J Biomol NMR (2007) 38:73–77
DOI 10.1007/s10858-007-9148-8

ARTICLE

Quantitative covariance NMR by regularization

Yanbin Chen · Fengli Zhang · David Snyder · Zhehong Gan · Lei Bruschweiler-Li · Rafael Brüschweiler

Received: 17 January 2007 / Accepted: 9 February 2007 / Published online: 10 March 2007

Abstract The square root of a covariance spectrum, which offers high spectral resolution along both dimensions requiring only few t_1 increments, yields in good approximation the idealized 2D FT spectrum provided that the amount of magnetization exchanged between spins is relatively small. When this condition is violated, 2D FT and covariance peak volumes may differ. A regularization method is presented that produces a modified covariance spectrum with cross-peak volumes that closely match their 2D FT analogues. The method is demonstrated for TOCSY spectra with variable mixing times.

Keywords Covariance NMR · Quantitative spectroscopy · Regularization · Ridge regression · 2D FT NMR · Total correlation spectroscopy (TOCSY)

Introduction

There is considerable interest in the development of new methods for the time-efficient acquisition and processing of multi-dimensional NMR data. Such methods are critical for high-throughput applications of biomolecular NMR spectroscopy in emerging fields such as structural proteomics (Vinarov and Markley 2005) and metabolomics (Lenz and Wilson 2007). In covariance NMR spectroscopy, a covariance matrix is constructed from a set of 1D spectra that belong to different evolution times t_1. After application of the matrix square root, a spectrum is obtained that approximates the idealized 2D Fourier transform spectrum (Brüschweiler 2004; Trbovic et al. 2004). Because the square root is by definition positive, negative eigenvalues present in the 2D FT spectrum will not be retrieved by the matrix square-root operation and, consequently, the resulting covariance peak volumes may differ from their 2D FT counterparts. This problem is addressed here by a regularization procedure that shifts the eigenvalues of the 2D FT matrix to positive values prior to covariance processing. The method is demonstrated for TOCSY spectra of strychnine and the decapeptide antamanide.

Y. Chen · D. Snyder · L. Bruschweiler-Li · R. Brüschweiler
Department of Chemistry and Biochemistry, Florida State University, Tallahassee 32306 FL, USA

F. Zhang · Z. Gan · R. Brüschweiler (✉)
National High Magnetic Field Laboratory, Florida State University, Tallahassee 32310 FL, USA
e-mail: bruschweiler@magnet.fsu.edu

Theory

The principle underlying covariance NMR is the conversion of a set of 1D spectra, collected using the canonical 2D FT NMR scheme (Aue et al. 1976), to a symmetric matrix $\mathbf{C}$ representing the covariances of pairs of resonances that are amplitude-modulated in t_1 (Brüschweiler and Zhang 2004). Using Parseval's theorem, it can be shown that $\mathbf{C}$ can be constructed from the corresponding 2D FT spectrum $\mathbf{F}$, up to a constant prefactor, as (Brüschweiler 2004; Trbovic et al. 2004)

$$\mathbf{C} = (\mathbf{F}^{\mathbf{T}}\mathbf{F})^{1/2} \quad (1)$$

The square root ensures that $\mathbf{C}$ approximates the idealized 2D FT spectrum, which is the 2D FT spectrum obtained using a large number of t_1 increments N_1, by suppressing false correlations that may be present in $\mathbf{F}^{\mathbf{T}}\mathbf{F}$ due to resonance overlaps.

The procedure works well for experiments for which the diagonal peaks tend to dominate the cross peaks as is the case for NOESY-type cross-relaxation type experiments (Jeener et al. 1979) and TOCSY experiments (Braunschweiler and Ernst 1983) at short mixing times (assuming that the spectra are phase-corrected as usual such that the diagonal peaks are positive). At longer mixing times, TOCSY cross peaks may be similar or larger in magnitude than the corresponding diagonal peaks. In such cases, the covariance spectrum can display discrepancies in the peak amplitudes with respect to the 2D FT spectrum.

This situation can be illustrated for an IS two-spin 1/2 system that is scalar coupled with a J-coupling constant J_{IS}. In a TOCSY experiment, during the mixing period of duration $\tau_m = 1/(2J_{IS})$ the two spins exchange their polarizations entirely (Braunschweiler and Ernst 1983) so that the corresponding 2D FT spectrum has two positive cross peaks but no diagonal peaks. The spectrum can be schematically represented by a 2×2 matrix

$$\mathbf{F} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2}$$

where the diagonal elements and off-diagonal elements denote the amplitudes of the diagonal peaks and cross peaks, respectively. The corresponding covariance spectrum is then

$$\mathbf{C} = (\mathbf{F}^{\mathbf{T}}\mathbf{F})^{1/2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{1/2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3}$$

which exhibits diagonal peaks but lacks both cross peaks. The discrepancy between $\mathbf{F}$ and $\mathbf{C}$ arises from the fact that $\mathbf{F}$ has one eigenvalue +1 and one eigenvalue −1, whereas the matrix square root selects by default the positive roots. From a statistical perspective, the peak amplitudes of the 1D spectra of the TOCSY experiment are genuinely *uncorrelated* over the evolution period t_1 because the two spins swap their identities during the mixing period. The situation is different in the presence of diagonal peaks, i.e. if a finite amount of magnetization stays on each of the spins. For example, if $\mathbf{F} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ then

$$\mathbf{C} = (\mathbf{F}^{\mathbf{T}}\mathbf{F})^{1/2} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}^{1/2} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{4}$$

i.e. the original 2D FT spectrum $\mathbf{F}$ is recovered. This is because the diagonal of $\mathbf{F}$ is sufficiently large so that none of the eigenvalues of $\mathbf{F}$ is negative.

The problem associated with negative eigenvalues of a general 2D FT spectrum $\mathbf{F}$ can be circumvented by the following regularization procedure. By adding a suitable diagonal part to $\mathbf{F}$ prior to the covariance calculation, most or all eigenvalues can be moved to positive values. In the simplest case, the diagonal part is the unity matrix $\mathbf{1}$ times a prefactor α. Thus, the regularized 2D FT matrix is

$$\mathbf{F}_\alpha = \mathbf{F} + \alpha\mathbf{1} \tag{5}$$

where $\mathbf{1}$ is the unity matrix and α is the regularization factor. Here and in the following it is assumed that $\mathbf{F}$ is a (generally non-symmetric) $N_2 \times N_2$ square matrix, which is achieved, for example, by appropriate zero-filling along the indirect dimension. One obtains for the regularized covariance matrix

$$\mathbf{C}_{reg} = \mathbf{C}_\alpha - \alpha\mathbf{1} \tag{6}$$

where

$$\mathbf{C}_\alpha = (\mathbf{F}_\alpha^{\mathbf{T}}\mathbf{F}_\alpha)^{1/2} = (\mathbf{F}^{\mathbf{T}}\mathbf{F} + \alpha(\mathbf{F}^{\mathbf{T}} + \mathbf{F}) + \alpha^2\mathbf{1})^{1/2} \tag{7}$$

This procedure, which is related to Tikhonov regularization or ridge regression (Tikhonov and Arsenin 1977), yields in the limit of a very large diagonal, $\alpha^2 \gg \|\mathbf{F}^{\mathbf{T}}\mathbf{F}\|$,

$$\mathbf{C}_{\mathrm{reg}} = (\mathbf{F}^{\mathbf{T}} + \mathbf{F})/2 \tag{8}$$

In this limit the regularized matrix $\mathbf{C}_{\mathrm{reg}}$ corresponds to the matrix obtained from direct symmetrization (Baumann et al. 1981). It should be noted that Eqs. 6, 7 operate in the frequency domain along both dimensions, which complicates regularization of data sets that are non-uniformly sampled along the indirect time domain, t_1.

For what follows it is useful to relate α to the sum of the negative real parts of the eigenvalues of $\mathbf{F}$

$$\alpha = -\alpha_0 \sum_{\mathrm{Re}(\lambda_j)<0} \mathrm{Re}(\lambda_j) \tag{9}$$

where λ_j is the jth eigenvalue of $\mathbf{F}$ and α_0 is a scaling factor. The real parts of λ_j enter because $\mathbf{F}$ is generally non-symmetric and therefore has complex eigenvalues. α_0 typically covers a range between 0 (no regularization) and 10^6 (very strong regularization).

Materials and methods

Strychnine was purchased from Sigma–Aldrich and an NMR sample was prepared at 6 mM concentration in deuterated chloroform. A sample of the cyclic decapeptide antamanide ([-Val-Pro-Pro-Ala-Phe-Phe-Pro-Pro-Phe-Phe-]) was prepared with 1 mM concentration in deuterated chloroform (Kessler et al. 1989).

A series of 2D TOCSY experiments (Braunschweiler and Ernst 1983) of strychnine was performed at 800 MHz

and 300 K using MLEV-17 (Bax and Davis 1985) with mixing times from 13 ms to 70 ms using TPPI-States with 128 increments along the indirect dimension t_1. A TOCSY experiment of antamanide was performed under the same conditions with 100 ms mixing time. All spectra were phase- and baseline corrected along ω_1 and ω_2 using NMRPipe (Delaglio et al. 1995). Covariance processing and regularization were performed using in-house MATLAB programs.

Results and discussion

The TOCSY series of strychnine were processed in three different ways: (i) by 2D FT, (ii) by standard covariance NMR (Eq. 1), and (iii) by the regularized covariance method (Eqs. 6, 7). Cross-peak volumes were determined by numerical peak integration. Figure 1 shows a section of the 2D spectra that contains the cross peak indicated by a rectangle between the two methylene protons 20a and 20b, which systematically differ between the 2D FT and covariance spectrum. The corresponding diagonal peaks are relatively weak in the 2D FT but strong in the covariance spectrum. The two protons represent in first order approximation an isolated 2-spin system to which Eqs. 2, 3 apply. A quantitative comparison between the 2D FT and covariance peak volumes for this cross peak is displayed in Fig. 1a. The cross-peak volumes scale in a non-linear way relative to each other for different mixing times τ_m, which is consistent with the prediction by Eq. 3. After regularization using Eqs. 6, 7 with $\alpha_0 = 100$ the covariance peak volume closely matches the one of the 2D FT spectrum over the whole range of mixing times (Fig. 1b). Thus, the regularization procedure removes the discrepancy between 2D FT and covariance NMR for this spin pair.

More subtle differences between 2D FT and covariance NMR spectra can occur in TOCSY spectra of larger spin systems. For the decapeptide antamanide, the 2D FT and covariance peak volumes show a good overall agreement, i.e. no large differences are observed as for the strychnine cross peak described above. Still, the agreement is not perfect as can be seen in Fig. 2a, c and Table 1. Generally, the correlation coefficient between covariance and 2D FT peak volumes decreases slightly when going from $N_1 = 512$ ($r = 0.98$) to $N_1 = 64$ ($r = 0.95$). Also, cross peaks that are close to the diagonal show a slightly poorer agreement than cross peaks that are further away.

Regularization leads to a clear improvement of the correlation ($r = 0.98$–1.00) and at the same time the average volume ratio is closer to 1.0. For example, for $N_1 = 128$ increments the correlation coefficient of the off-diagonal covariance peak volumes with respect to the 2D FT peak volumes is $r = 0.965$ before and $r = 0.996$ after

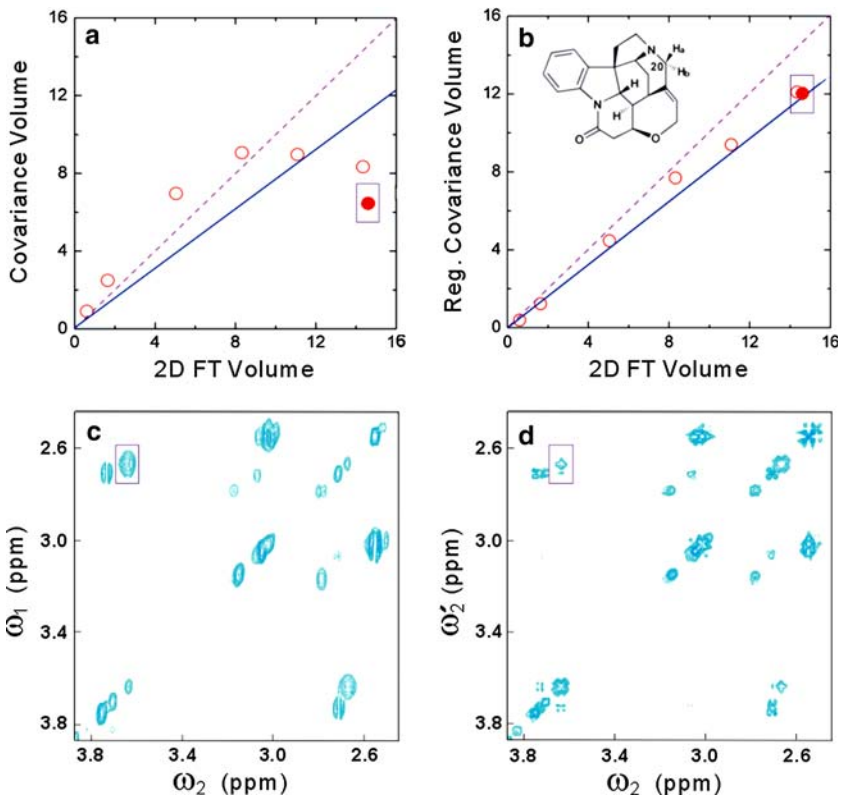

Fig. 1 Relative cross peak volumes of strychnine as a function of mixing time τ_m in TOCSY spectra. The mixing times are 13.3, 22.6, 31.9, 41.2, 50.6, 59.9, and 69.2 ms. (**a**) Volume ratio between methylene H_a–H_b cross peak of carbon 20 between covariance and 2D FT spectra as a function of τ_m. (**b**) Same as (**a**) using regularized covariance spectrum. The inset shows the structure of strychnine with the proton pair labeled. The highlighted circle in (**a**) and (**b**) corresponds to the framed cross peak of the (**c**) 2D FT and (**d**) covariance spectra with $\tau_m = 31.9$ ms. The solid blue lines represent the average slopes

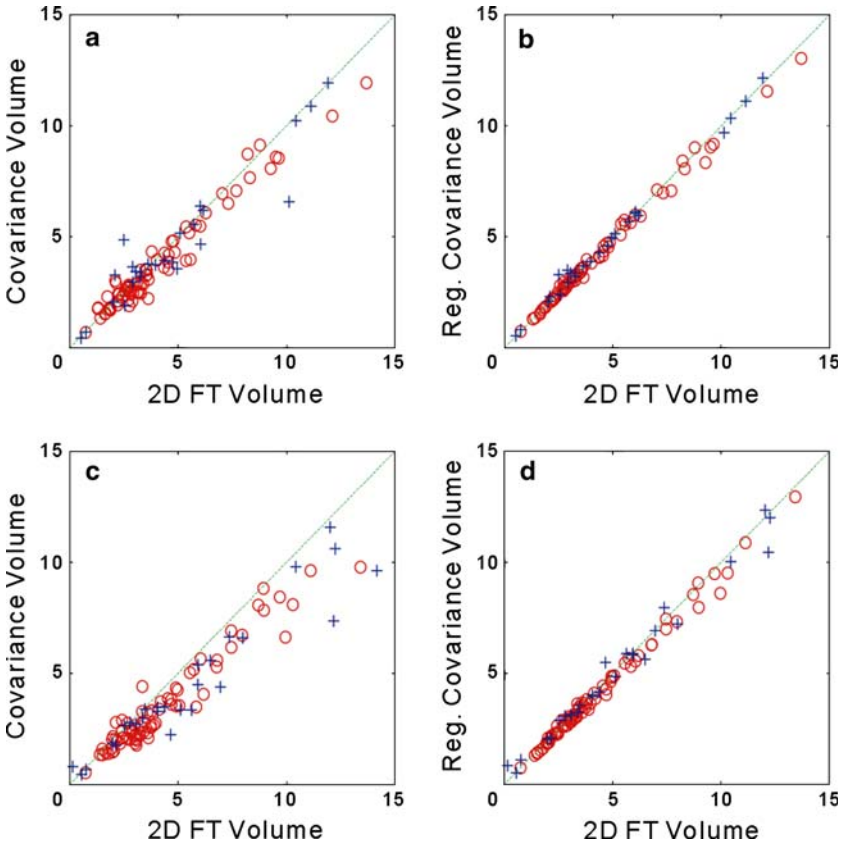

Fig. 2 Correlation plots of cross-peak volumes of covariance and 2D FT TOCSY spectra of antamanide for τ_m = 100 ms. Panels a, c show correlations between 2D FT and standard covariance spectra, and Panels b, d show the correlations between 2D FT and regularized covariance spectra. For Panels a, b, N_1 = 256, whereas for Panels c, d, N_1 = 128. The 'o' symbols represent off-diagonal cross peaks, while the '+' symbols represent near-diagonal cross peaks

Table 1 Cross-peak volume comparison between 2D FT and covariance TOCSY spectra before and after regularization (α_0 = 100). A total of 73 off-diagonal cross peaks and 19 near-diagonal cross peaks are evaluated for each spectrum

N_1		Avg. ratio ± std.[a]	r[b]	Avg. ratio ± std.[c]	r[d]
512	Off[e]	1.05 ± 0.17	0.982	1.02 ± 0.04	0.997
	Near[f]	1.12 ± 0.22	0.961	0.98 ± 0.05	0.999
256	Off	1.08 ± 0.16	0.980	1.02 ± 0.04	0.997
	Near	1.14 ± 0.22	0.970	1.00 ± 0.04	0.998
128	Off	1.24 ± 0.21	0.965	1.04 ± 0.05	0.996
	Near	1.27 ± 0.21	0.961	1.00 ± 0.10	0.983
96	Off	1.23 ± 0.20	0.967	1.04 ± 0.06	0.996
	Near	1.32 ± 0.23	0.947	1.09 ± 0.25	0.969
64	Off	1.13 ± 0.23	0.960	1.02 ± 0.14	0.989
	Near	1.26 ± 0.48	0.954	1.06 ± 0.36	0.984

[a] Ratio between 2D FT cross-peak volume and covariance cross-peak volume without regularization

[b] Pearson correlation coefficient between 2D FT and covariance cross-peak volume without regularization

[c] Ratio between 2D FT and covariance cross-peak volume with regularization

[d] Pearson correlation coefficient between 2D FT and covariance cross-peak volume with regularization

[e] Off-diagonal cross peaks that are $\geq$ 0.352 ppm away from diagonal

[f] Near-diagonal cross peaks that are between 0.059 and 0.352 ppm away from diagonal

regularization, while the average peak volume ratio goes from 1.24 to 1.04.

Generally, cross peaks in close vicinity to the diagonal ('near-diagonal peaks') exhibit a slightly poorer agreement than cross peaks that lie further away from the diagonal ('off-diagonal peaks') because for low N_1 values the covariance statistics is worse for resonance pairs whose chemical shift difference is small (Chen et al. 2006).

2D FT spectra with low numbers of t_1 increments have intrinsically low resolution along ω_1 manifested in the form of large line widths along this dimension. By contrast, the corresponding covariance spectrum retains high resolution along both dimensions. When the covariance spectrum is regularized a star-like peak shape emerges as the regularization parameter α_0 is increased. This effect is demonstrated in Fig. 3 for the TOCSY spectrum of antamanide with N_1 = 96 and α_0 = 0, 1.5, 10, and 100. At α_0 = 1.5 the onset of the star shapes is visible (Fig. 3b). At α_0 = 10, the star shapes are almost completely formed (Fig. 3c) and they are reminiscent of the 'star effect' of Lorentzian line shapes in 2D NMR (Bodenhausen et al. 1977). In the limit of very large α_0 the regularized covariance spectrum approaches $(\mathbf{F}^T + \mathbf{F})/2$ (Eq. 8). Thus, each cross peak, elongated along ω_1, is superimposed with its transposed, elongated along ω_2, which creates the star-like appearance. The star shapes are most pronounced for

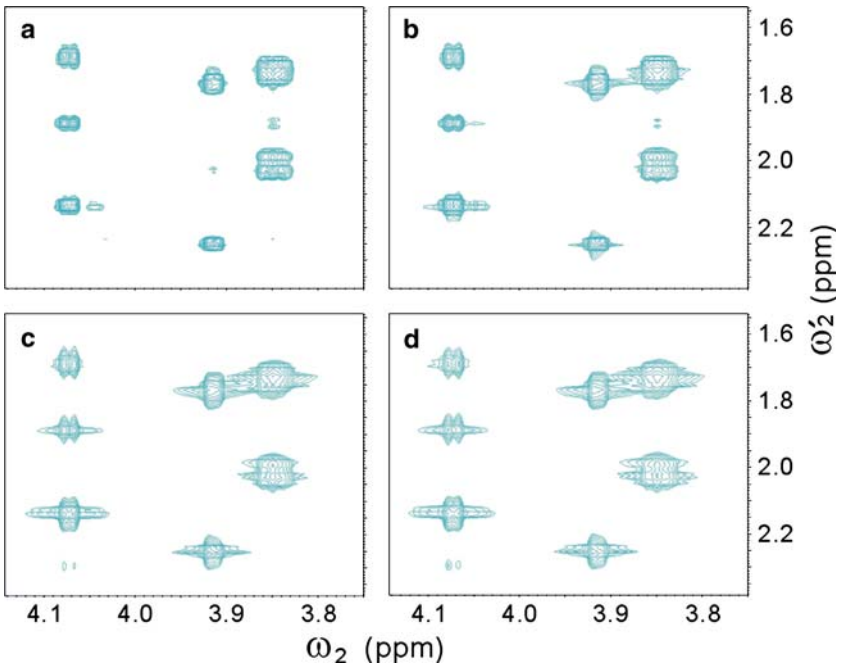

Fig. 3 Star-shape effect of TOCSY cross peaks monitored as a function of regularization for antamanide with $N_1 = 96$ increments. In Panel (**a**) $\alpha_0 = 0$ (no regularization), (**b**) $\alpha_0 = 1.5$, (**c**) $\alpha_0 = 10$, and (**d**) $\alpha_0 = 100$

small N_1 values and disappear for a sufficiently large N_1 irrespective of α_0. Because $(\mathbf{F}^T + \mathbf{F})/2$ does not provide resolution enhancement over the original 2D FT spectrum **F**, for the analysis of densely populated peak regions the scaling parameter α_0 should be kept moderately small to minimize spectral overlap. Because of the computational efficiency of covariance processing, α_0 can be optimized for different regions of the spectrum, if necessary, until the peak volumes of interest do not change any longer relative to the rest of the spectrum.

Acknowledgments This work was supported by the National Institutes of Health (grant GM 066041).

References

Aue WP, Bartholdi E, Ernst RR (1976) 2-dimensional spectroscopy—application to nuclear magnetic-resonance. J Chem Phys 64:2229–2246

Baumann R, Wider G, Ernst RR, Wüthrich K (1981) Improvement of 2D NOE and 2D correlated spectra by symmetrization. J Magn Reson 44:402–406

Bax A, Davis DG (1985) Mlev-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J Magn Reson 65:355–360

Bodenhausen G, Freeman R, Niedermeyer R, Turner DL (1977) Double fourier transformation in high-resolution NMR. J Magn Reson 26:133–164

Braunschweiler L, Ernst RR (1983) Coherence transfer by isotropic mixing—application to proton correlation spectroscopy. J Magn Reson 53:521–528

Brüschweiler R (2004) Theory of covariance nuclear magnetic resonance spectroscopy. J Chem Phys 121:409-414

Brüschweiler R, Zhang F (2004) Covariance nuclear magnetic resonance spectroscopy. J Chem Phys 120:5253–5260

Chen Y, Zhang F, Bermel W, Brüschweiler R (2006) Enhanced covariance spectroscopy from minimal datasets. J Am Chem Soc 128:15564–15565

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) Nmrpipe—a multidimensional spectral processing system based on unix pipes. J Biomol NMR 6:277–293

Jeener J, Meier BH, Bachmann P, Ernst RR (1979) Investigation of exchange processes by 2-dimensional NMR-spectroscopy. J Chem Phys 71:4546–4553

Kessler H, Müller A, Pook KH (1989) Assignment of all proton, carbon, and nitrogen NMR signals of antamanide in chloroform solution. Liebigs Annalen der Chemie 903–912

Lenz EM, Wilson ID (2007) Analytical strategies in metabonomics. J Proteome Res 6:443-458

Tikhonov AN, Arsenin VA (1977) Solution of ill-posed problems. Washington, Winston & Sons

Trbovic N, Smirnov S, Zhang F, Brüschweiler R (2004) Covariance NMR spectroscopy by singular value decomposition. J Magn Reson 171:277–283

Vinarov DA, Markley JL (2005) High-throughput automated platform for nuclear magnetic resonance-based structural proteomics. Expert Rev Proteomics 2:49–55