for noiseless signals (see also [8]). When random Gaussian noise of a standard deviation (SD) comparable to that found in measured data is added, once again the FPT extracts all the simulated peak parameters to within several decimal places of accuracy, that is to a level of accuracy greater than what is actually required in practice [9, 11]. Testing on simulated data that are similar to realistically measured signals establishes the fidelity of the FPT for controllable input data and is therefore of crucial importance.

Clinical magnetic resonance spectroscopy (MRS) is a synonym for *in vivo* NMR. Thus, MRS signals are of the same type as those encoded by NMR in chemistry. It follows, thereby, that the methods established in NMR chemistry should be applicable to MRS. The recently published book and series of papers using the FPT applied to *in vivo* time signals encoded via MRS [1, 5, 15, and 16] indeed confirm this to be the case. *In vivo* NMR signals, irrespective of the tissue of origin, are of both Lorentzian and non-Lorentzian types. The free induction decay curves are obviously time signals that display exponentially damped sinusoidal oscillations. Therefore, methods validated in one source of *in vivo* NMR time signals (e.g. brain) can be justifiably expected to be applicable to the corresponding data from other organs (i.e. breast, prostate, etc.). We are now performing this type of validation, the results of which will be soon forthcoming. The current achievement using *in vivo* NMR signals encoded using e.g. 9.4 T and ultra-short echo time in experimental animals has been to estimate the concentrations of some 18 metabolites [18]. However, with clinical scanners at 1.5 T, in practice, only about five metabolite concentrations are estimated by means of post-processing fitting from the spectra processed using the FFT. Clearly, even the aim of quantifying several dozen metabolites is very modest, in light of the fact that tissues contain far more resonances that could be detected and quantified. From the corresponding *in vitro* findings, it has been repeatedly shown that there is a wealth of clinically relevant information, particularly for distinguishing malignant and normal tissues, as well as for gaining deeper insights into molecular mechanisms. Moreover, *in vitro* NMR has often demonstrated that the most important information for detecting malignant lesions is found in closely overlapping resonances, some of which decay rapidly and therefore can only be

detected at short acquisition times. This is the case for breast tissue, as well be discussed later in the next section.

2. NMR chemistry for early breast cancer diagnostics: current state of the art

2.1. Importance of magnetic resonance-based methods for early breast cancer diagnostics

There is increasing interest in magnetic resonance (MR)-based modalities for early breast cancer detection and screening, particularly among younger women at high risk for this malignancy [19]. A key advantage of MR-based diagnostics is the lack of exposure to ionizing radiation to breast, which is a radiosensitive tissue. This is especially critical in light of the heightened radiosensitivity for women with genetic risk for breast cancer, i.e. with *BRCA* germline mutations, Li Fraumeni syndrome (p53 tumor suppressor gene mutations), as well as those who are heterozygous for ataxia–telangiectasia [20, 21].

Contrast-enhanced magnetic resonance imaging (MRI) with fat-suppressed T_1-weighting has a sensitivity between 95% and 100% for detection of breast cancer [22]. It is particularly valuable for identifying malignancy in dense breasts, commonly seen in younger women (as well as in association with combined hormone replacement therapy) among whom cancers, unless calcified, are difficult to perceive using mammography [23, 24]. Breast MRI is considered superior to mammography for detecting multifocal or multi-centric cancers, as well as tumor recurrence and response to chemotherapy. However, false negative findings using MRI have been reported for small tumors, particularly if they do not selectively take up the contrast agent. Moreover, MRI cannot reveal microcalcifications [25], and occasionally misses invasive ductal and lobular carcinomas, although non-detection with MRI is more common with *in situ* ductal carcinoma [26]. The major problem with MRI, however, is that despite excellent spatial resolution and generally superior sensitivity, it often has limited specificity, thus sharing with mammography a high-false positive rate (approximately 50%) (reports range from 37 to 97%) [27, 28]. In a recent prospective study among 1909 women with a genetic or familial predisposition to breast cancer, MRI showed better sensitivity for breast cancer, but lower specificity than mammography, i.e. MRI generated more uncertain findings, requiring follow-up or additional investigations [29]. Intensive surveillance programs with a large number of false positive findings[1] may impact unfavorably upon quality of life [30, 31]. Thus, questions remain about the appropriateness of breast MRI as a screening

[1] False positive is defined as a finding which appears to indicate the presence of a disease, when the disease is not actually present. This occurs with diagnostic methods that have poor specificity. False negative is defined as apparent absence of disease, when the disease is actually present. This occurs with diagnostic methods that have poor sensitivity.

tool in asymptomatic, high-risk patients, for whom the need to improve specificity has been particularly underscored [22].

2.2. In vivo NMR chemistry findings in breast cancer

2.2.1. Total choline

While MRI yields anatomical/morphological images, *in vivo* NMR provides complementary biochemical and physiologic information in the form of spectra. ^{1}H MRS based upon the presence or absence of a composite choline resonance increases the specificity of diagnosing breast cancer [27]. In a meta-analysis of five clinical studies using *in vivo* one-dimensional (1D) single-voxel ^{1}H MRS to evaluate 100 malignant and 53 benign breast lesions, sensitivity, and specificity of 83% (95% Confidence intervals (CI) = 73–89%) and 85% (95% CI = 71–93%) were achieved, respectively. Better diagnostic accuracy was attained among the 20 women up to age 40. In these younger women, all 11 malignancies were correctly diagnosed. A more recent single-voxel ^{1}H MRS study among 30 women with positive mammogram and rapid contrast enhancement on breast MRI similarly showed that the presence of choline (with SNR $\geq$ 2) improved the specificity for breast cancer (to 87.5%), with three false positive results [32]. In another recent single-voxel ^{1}H MRS study [33] using a very long-echo time (TE) of 288 ms (plus inversion recovery sequence to attenuate the fat signal, as well as water suppression by chemical selective saturation and improved shimming of the static magnetic field), the presence or absence of choline accurately detected 19/19 breast cancers and 16/16 benign breast lesions. Among the patients with cancer, the choline SNR varied from 2.4 to 12.7, with a mean value of 5.4.

Molecular imaging can be accomplished by combining MRI and MRS to yield magnetic resonance spectroscopic imaging (MRSI). Rather than selecting a single voxel from three orthogonal slices to encompass a specific volume as in MRS, a spectrum in MRSI is obtained at each point of selected grids that can be of various sizes. Thereby, full volumetric coverage can be achieved with MRSI. The advantages of MRSI for breast cancer diagnostics, relative to MRS, have been emphasized [31]. A single voxel of tissue pre-selected via MRS might not be sufficiently representative of the malignant process. However, MRSI can help assess, e.g. the degree of micro-invasion of surrounding healthy tissue, aid in determining tumor grade and evaluating the success of therapy. In the first reported investigation [34] of breast cancer using MRSI,[2] the choline SNR yielded a sensitivity of 87% and specificity of 85% among eight patients with malignant lesions and seven with benign lesions. For an additional three patients, MRSI was a technical failure. The mean choline SNR among the

[2]This MRSI investigation [19] was carried out at 1.5 T, shimming was performed to optimize field homogeneity, three sequential **CHE**mical **S**hift **S**elective (CHESS) pulses were used for water suppression, lipid signals were attenuated with inversion pulse and TE = 272 ms was selected.

patients with breast cancer was 6.2 ± 2.1 compared to 2.4 ± 0.7 among those with benign breast pathology. The distinguishing cut-point for choline SNR was 4.1, with the highest benign value being 4, and the lowest malignant value being 4.1. The authors from Ref. [34] note that unambiguous detection of the choline resonance was difficult at times, due to overlap with sidebands of residual water or lipid resonances, and emphasized the need for improved data processing methods to quantify choline in breast lesions.

Using a high-static magnetic field (4 T) and optimized surface coils with single voxel *in vivo* ^{1}H MRS, attempts were made to quantify total choline by an algorithm that fitted one peak at a time over a narrow frequency band [35]. The average of spectra at different TE values was used to try to separate non-coupled metabolite resonances from the lipid-induced sidebands [36]. This method was applied to 500 spectra from patients with breast cancer, with benign breast lesions and with normal breast tissue. Total choline was significantly higher in the malignant compared to benign lesions, but there was also overlap in the respective ranges. Moreover, total choline was undetectable in several breast cancers, and elevated in some benign lesions, and even in a few normal breasts. The total choline amplitude appeared to be influenced by the lipid content of the voxel, either through baseline artifacts not suppressed by TE averaging or by a true resonance at 3.25 ppm (parts per million). In smaller voxels, the fitting error was also noted to be greater [35].

2.2.2. Water to fat ratios in breast cancer

The unsuppressed water to lipid ratio has also been compared in breast cancer and normal mammary tissue with *in vivo* ^{1}H MRS. The ratio of water at 4.7 ppm to the major fat peak at 1.3 ppm at TE = 135 ms was significantly higher at breast voxels containing cancer compared to contralateral unaffected breasts of 17 patients, as well as relative to 14 healthy referents [37]. However, there was a very large SD of the malignant ratios (6.0 ± 6.9) suggesting a substantial overlap with the normal values (0.36 ± 0.25). When overlapping metabolites cannot be resolved in 1D, then two-dimensional (2D) MRS has been applied in attempts to separate the resonances using cross-correlation plots. Results applying *in vivo* 2D MRS also revealed a significantly higher diagonal peak volume of unsuppressed water relative to fat at 1.4 ppm among 21 patients with invasive breast carcinoma compared to 14 healthy referents with fatty breasts, but no ranges were given [38]. The diagonal peak volume ratios of water to methyl fat (0.9 ppm) and to olefinic fat (5.4 ppm) were also significantly higher in the cancerous tissue using 2D MRS, as were the water to cross peak volumes of unsaturated fatty acid at the 2D frequency locations (2.1 ppm, 5.4 ppm) as well as (2.9 ppm, 5.4 ppm) and triglyceride fat at (4.3 ppm, 5.3 ppm) [38]. The high water to fat ratios seen in breast cancers are considered to reflect the presence of edema, which correlates clinically with lymphatic blockage and involvement of subdermal

lymphatics by the cancer. Some (but not all) of the patients followed by *in vivo* 1D MRS showed a drop in this ratio after chemotherapy; this response appeared to correlate with response to the neoadjuvant therapy [37].

2.3. In vitro NMR chemistry findings in breast cancer

2.3.1. Individual chemical metabolites

As noted, the high resolution of *in vitro* NMR applied to extracted specimens can provide a greater insight into the metabolic activity of cancerous breast tissue. Resolution enhancement with *in vitro* NMR is achieved by very strong magnetic fields (of the order of 10 T) applied to excised tissue specimens. Analysis of excised breast cancers reveals that the composite choline peak contains a number of water-soluble metabolites such as phosphocholine (PC), glycerophosphocholine (GPC), betaine, analogous compounds containing the ethanolamine head group and taurine, as well as choline itself [39]. Using tracer kinetics and ^{13}C NMR and ^{31}P NMR to examine the biochemical mechanisms underlying the high levels of water-soluble choline metabolites seen in breast cancer, two non-intersecting pathways were identified: phosphorylation and oxidation of choline, that were augmented with malignant transformation of mammary cells, with increased synthesis of PC and betaine, and suppression of choline-derived ether lipids [39].

Gribbestad et al. [40] conducted an *in vitro* ^{1}H NMR study[3] comparing 14 extracts of malignant breast tissue and one fibroadenoma to non-involved breast from the same group of patients. We subsequently performed logistic regression analysis of these data to ascertain the sensitivity and specificity of individual metabolite concentrations for identifying breast cancer [31, 41]. We found that only lactate showed 100% diagnostic accuracy both with and without inclusion of the fibroadenoma. The diagnostic accuracy of total choline[4] was marginally lower than several of the individual metabolites, including some of its own constituents, as shown in table 1. Paired analysis ("*t*"-test) revealed a significant difference in all metabolite concentrations when comparing non-infiltrated and malignant breast tissue (always higher in the latter). The findings for lactate are illustrated in figure 1.

In the study from Ref. [40] GPC, PC, phosphoethanolamine, total choline, and lactate were elevated, as well, in the fibroadenoma compared to the non-infiltrated tissue of the same patient. Moreover, most of the calculated metabolite concentrations were at least one SD greater than the mean values for normal breast tissue. In contrast, the calculated concentration of myoinositol

[3]Gribbestad et al. [40] performed their study using a BRUKER Avance DRX600 spectrometer at 14.1 T.

[4]The composite or total choline peak is comprised of choline (3.21 ppm), PC (3.22 ppm), and GPC (3.23 ppm).

Table 1
Individual metabolite concentrations with superior diagnostic accuracy for identifying breast cancer compared to total choline – data from Gribbestad et al. [40], logistic regression calculations from [31, 41]. The non-malignant samples include 1 fibroadenoma plus 12 samples from normal breast tissue. The chemical shift is denoted in ppm (parts per million). Specificity is denoted by Spec and Sensitivity by Sens.

Normal Breast versus Breast Cancer				Non-malignant Breast versus Breast Cancer			
	Predicted Result				Predicted Result		
Observed Result	Normal Breast	Breast Cancer	(% correct)	Observed Result	Non-Cancer	Breast Cancer	(% Correct)
Lactate (1.33 ppm)							
Normal breast	12	0	100% (Spec)	Non-cancer	13	0	100% (Spec)
Breast cancer	0	14	100% (Sens)	Breast cancer	0	14	100% (Sens)
Alanine (1.47 ppm)							
Normal breast	12	0	100% (Spec)	Non-cancer	13	0	100% (Spec)
Breast cancer	1	13	92.9% (Sens)	Breast cancer	1	13	92.9% (Sens)
Choline (3.21 ppm)							
Normal breast	12	0	100% (Spec)	Non-cancer	13	0	100% (Spec)
Breast cancer	0	14	100% (Sens)	Breast cancer	2	12	85.7% (Sens)
Phosphocholine (3.22 ppm)							
Normal breast	12	0	100% (Spec)	Non-cancer	13	0	100% (Spec)
Breast cancer	1	13	92.9% (Sens)	Breast cancer	1	13	92.9% (Sens)
Phosphoethanolamine (3.22 ppm)							
Normal breast	12	0	100% (Spec)	Non-cancer	13	0	100% (Spec)
Breast cancer	1	13	92.9% (Sens)	Breast cancer	1	13	92.9% (Sens)
Taurine (3.27 ppm)							
Normal breast	12	0	100% (Spec)	Non-cancer	13	0	100% (Spec)
Breast cancer	1	13	92.9% (Sens)	Breast cancer	1	13	92.9% (Sens)
Total choline (Choline + Phosphocholine + Glycerophosphocholine)							
Normal breast	11	1	91.7% (Spec)	Non-Cancer	13	0	100% (Spec)
Breast cancer	1	13	92.9% (Sens)	Breast Cancer	1	13	92.9% (Sens)

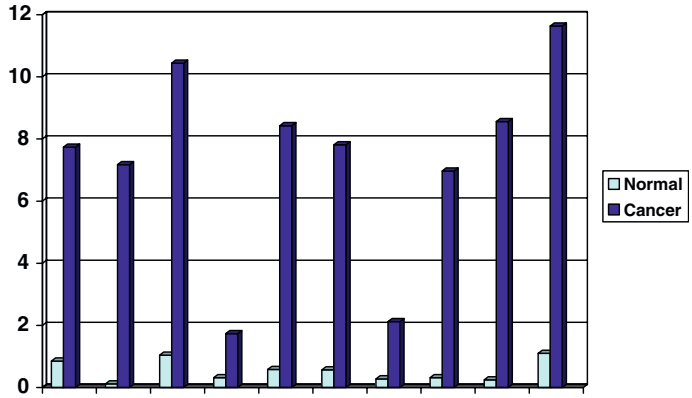

Figure 1. Case-by-case analysis of calculated lactate concentrations in normal versus cancerous breast tissue (from data by Gribbestad et al. [40]).

Along the abscissa are 10 individual cases, each separated by a tick mark. Calculated lactate concentrations are shown along the ordinate, in μmol/g wet weight (ww). The first value shown for an individual case is the calculated lactate concentration in non-involved breast tissue (denoted as Normal). The second value shown for an individual case is the calculated lactate concentration in the malignant breast lesion (denoted as Cancer). These data are shown only for those cases in which both concentrations were available. In two cases, there were two malignant sampled concentrations; the mean of these two values is shown in each case.

was nearly the same (0.465 and 0.448 μmol/g ww) for the fibroadenoma and for the non-infiltrated tissue, respectively, of the same patient and showed the smallest departure from the mean for normal breast tissue (+0.52 μmol/g ww of the SD) [31].

On the basis of these data from a fairly small sample (with substantial missing data for a few metabolite concentrations in malignant tissues), definitive conclusions cannot be drawn about which metabolites are optimal for detecting breast cancer and distinguishing this from normal tissue or benign lesions. Nevertheless, several metabolites (most notably lactate) showed promise with respect to diagnostic accuracy. On the other hand, total choline, upon which most *in vivo* ^{1}H MRS diagnoses are customarily based, had marginally lower sensitivity and specificity than several other metabolites. Myoinositol also provided some insight that could be helpful for identifying benign breast lesions. Viewed together, these analyses confirm that a rich "window" of information is provided by *in vitro* ^{1}H NMR study of metabolite concentrations in malignant versus non-cancerous breast tissue.

A more recent study [42] using *in vitro* 2D MRS to compare 11 involved and 12 uninvolved lymph nodes from patients with breast cancer provide some corroboration of the above findings from Refs. [31, 41]. Namely, the concentrations of PC and GPC were significantly higher in involved compared to non-involved nodes. This was attributed to increased membrane synthesis in cancer

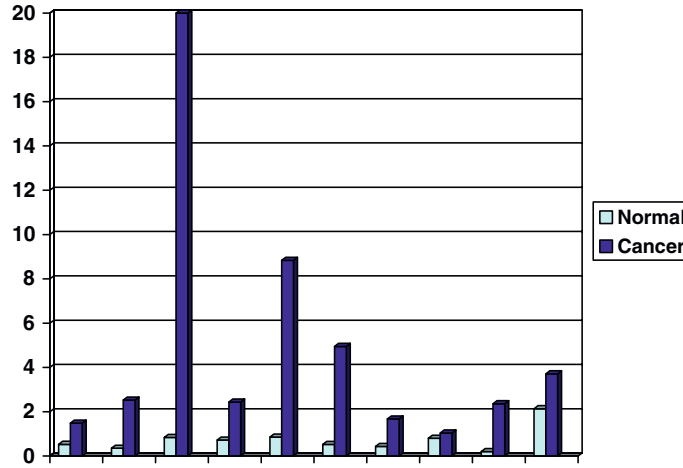

Figure 2. Case-by-case analysis of the ratio of the calculated phosphocholine (PC) concentration to calculated glycerophosphocholine (GPC) concentration in normal versus cancerous breast tissue (from data by Gribbestad et al. [40]).

Along the abscissa are 10 individual cases, each separated by a tick mark. The ratios of the calculated PC to GPC concentrations are shown along the ordinate, in dimensionless units. The first value shown for an individual case is the ratio of calculated PC to GPC concentrations in non-involved breast tissue (denoted as Normal). The second value shown for an individual case is the ratio of calculated PC to GPC concentrations in the malignant breast lesion (denoted as Cancer). These data are shown only for those cases in which both concentrations were available. In two cases, there were two malignant sampled concentrations; the mean of these two values is shown in each case.

cells, suggesting that metastatic breast cancer cells were present in the lymph nodes. There was also a highly significant difference ($p = 0.0001$) between the lactate concentrations in involved and non-involved nodes [42]. The elevated lactate reflects the presence of cancer cells whose energy source is from the anaerobic glycolytic pathway. Animal models of breast cancer also support the importance of assessing the rate of glycolysis and lactate clearance with respect to the diagnosis and prognosis of breast cancer [43]. Thus far, however, neither 1D nor 2D clinical *in vivo* ^{1}H NMR has included lactate as a metabolic marker of breast cancer.

2.3.2. Relationships of the chemical metabolites

In our analyses of the data of Ref. [40] many of the metabolite concentrations in the malignant tissues were significantly correlated. However, alanine was not correlated with phosphoethanolamine or with GPC, nor was choline concentration correlated with those of several other metabolites in the malignant tissues. Furthermore, principle components analysis revealed that those metabolites with the strongest diagnostic accuracy did not consistently load with the others.

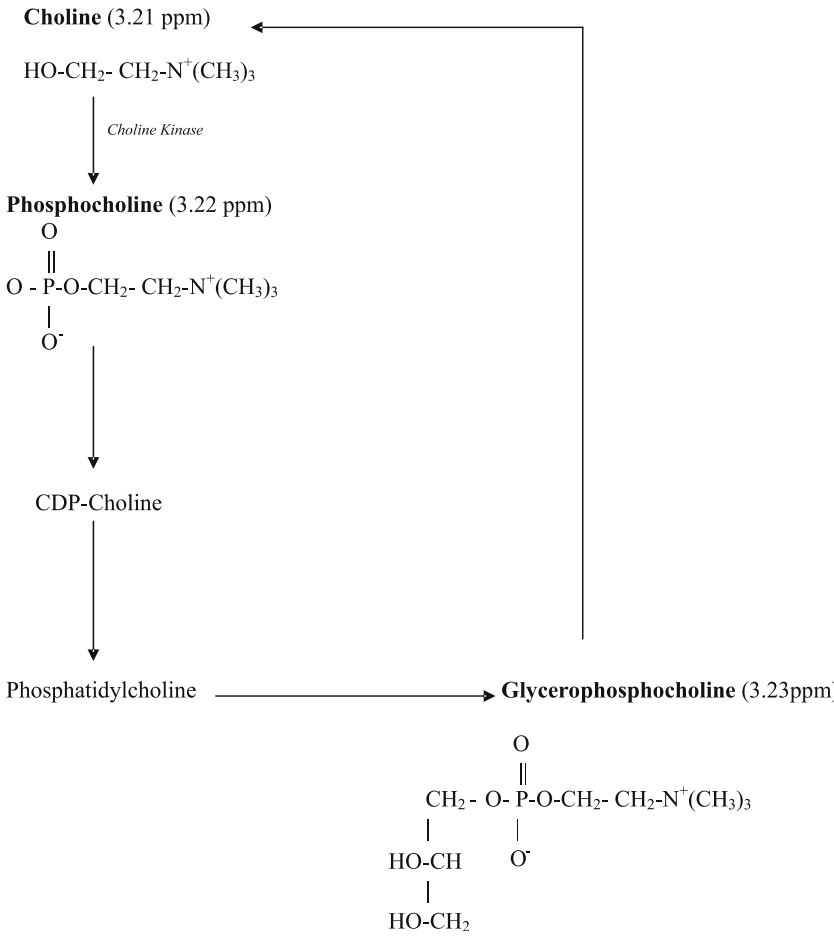

Figure 3. Major steps in choline metabolism in mammalian cells through the cytosine diphosphate (CDP)-choline pathway [45–47], emphasizing ^{1}H NMR visible compounds. Chemical shift is denoted in parts per million (ppm).

We also analyzed the ratio of PC to GPC from Ref. [40]. Unlike most of the individual metabolites, the PC/GPC ratio was highly skewed and kurtotic (3.9 and 16.9, respectively), so that only non-parametric statistical methods were used. The Wilcoxon matched pairs test revealed a significantly higher PC/GPC ratio in the breast cancer samples compared to the normal tissue from the same patient ($p = 0.005$). These paired findings for the PC/GPC ratio are displayed in figure 2. Using the Mann–Whitney test with all the measured data, the breast cancer samples showed a significantly higher PC/GPC ratio compared to the normal, non-infiltrated tissue ($p = 0.000029$, 2-sided exact p). These findings corroborate human breast cell line research, indicating that malignant transformation is associated with a so-called "glycerophosphocholine to phosphocholine

switch" [44], related, *inter alia*, to over-expression of the enzyme choline kinase responsible for PC synthesis [39, 45], and also reflecting altered membrane choline phospholipid metabolism. Figure 3 presents the major steps in choline metabolism in mammalian cells through the cytosine diphosphate (CDP)-choline pathway [45–47]. Therein, emphasis is given to the ^{1}H NMR visible compounds, namely: choline (3.21 ppm), PC (3.22 ppm), and GPC (3.23 ppm), underscoring the clinico-biological importance of analyzing the relationship among these closely overlapping resonances. On the other hand, by summing these three metabolites as "total choline", as is currently done with *in vivo* NMR, substantial information for breast cancer diagnostics is missed.

3. Potential relevance of Padé optimized *in vivo* NMR for early breast cancer diagnostics

Based upon these *in vitro* findings, it is natural to expect that a richer "window" of informational content ought also to be available for extraction from NMR data encoded from the breast *in vivo*. This has generally been missing from the literature thus far. One of the key obstacles to a greater use of MRS in breast cancer diagnostics has been the lack of a uniform approach to data analysis. Data compatibility is needed in order to make multi-center comparisons that are indispensable for wider routine application of *in vivo* NMR in oncology [48]. Differences in data processing methods, rather than real differences, are considered as the major contributor to deviations among various clinical results in MRS [49]. At loggerheads with the need for data compatibility is the requirement within the FFT for fitting, which can lead both to spurious peaks (over-modeling or overfitting) and true metabolites being undetected (under-modeling or underfitting). This is unacceptable for medical diagnostics, and also renders inter-study comparisons tenuous, at best, unless, e.g. the same *in vitro* basis set is used to predetermine the number of metabolites. Such fitting is based upon prior-knowledge/measurement of *in vitro* data, before analysis of the actual *in vivo* spectrum has been attempted [18]. Although claims have been made that fitting procedures can be automatic and objective, their major pitfalls and inherent subjectivity have been highlighted [1]. If one does not include an *in vitro* metabolite in the basis set of, e.g. the so-called Linear Combination Model of *in vitro* spectra [18], then one is going to have a very bad fit precisely at the frequency location where the missing metabolite is expected to occur in the studied *in vivo* spectrum. This leads to subjectivity of all estimators that guess the number of metabolites. These problems are most pronounced with respect to overlapping resonances, which have been shown to be of utmost importance in breast cancer diagnostics, e.g. at 1.3–1.4 ppm where lipids and lactate overlap, and at 3.2 ppm where the difference between PC and GPC is at the level of 0.01 ppm. As elaborated here and in the cited references, not only can the FPT retrieve

overlapping resonances with fidelity, but it also provides unequivocal quantification. As noted, breast cancer diagnostics using *in vivo* NMR have relied mainly upon assessments of the composite choline peak. Notwithstanding the need to expand the number of metabolites upon which this diagnosis is made, accurate quantification of total choline and its components via the FPT would represent an important breakthrough. Moreover, the improved resolution and SNR provided by the FPT could help detect low concentrations of total choline, as well as offer the possibility of detecting potentially informative resonances with short T_2-relaxation times, that will have decayed at longer TE, such as myoinositol, whose concentrations yielded the clearest distinction between breast cancer and fibroadenoma in our analysis [31] of *in vitro* MRS data from Ref. [40]. Furthermore, our initial results for 2D FPT with respect to 2D MRS, using cross-correlation diagrams reveal marked improvement in resolution relative to 2D FFT [50]. In 2D MRS one of the time axes is usually long, but the other must be short to maintain a reasonable scanning time. This is where estimators that can extract information from short-time signals become indispensable. The FPT has been shown to fulfill this critical task, which is virtually unfeasible by attempts to extend the 1D fitting algorithms to two-dimensions. The possibilities of, e.g. 2D J-resolved MRS to extract vital information for tumor detection have been underscored. With the aid of the Padé-based data analysis, these advantages of 2D MRS could be optimally utilized for breast cancer diagnostics.

Work is on-going to apply the FPT to *in vivo* MRS signals encoded from patients with breast and other cancers, and compare these to findings from non-malignant tissue, aiming at optimizing cancer detection, especially at an early stage. Another vital component is the use of the FPT for quantitative analysis of *in vivo* NMR signals from non-malignant lesions that have thus far presented differential diagnostic dilemmas, notably benign tumors, infectious or inflammatory lesions. It will be particularly important to consider benign breast conditions that are difficult to distinguish from breast cancer using other non-invasive diagnostic modalities. These difficult differential diagnoses include, e.g. ductal hyperplasia, fibroadenoma, and fibrocystic changes; it should also be noted that the lactating breast has high concentrations of choline. Insofar as the expected improvements in diagnostic accuracy are achieved by Padé-based *in vivo* NMR, this could then be applied in younger women at high risk for breast cancer with the aim of determining the suitability of this newly emerging methodology for surveillance and screening.

Acknowledgments

This work was supported by the following four Swedish funds: the Stockholm's County Council through FoUU at the Karolinska Hospital, the Swedish

Scientific Research Council (Vetenskapsrådet), the Swedish Cancer Society (Cancerföreningen), and King Gustav the Fifth's Jubilee Foundation.

References

[1] Dž. Belkić, *Quantum Mechanical Signal Processing and Spectral Analysis* (Institute of Physics Publishing, Bristol, 2004).
[2] Dž. Belkić and K. Belkić, Int. J. Quantum Chem. 105 (2005) 493.
[3] W.W.F. Pijnappel, A. van den Boogaart, R. De Beer, and R. D. van Ormondt, Magn. Reson. 97 (1992) 122.
[4] B. Porat, *A Course in Digital Signal Processing* (Wiley, New York, 1997).
[5] Dž. Belkić, Nucl. Instrum. Meth. Phys. Res. A 525 (2004) 366.
[6] Dž. Belkić and K. Belkić, Phys. Med. Biol. 50 (2005) 4385.
[7] Dž. Belkić and K. Belkić, Phys. Med. Biol. 51 (2006) 1049.
[8] Dž. Belkić, Exact quantification of time signals in Padé-based magnetic resonance spectroscopy, Phys. Med. Biol. 51 (2006) 2633.
[9] Dž. Belkić, P.A. Dando, J. Main and H.S. Taylor, J. Chem. Phys. 113 (2000) 6542.
[10] Dž. Belkić, P.A. Dando, H.S. Taylor and J. Main, Chem. Phys. Lett. 315 (1999) 135.
[11] Dž. Belkić, P.A. Dando, H.S. Taylor, J. Main and S-K. Shin, J. Phys. Chem. A. 104 (2000) 11677.
[12] M. Deschamps, I. Burghardt, C. Derouet, G. Bodenhausen and Dž. Belkić, J. Chem. Phys. 113 (2000) 1630.
[13] J. Main, P.A. Dando, Dž. Belkić and H.S. Taylor, Europhys. Lett. 48 (1999) 250.
[14] J. Main, P.A. Dando, Dž. Belkić and H.S. Taylor, J. Phys. A. 33 (2000) 1247.
[15] Dž. Belkić, Nucl. Instrum. Meth. Phys. Res. A 525 (2004) 379.
[16] Dž. Belkić, Nucl. Instrum. Meth. Phys. Res. A 525 (2004) 372.
[17] Dž. Belkić, MAGMA 15 (2002) 36.
[18] J. Pfeuffer, I. Tkáč, S.W. Provencher and R. Gruetter, J. Magn. Reson. 141 (1999) 104.
[19] C.K. Kuhl, W. Kuhn and H. Schild, The Breast 14 (2005) 480.
[20] H. Kuni, I. Schmitz-Feuerhake and H. Dieckmann, Gesundheitswesen 65 (2003) 44.
[21] M.P. Laderoute, Br. J. Cancer 90 (2004) 278.
[22] J.A. Smith and E. Andreopoulou, Ann. Oncol. 15 (Suppl. 1) (2004) i18.
[23] D.B. Kopans, *Breast Imaging*, 2nd ed. (Lippincott-Raven Publishers, Philadelphia, 1998).
[24] G.A. Greendale, B.A. Reboussin, S. Slone, C. Wasilauskas, M.C. Pike and G. Ursin, J. Natl. Cancer Inst. 95 (2003) 30.
[25] S.J. Nass, C. Henderson and J.C. Lashof (eds.), *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancers* (National Academy Press, Washington, D.C., 2001).
[26] E.A. Morris, Radiol. Clin. N. Am. 40 (2002) 443.
[27] R. Katz-Brull, P.T. Lavin and R.E. Lenkinski, J. Natl. Cancer Inst. 9 (2002) 1197.
[28] C. Kuhl, S. Schrading, C.C. Leutner, N. Morakkabati-Spitz, E. Wardelmann, R. Fimmers, W. Kuhn and H.H. Schild, J. Clin. Oncol. 23 (2005) 8469.
[29] M. Kriege, C.T.M. Brekelmans, C. Boetes, P.E. Basnard, H.M. Zonderland and I.M. Obdeijn et al., N. Engl. J. Med. 351 (2004) 427.
[30] M. Robson, Clin. Breast Cancer 5 (2004) 260.
[31] K. Belkić, Isr. Med. Assoc. J. 6 (2004) 610.
[32] W. Huang, P.R. Fisher, K. Dulaimy, L.A. Tudorica, B. O'Hea and T.M. Button, Radiology 232 (2004) 585.

[33] J.-K. Kim, S.-H. Park, H.M. Lee, Y.-H. Lee, N.-K. Sung, D.-S. Chung and O.-D. Kim, The Breast 12 (2003) 179.
[34] M.A. Jacobs, P.B. Barker, P.A. Bottomley, Z. Bhujwalla and D.A. Bluemke, J. Magn. Reson. Imaging 19 (2004) 68.
[35] P.J. Bolan, S. Meisamy, E.H. Baker, J. Lin, T. Emory and M Nelson et al., Magn. Reson. Med. 50 (2003) 1134.
[36] P.J. Bolan, L. DelaBarre, E.H. Baker, H. Merkle, L.I. Everson, D. Yee and M. Garwood, Magn. Reson. Med. 48 (2002) 215.
[37] N.R. Jagannathan, M. Singh, V. Govindaraju, P. Raghunathan, O. Coshic and P.K. Julka et al., NMR Biomed. 11 (1998) 414.
[38] M.A. Thomas, N. Wyckoff, K. Yue, N. Binesh, S. Banakar and H.-K. Chung et al., Technol. Cancer. Res. Treat. 4 (2005) 99.
[39] R. Katz-Brull, D. Seger, D. Rivenson-Segal, E. Rushkin and H. Degani, Cancer Res. 62 (2002) 1966.
[40] I.S. Gribbestad, B. Sitter, S. Lundgren, J. Krane and D. Axelson, Anticancer Res. 19 (1999) 1737.
[41] K. Belkić, Nucl. Instru. Meth. Phys. Res. A. 525 (2004) 313.
[42] U. Sharma, A. Mehta, V. Seenu and N.R. Jagannathan, Magn. Reson. Imag. 22 (2004) 697.
[43] D. Rivenzon-Segal, R. Margalit and H. Degani, Am. J. Physiol. Endocrinol. Metab. 283 (2002) E623.
[44] E.O. Aboagye and Z.M. Bhujwalla, Cancer Res. 59 (1999) 80.
[45] K. Glunde, C. Jie and Z.M. Bhujwalla, Cancer Res. 64 (2004) 4270.
[46] R Katz-Brull, R. Margalit and H. Degani, Magn. Reson. Med. 46 (2001) 31.
[47] N.M. Loening, A.M. Chamberlin, A.G. Zepeda, R. Gilberto Gonzalez and L.L. Cheng, NMR Biomed. 18 (2005) 413.
[48] C. Arús, MAGMA 15 (2002) 38.
[49] E.R. Danielsen and B. Ross, *Magnetic Resonance Spectroscopy Diagnosis of Neurological Diseases* (Marcel Dekker Inc., New York, 1999).
[50] Dž. Belkić, High-Resolution Parametric Estimation of Two-Dimensional Magnetic Resonance Spectroscopy, *20th Annual Meeting of European Soc. Magn. Res. Med. Biol. (ESMRMB), Abstract Number 365 (CD)*, Rotterdam (Netherlands), September 18–21, 2003.