Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 169 (2004) 215–224
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Ultra-high resolution 3D NMR spectra from limited-size data sets

Jianhan Chen,[a] Daniel Nietlispach,[b] A.J. Shaka,[a,*] and Vladimir A. Mandelshtam[a,*]

[a] *Chemistry Department, University of California, Irvine, CA 92697, USA*
[b] *Department of Biochemistry, University of Cambridge, CB2 1GA, UK*

Received 13 February 2004; revised 21 April 2004
Available online 4 June 2004

Abstract

The advantage of the filter diagonalization method (FDM) for analysis of triple-resonance NMR experiments is demonstrated by application to a 3D constant time (CT) HNCO experiment. With a ^{15}N-,^{13}C-labeled human ubiquitin sample (1.0 mM), high spectral resolution was obtained at 500 MHz in 25 min with only 6–8 increments in each of the CT dimensions. This data set size is about a factor of 50–100 smaller than typically required, yet FDM analysis results in a fully resolved spectrum with a sharp peak for each HNCO resonance. Unlike Fourier transform (FT) processing, in which spectral resolution in each dimension is inversely proportional to the acquisition time in this dimension, FDM is a true multi-dimensional method; the resolution in *all* dimensions is determined by the total information content of the *entire* signal. As the CT dimensions of the 3D HNCO signal have approximate time-reversal symmetry, they can each be doubled by combining the usual four hyper-complex data sets. This apparent quadrupling of the data is important to the success of the method. Thus, whenever raw sensitivity is not limiting, well-resolved *n*-dimensional spectra can now be obtained in a small fraction of the usual time. Alternatively, to maximize sensitivity, evolution periods of faster relaxing nuclei may be radically shortened, the total required resolution being obtained through chemical shift encoding of other, more slowly relaxing, spins. Improvements similar to those illustrated with a 3D HNCO spectrum are expected for other triple-resonance spectra, where CT evolution in the indirect dimensions is implemented.

Keywords: Filter diagonalization; FDM2K; HNCO; Triple-resonance experiments; Nonlinear processing; Protein assignment; Resolution enhancement

1. Introduction

To obtain a highly resolved three-dimensional (3D) Fourier transform (FT) frequency spectrum from a 3D time domain signal, data must be recorded for sufficiently long enough in each of the three time dimensions. The resolution is determined by the Fourier uncertainty principle, which applies to each dimension separately

$$\delta\nu_p \sim 1/N_p\tau_p, \tag{1}$$

where $\delta\nu_p$, N_p, and τ_p are the frequency resolution, the number of data points, and the sampling interval in the pth dimension ($p = 1, 2, 3$), respectively. Resolution in each dimension is independent of that in any others, a consequence of the multi-dimensional FT being simply a sequence of one-dimensional FTs applied to the rows and columns of the multi-dimensional data array. As such, information contained in the raw data may not be exploited to its full extent, e.g., information obtainable from joint correlations amongst signal evolution in different dimensions cannot be considered. To achieve the required resolution in FT spectra of biological systems, data sets are therefore typically obtained by collecting many time points in each indirect dimension, a process taking up to several days. This problem motivated recent attempts to develop alternative experimental and processing schemes that could in principle yield adequate multidimensional spectra by combining data from lower dimensional experiments (see, e.g. [1,2]). As such it was demonstrated [1] that a 5D NMR spectrum can be constructed.

* Corresponding authors. Fax: 1-949-824-8571.
E-mail addresses: ajshaka@uci.edu (A.J. Shaka), mandelsh@uci.edu (V.A. Mandelshtam).

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.04.017

For NMR signals it is believed that an appropriate parametric model for the data, such as assuming Lorentzian line shapes is *in principle* beneficial. Methods such as maximum entropy, linear prediction, and others can potentially yield higher spectral resolution than that given in Eq. (1) (see, e.g. [3]). However, difficulties in correctly formulating the multi-dimensional spectral analysis problem, very unfavorable numerical scaling with the data size, and/or the very ill-conditioned nature of the corresponding inverse problem may make these methods only applicable to short time dimensions, typically the indirect dimensions.

By contrast, the filter diagonalization method (FDM) discussed recently [4–8] is a true multi-dimensional method that in essence fits the entire data set to the multi-dimensional Lorentzian model. The large-scale, highly non-linear spectral or parameter estimation problem is efficiently solved by matrix linear algebra, diagonalizing small matrices computed from the data. When the data exactly conforms to the assumed model, the FDM solution is exact. The problems associated with large data sets are overcome by implementing the spectral inversion locally in the frequency domain, a "divide-and-conquer" approach that leads to small matrices of controllable size. The resulting generalized eigenvalue problems are still ill-conditioned, requiring a regularization. Among several possible regularization approaches the simplest method (FDM2K) was proposed in [7].

In conventional multi-dimensional NMR experiments, either N- and P-type phase-modulated signals or cosine and sine amplitude-modulated signals are acquired for each of the indirect dimensions to give a pure-phase absorption-mode FT spectrum. Thus, 2^n complementary data sets are required to return a single n-dimensional absorption-mode FT spectrum. FDM processing, however, only requires a single phase-modulated signal to compute an absorption-mode spectrum. As such, the required data formats for FDM and FT processing are not identical, necessitating some data manipulation.

Many protein backbone triple-resonance NMR experiments employ a constant time (CT) scheme [9,10] embedding chemical shift evolution into existing coherence transfer periods, shortening the duration of the pulse sequence and improving sensitivity. Similarly, J modulation from coupling constants can be suppressed by a CT period without band-selective RF pulses. As shown previously [12], CT signals have properties making them particularly suitable for FDM analysis. First, CT signals have essentially perfect Lorentzian line shapes in the CT dimensions. Second, a doubling scheme exists for two-dimensional CT signals, which allows FDM to process two hypercomplex data sets as a single data set of twice the size. The sudden onset of convergence with nonlinear methods means that doubling the signal size can give a dramatic improvement in the obtainable resolution. Further, with the doubling scheme, optimal data formats for FDM and FT spectra become transparently compatible. In the case of the CT-HNCO [11] experiment tested here, the $^{13}C'$ evolution period was therefore converted into a short CT period. In conjunction with the ^{15}N CT evolution, an effective quadrupling of the 3D data size was obtained. This was expected to have a dramatic impact on the resolving power of FDM, and is indeed so.

In CT dimensions the maximum signal length is limited by half of the total constant period $2T$, the obtainable FT resolution is thus determined by $1/T$ (vide infra). Moreover, as there is negligible decay along the CT dimensions, the truncated time domain data Fourier transforms into an unfavorable sinc-modulated signal. Apodization reduces these truncation artifacts, but further broadens the lines. A longer constant-time period is therefore preferred to improve resolution. However, a longer CT period reduces sensitivity, as during the CT period the observed signal is attenuated by a factor $\sim e^{-2T/T_2}$, T_2 being the traverse relaxation time constant. Thus, a short CT period is preferred to maximize sensitivity, an increasingly important factor with faster transverse relaxation. Accordingly, a compromise between resolution and sensitivity is necessary when working with the conventional FT spectrum. Such a trade-off may no longer be necessary when FDM processing is used: it is possible to use the information encoded in a long dimension to enhance the resolution in the short dimensions significantly.

We show that very high quality spectra can be obtained from a rapid 3D HNCO experiment employing short CT chemical shift evolution in both $^{13}C'$ and ^{15}N indirect dimensions. The results are compared with the corresponding FT spectra, conclusively demonstrating the dramatic benefit that comes from the multi-dimensional character of the FDM approach.

2. Materials and methods

2.1. CT-HNCO experiment

The standard HNCO [11] experiment was slightly modified to implement CT evolution in *both* indirect dimensions as shown in the pulse sequence of Fig. 1. Data sets were recorded at 298 K on a Bruker DRX500 spectrometer equipped with a 5 mm HCN/z-gradient probe on a sample of uniformly ^{15}N-,^{13}C-labeled human ubiquitin in 95% H_2O. Eight scans were acquired per increment on the 1.0 mM sample. Standard "States-TPPI" phase cycling was implemented in the $^{13}C'$ dimension, giving a sine/cosine amplitude modulated pair of data sets and gradient selection was used in the ^{15}N dimension, giving an N/P-type phase mod-

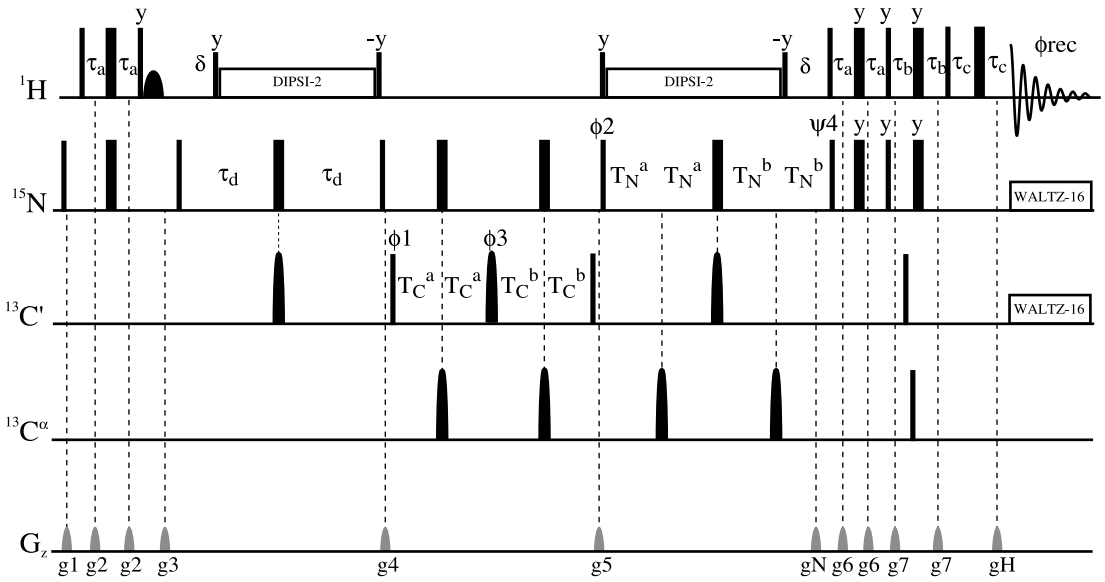

Fig. 1. Experimental scheme for the CT-HNCO pulse sequence. Square 90° and 180° pulses are indicated by black vertical lines, respectively, with the phases, if different from x, indicated above the pulses. The ^{1}H, ^{13}C′, and ^{15}N carriers are set to 4.75 (water), 176.0, and 120.0 ppm, respectively. All proton pulses are applied with a field strength of 24 kHz with the exception of the shaped 90° pulse, which has the E-SNOB [13] profile (2.1 ms, 0.27 kHz) and the DIPSI-2 [14] decoupling sequence including the two surrounding 90° pulses (4.9 kHz). ^{15}N pulses employ a 6.3 kHz field. At 500 MHz, ^{13}C square pulses have a length of 64.2 μs to minimize excitation of the ^{13}C$^\alpha$ region. The shaped 180° ^{13}C′ and ^{13}C$^\alpha$ pulses are applied with a G3 [15] profile at a peak field strength of 7.03 kHz and have a duration of 512 μs. ^{13}C$^\alpha$ pulses are applied phase-modulated at 56 ppm. Decoupling is achieved with a WALTZ-16 [16] sequence for ^{15}N (1.0 kHz field strength) and ^{13}C′ (300 μs 90° pulses with SEDUCE-1 [17] profile, 3.6 kHz peak power). The delays used are $\tau_a = 2.25$ ms, $\tau_b = 2.75$ ms, $\tau_c = 0.7$ ms, $\tau_d = 12.4$ ms, $\delta = 5.4$ ms, $T_C^a = (T_C - t_1)/2$, $T_C^b = (T_C + t_1)/2$, $T_N^a = (T_N - t_2)/2$, $T_N^b = (T_N + t_2)/2$ with the CT periods $2T_C = 8.0$ ms and $2T_N = 24.8$ ms. The phase cycling is as follows: $\phi_1 = (x, -x)$, $\phi_2 = 2(x), 2(-x)$, $\phi_3 = 4(x), 4(-x)$, and $\phi_{rec} = (x, -x, -x, x)$. Frequency discrimination in the t_1(^{13}C′) dimension is achieved using States-TPPI [18] of ϕ_1 while in the t_2(^{15}N) dimension a phase-sensitive spectrum is obtained by recording a second FID for each increment of t_2, with the pulse phase ψ_4 inverted and the amplitude of the gradient gN reversed [19]. For every successive t_2 increment ϕ_2 and ϕ_{rec} are inverted. Durations and strengths of the gradients (sine-bell) are $g1$ =(0.8 ms, 6 G/cm), $g2$ =(0.65 ms, 2.5 G/cm), $g3$ =(1.3 ms, 7 G/cm), $g4$ =(1.1 ms, 4.5 G/cm), $g5$ =(0.8 ms, 8 G/cm), $g6$ =(0.65 ms, 12 G/cm), $g7$ =(0.65 ms, 4 G/cm), gN =(2.0 ms, 20 G/cm) and gH =(0.6 ms, 6.8 G/cm).

ulated pair of data sets. The CT period was fixed at $2T = 24.8$ ms in the ^{15}N dimension, to maximize the magnetization transfer. A short constant time period $2T = 8.0$ ms was used in the ^{13}C′ dimension to minimize the intensity loss by transverse relaxation. A maximum of 6 ^{13}C′ increments were acquired over a spectral width of 1 kHz. While up to 48 increments in the ^{15}N dimension would have been compatible with the 2 kHz ^{15}N spectral width, only 8 increments were actually acquired, minimizing NMR instrument time, and allowing study of the lower signal size limit required for FDM analysis to resolve all the resonances. The detected proton FID was 1024 complex points over a 10 kHz spectral width. Consequently, a 3D signal $N_1 \times N_2 \times N_3 = 6 \times 8 \times 1024$ (for each of the four data sets) was obtained. The total acquisition time was 25 min.

2.2. A quadrupling scheme for 3D experiments with two CT dimensions

For a 3D NMR signal with two CT dimensions, the doubling scheme [12] can be applied independently to each CT dimension. The same four data sets acquired in the conventional set-up (to construct the absorption-mode FT spectrum) can be combined together and processed by FDM2K as a *single* data set, four times as large. The doubling scheme thus becomes a *quadrupling scheme*.

If each 3D resonance is described by three *complex* frequencies ω_{pk} ($p = 1, 2, 3$) and a complex amplitude d_k, the four hyper-complex data sets can be written as

$$\begin{aligned} S_{CN}(t_1,t_2,t_3) &= \sum_k d_k \cos(\omega_{1k}t_1)e^{-i\omega_{2k}t_2}e^{-i\omega_{3k}t_3}, \\ S_{SN}(t_1,t_2,t_3) &= \sum_k d_k \sin(\omega_{1k}t_1)e^{-i\omega_{2k}t_2}e^{-i\omega_{3k}t_3}, \\ S_{CP}(t_1,t_2,t_3) &= \sum_k d_k \cos(\omega_{1k}t_1)e^{i\omega_{2k}t_2}e^{-i\omega_{3k}t_3}, \\ S_{SP}(t_1,t_2,t_3) &= \sum_k d_k \sin(\omega_{1k}t_1)e^{i\omega_{2k}t_2}e^{-i\omega_{3k}t_3} \end{aligned} \tag{2}$$

with $t_p \in [0, T_p]$ ($p = 1, 2, 3$). Only the proton frequencies ω_{3k} are in fact complex, while ω_{1k} and ω_{2k} are essentially real reflecting negligible decay in the CT dimensions. The latter property implies the following symmetries with respect to t_1 and t_2:

$$
\begin{aligned}
&S_{\mathrm{CN}}(-t_1,t_2,t_3) = S_{\mathrm{CN}}(t_1,t_2,t_3),\\
&S_{\mathrm{CP}}(-t_1,t_2,t_3) = S_{\mathrm{CP}}(t_1,t_2,t_3),\\
&S_{\mathrm{SN}}(-t_1,t_2,t_3) = -S_{\mathrm{SN}}(t_1,t_2,t_3),\\
&S_{\mathrm{SP}}(-t_1,t_2,t_3) = -S_{\mathrm{SP}}(t_1,t_2,t_3),\\
&S_{\mathrm{CN}}(t_1,-t_2,t_3) = S_{\mathrm{CP}}(t_1,t_2,t_3),\\
&S_{\mathrm{SN}}(t_1,-t_2,t_3) = S_{\mathrm{SP}}(t_1,t_2,t_3).
\end{aligned}
$$

Thus, by defining

$$
S(t_1,t_2,t_3) = \begin{cases}
S_{\mathrm{CN}}(t_1,t_2,t_3) - \mathrm{i}S_{\mathrm{SN}}(t_1,t_2,t_3), & t_1 \geqslant 0,\ t_2 \geqslant 0,\\
S_{\mathrm{CN}}(t_1,t_2,t_3) + \mathrm{i}S_{\mathrm{SN}}(t_1,t_2,t_3), & t_1 < 0,\ t_2 \geqslant 0,\\
S_{\mathrm{CP}}(t_1,t_2,t_3) - \mathrm{i}S_{\mathrm{SP}}(t_1,t_2,t_3), & t_1 \geqslant 0,\ t_2 < 0,\\
S_{\mathrm{CP}}(t_1,t_2,t_3) + \mathrm{i}S_{\mathrm{SP}}(t_1,t_2,t_3), & t_1 < 0,\ t_2 < 0.
\end{cases} \tag{3}
$$

a single purely phase-modulated signal can be constructed

$$
S(t_1,t_2,t_3) = \sum_k d_k \mathrm{e}^{-\mathrm{i}\omega_{1k}t_1}\,\mathrm{e}^{-\mathrm{i}\omega_{2k}t_2}\,\mathrm{e}^{-\mathrm{i}\omega_{3k}t_3} \tag{4}
$$

with $t_1 \in [-T_1, T_1]$, $t_2 \in [-T_2, T_2]$, $t_3 \in [0, T_3]$, i.e., twice as long in each of the CT dimensions. A similar procedure can be designed to accommodate any combination of frequency discrimination schemes in the indirect dimensions.

These symmetries cause the FT spectrum from such a signal, computed by a discrete Fourier transform over the symmetric time domains $[-T_1, T_1]$ and $[-T_2, T_2]$, to automatically possess absorption line shapes in the first and second dimensions. However, the corresponding spectral resolutions are still determined by $1/T_1$ and $1/T_2$ rather than $1/2T_1$ and $1/2T_2$. That is, extending the signal with time-reversal symmetry into the negative time domain is not advantageous for FT spectra. However, for a nonlinear method like FDM2K, the new signal contains roughly four times more information that can be fully utilized.

2.3. FDM2K: the regularized filter diagonalization method

The theory of FDM has been described in detail previously and is only outlined here (for a review see [8]). The experimental time signal is defined as a 3D array $S(n_1\tau_1, n_2\tau_2, n_3\tau_3)$ with $n_1 = -N_1, \ldots, N_1$, $n_2 = -N_2, \ldots, N_2$, and $n_3 = 0, \ldots, N_3$; τ_1, τ_2, and τ_3 are the sampling intervals used in the experiment. The derivation starts by assuming there exists an auxiliary complex symmetric space, three linear operators $\hat{U}_p$ $(p = 1, 2, 3)$ acting in this space, and a state vector ϕ, that allow the data to be represented in the form

$$
S(n_1\tau_1, n_2\tau_2, n_3\tau_3) = \phi^{\mathrm{T}} \hat{U}_1^{n_1} \hat{U}_2^{n_2} \hat{U}_3^{n_3} \phi. \tag{5}
$$

The operators are assumed to commute

$$
\hat{U}_p \hat{U}_{p'} = \hat{U}_{p'} \hat{U}_p
$$

and be symmetric

$$
\hat{U}_p = \hat{U}_p^{\mathrm{T}}.
$$

It can be shown that for purely phase modulated data this ansatz is equivalent to the Lorentzian assumption of Eq. (4). Note that assuming the operators commute makes explicit use of any correlation between signal evolution in different time dimensions.

A number of different spectra can be computed using FDM. Here we show how to estimate 2D projections of the 3D spectrum, for example, the 2D infinite-time discrete Fourier transform (DFT) along the first and second dimensions

$$
I(\omega_1,\omega_2) := \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} S(n_1\tau_1, n_2\tau_2, n_3, 0) z_1^{-n_1} z_2^{-n_2} \tag{6}
$$

with $z_1 = \mathrm{e}^{-\mathrm{i}\tau_1\omega_1}$ and $z_2 = \mathrm{e}^{-\mathrm{i}\tau_2\omega_2}$.

Inserting the ansatz (5) and summing the geometric series analytically, we obtain the resolvent formula

$$
I(\omega_1,\omega_2) = \phi^{\mathrm{T}} (1 - \hat{U}_1/z_1)^{-1} (1 - \hat{U}_2/z_2)^{-1} \phi. \tag{7}
$$

This is still a formal expression as the auxiliary objects, $\hat{U}_p$ and ϕ, are not known explicitly. However, it can be shown (see, e.g. [8]) that Eq. (7) can be evaluated by pure linear algebra. In the framework of FDM2K [7] this is realized by first selecting a small 3D frequency window. Practically, it needs to be small only in the long dimensions, i.e., the proton dimension in the present case. Then one constructs three data matrices $\mathbf{U}_p$ $(p = 0, 1, 2)$ from the entire 3D data set $S(n_1\tau_1, n_2\tau_2, n_3\tau_3)$, and solves the two regularized generalized eigenvalue problems

$$
\mathbf{U}_0^{\dagger} \mathbf{U}_p \mathbf{B}_{pk} = u_{pk} (\mathbf{U}_0^{\dagger} \mathbf{U}_0 + q^2) \mathbf{B}_{pk} \quad (p = 1, 2) \tag{8}
$$

for the eigenvectors $\mathbf{B}_{pk}$ and eigenvalues u_{pk}. Here $\mathbf{U}_0^{\dagger}$ is the Hermitian conjugate of $\mathbf{U}_0$.

The matrix $(\mathbf{U}_0^{\dagger}\mathbf{U}_0 + q^2)$ in the right-hand side of Eq. (8) is positive definite with the condition number depending on q^2. This effectively regularizes the generalized eigenvalue problem. For a sufficiently large value of q the artifacts caused by the ill-conditioned nature of $\mathbf{U}_p$ matrices are eliminated. Further increase of q leads to broadening the lines, the weaker peaks being damped more rapidly than the strong ones (see [7] for more details). To set the value of the regularization parameter q we use the formula

$$
q = \bar{q}(N_1+1)(N_2+1)(N_3+1) \times \left[\sum_{n1=0}^{N_1} \sum_{n2=0}^{N_2} \sum_{n3=0}^{N_3} |S(n_1\tau_1, n_2\tau_2, n_3, 0)|^2 \right]^{1/2}, \tag{9}
$$

that relates it to the scaled regularization parameter $\bar{q}$. The values of $\bar{q}$ are typically chosen from the range $[0.01; 0.1]$. Because the results only weakly depend on q, a few calculations are sufficient to optimize its choice [7,12].

The infinite-time DFT spectrum in the selected frequency domain is then estimated by

$$I(\omega_1,\omega_2)=\sum_{k,k'}\frac{d_{kk'}}{(1-u_{1k}/z_1)(1-u_{2k'}/z_2)} \tag{10}$$

with

$$d_{kk'}=\mathbf{C}^{\mathrm{T}}\mathbf{B}_{1k}\mathbf{B}_{1k}^{\mathrm{T}}\mathbf{B}_{1k'}\mathbf{B}_{1k'}^{\mathrm{T}}\mathbf{C} \tag{11}$$

and **C**, a column vector, defined by the full 3D data array [8]. The spectral parameters, the amplitudes $d_{kk'}$ and the poles $u_{1k}=\mathrm{e}^{-\mathrm{i}\tau_1\omega_{1k}}$ and $u_{2k}=\mathrm{e}^{-\mathrm{i}\tau_2\omega_{2k}}$, completely define the parameters of the NMR peaks, in principle. However, as discussed previously [8], even in the simple 1D case a meaningful line list construction may not be straightforward, as single peaks may be represented by a superposition of several Lorentzian peaks. In the 2D spectral estimation expression (10) there are many more terms than peaks. Although most of them are numerically negligible, their presence further complicates the problem of a relevant line list construction. Note, however, that these complications have no critical effect on the appearance of the FDM2K spectral estimation.

Eq. (10) gives a phase-twist spectrum consistent with the standard complex DFT of the data, albeit potentially having much better resolution than the latter, the transform-limited line widths being replaced by estimates of the actual line widths. Thus, Eq. (10) can be used in the standard framework to obtain a double-absorption spectrum. However, essentially the same goal can be achieved by the short cut of replacing the complex Lorentzian lineshapes in Eq. (10) by any other suitable form, e.g., the double-absorption Lorentzian lineshapes

$$A(\omega_1,\omega_2)=\sum_{k,k'}\mathrm{Re}\ \{d_{kk'}\}\mathrm{Re}\left\{\frac{1}{1-u_{1k}/z_1}-\frac{1}{2}\right\}\times\mathrm{Re}\left\{\frac{1}{1-u_{2k'}/z_2}-\frac{1}{2}\right\}, \tag{12}$$

or even double-absorption Gaussian lineshapes

$$A_G(\omega_1,\omega_2)=\pi\ln(2)\sum_{k,k'}\frac{\mathrm{Re}\ \{d_{kk'}\}}{\gamma_{1k}\gamma_{2k'}}\times\exp\left[-\ln(2)\left(\frac{\omega_1-v_{1k}}{\gamma_{1k}}\right)^2\right]\times\exp\left[-\ln(2)\left(\frac{\omega_2-v_{2k'}}{\gamma_{2k'}}\right)^2\right], \tag{13}$$

where $v_{pk}=\mathrm{Re}\ \omega_{pk}$ define the line positions and $\gamma_{pk}=-\mathrm{Im}\ \omega_{pk}$, their widths. The last expression converts the star-shaped contours of the 2D Lorentzian peaks into clean elliptical contours that are more easily discerned, while conserving the peak volume and full width at half-height. As there is hardly any decay in the CT dimensions, theoretically γ_{1k} and γ_{2k} must be nearly zero. Numerically they are very small and may have arbitrary sign. In such a case smooth spectra are obtained using Eqs. (12) and (13) with γ_{1k} and γ_{2k} replaced by sufficiently large positive number Γ.

Any other 2D spectral projection, e.g., along the second and third dimensions

$$I(\omega_2,\omega_3):=\sum_{n_2=1}^{\infty}\sum_{n_3=1}^{\infty}S(0,n_2\tau_2,n_3\tau_3)z_2^{-n_2}z_3^{-n_3}, \tag{14}$$

can be constructed by modifying Eqs. (7), (8), and (10)–(13) accordingly.

The FT spectral estimation by Eq. (6) requires only a 2D slice ($n_3=0$) of the 3D data set. Theoretically, the same 2D spectrum may be obtained by first calculating the 3D FT and then simply integrating over the third frequency variable. This reduces noise, but does not affect the resolution of the two-dimensional plane. With 3D FDM2K, however, the maximum information contained in the whole 3D array is utilized. This makes an enormous difference in the case when both N_1 and N_2 are small and N_3 is large. According to Eq. (1) a high resolution FT spectral estimation requires both N_2 and N_3 to be sufficiently large, while the FDM2K spectral reconstruction depends on the product $(2N_1)\times(2N_2)\times N_3$ being sufficiently large.

The explicit expressions to evaluate the $\mathbf{U}_p$ matrices and the vector **C** have been published [8]. The change in the data format to $t_p\in[-T_p;T_p]$ $(p=1,2)$, rather than the usual format $t_p\in[0;T_p]$, only affects the definition of time zero for t_1 and t_2 and is easily accommodated by introducing appropriate phase-correction terms to the matrix elements of $\mathbf{U}_p$ and **C** [12,20].

3. Results and discussion

The FT spectra were obtained using cosine weighting functions along all dimensions to reduce truncation artifacts. The four 3D data sets, each of size $N_1\times N_2\times N_3=6\times 8\times 1024$ (^{13}C′, ^{15}N, and ^{1}H) were zero-filled to $128\times 128\times 2048$ to provide sufficient digital resolution for smooth contouring. The FDM spectra were obtained using the FDM2K algorithm with an optimized regularization level adjusted to produce stable, resolved results [7]. With the *quadrupling scheme* the four sets were combined together to yield a single phase modulated 3D signal with an effective matrix size of $11\times 15\times 1024$, of which $10\times 14\times 1024$ could be used. In the proton dimension, fully decayed signals were acquired, meaning the tail contains predominately noise. Using FT processing, the extra points in the time domain in combination with proper apodization will marginally improve digital resolution without sacrificing sensitivity. However, experience has shown that the noisy tail does not contain much useful *information*, and so is best not

included in FDM calculations. The results shown here were computed using only the first 500 complex data points in the proton dimension. In the proton dimension signals are spread over a small area so that a reduced frequency window of 2000 Hz was kept, while for the two indirect dimensions the entire spectral range was used. The basis sizes along the $^{13}C'$ and ^{15}N dimensions were $Nb_1 = 10/2 = 5$ and $Nb_2 = 14/2 = 7$, respectively. For each local frequency window in the proton dimension $Nb_3 \sim 500/2 \times 2000\,\mathrm{Hz}/SW = 14$ basis functions were used. Thus, the total size of the basis for each 3D window was $5 \times 7 \times 14 = 490$. Six overlapping windows were used to cover the spectral regions depicted in Figs. 2–4. The total calculation took about 40 min on a Linux PC equipped with an AMD XP 1800+ CPU. Using between 400 and 600 proton points produced essentially identical spectra, pointing to convergence of the method.

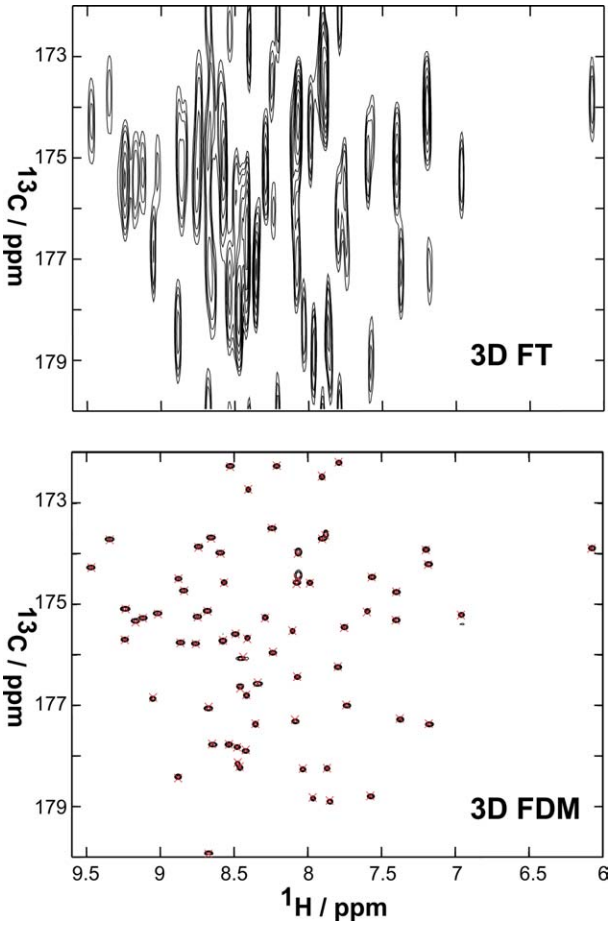

Fig. 2. Comparison of 2D (^1H, $^{13}C'$) projections of the 3D FT and FDM spectra obtained from a 3D CT-HNCO signal with 6×8 increments along the $^{13}C'$ and ^{15}N CT dimensions. The red crosses indicate the HNCO peak positions predicted from high resolution 3D FT spectra using far more data (see text). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

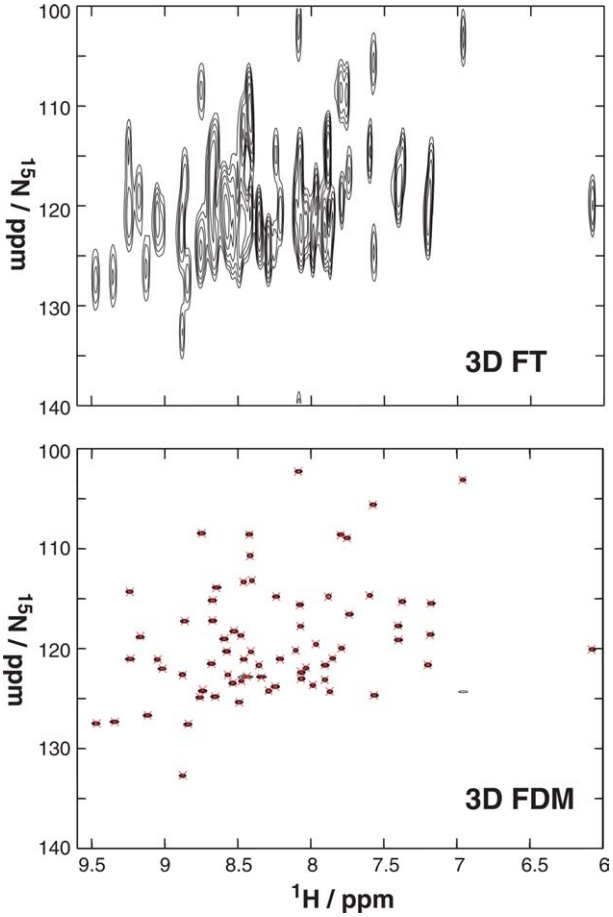

Fig. 3. Corresponding (^1H, ^{15}N) projections of the 3D HNCO data sets shown in Fig. 2.

As with FT processing, FDM can produce full 3D spectra, although when the resolution is very high, representative 2D projections may be preferred, as the high-resolution 3D spectrum is of enormous size. The corresponding contour plots of the three orthogonal 2D projections computed by Eq. (13) are shown in Figs. 2–4. As the lines can be arbitrarily narrow in the CT dimensions, they were properly smoothed for better plotting by enforcing minimum line widths of 4 and 6 Hz in the ^{13}C and ^{15}N dimensions, respectively. To assess the accuracy and reliability of the FDM2K projections, the ^1H, ^{15}N, and $^{13}C'$ chemical shifts of the peaks in the CT-HNCO were independently verified using a corresponding high resolution 3D FT spectrum. The sequential backbone assignments were obtained from 3D HNCA/HN(CO)CA and CBCA(CO)NH/HNCACB experiments. The deduced HNCO peak positions of the FT spectrum are indicated by the red crosses in Figs. 2–4. FDM peak positions agree extremely well with the high resolution FT data, suggesting good reliability of the method. All expected peaks, including weak signals, can be observed in the spectrum recorded in 25 min. (The unmarked peak in the FDM spectra at approxi-

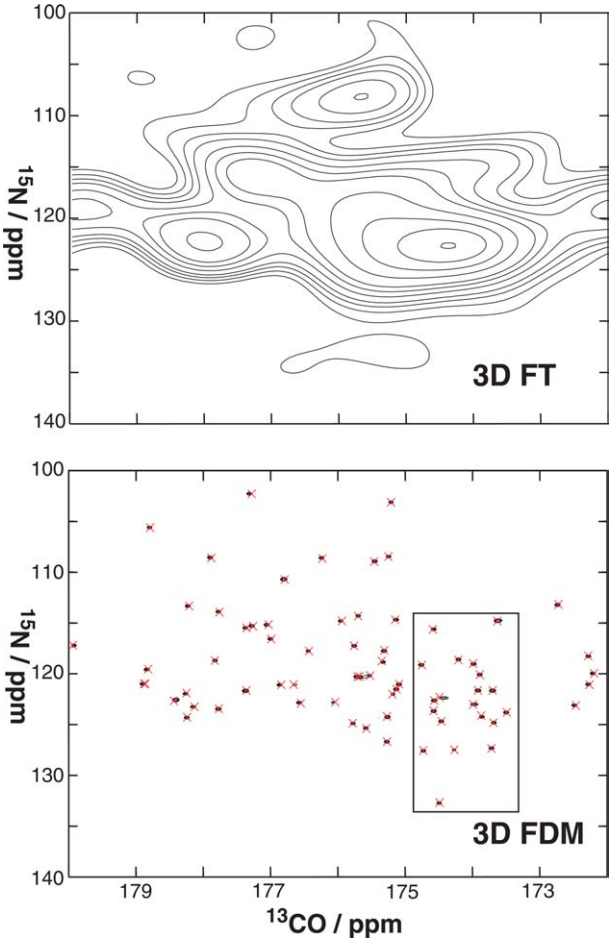

Fig. 4. Same as in Fig. 2 but showing the (^{13}C′, ^{15}N) projections. In the FDM spectrum each correlation is distinctly resolved, while the FT spectrum does not allow individual peaks to be distinguished.

mately 7 ppm of the ^{1}H chemical shift also appears, albeit broader and with lower intensity in the FT spectrum: it arises from a rather poorly decoupled arginine side chain resonance.) It is noteworthy that the large majority of the peaks are fully resolved in the FDM projections even in the (^{13}C′, ^{15}N) plane (lower panel of Fig. 4 and its expansion, Fig. 5), for which the size of the data is only 6×8 points. For such a small data size, one typically sees only completely unresolved spectra with vast amounts of overlap, as visualized in the upper panel of Fig. 4. This remarkable result shows how the resolution in every dimension of the multi-dimensional FDM2K spectrum is defined by the total size of the data matrix. It is important to emphasize that this dramatic resolution enhancement through FDM2K is the result of a combination of several factors: (i) the inherent multi-dimensional character of the method; (ii) enforcement of Lorentzian lineshapes with the CT evolution protocol; and (iii) the data size quadrupling scheme. Under these conditions, the small number of points in the indirect dimensions is still sufficient to provide a

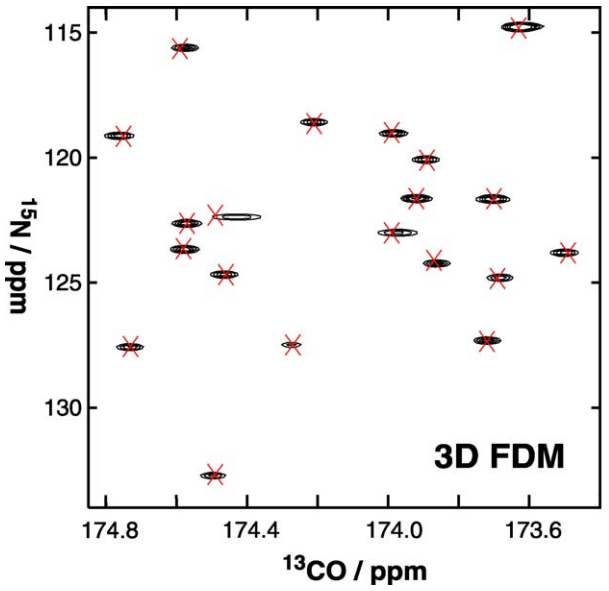

Fig. 5. Expansion of the rectangular region of the (^{13}C′, ^{15}N) projection shown in Fig. 2.

highly resolved spectrum, allowing one simply to reduce the total experiment time and/or to use short CT-periods to prevent excessive sensitivity losses through fast transverse relaxation.

3.1. Validation of the FDM2K Results

In practice, it may be difficult to validate the FDM results directly, particularly in the absence of a comparable reference. The resolution can be so enormously improved that it is difficult to compare the spectra with those that are usually obtained by other methods. However, there may be ways to validate the results indirectly [20]. For example, as mentioned above, the results are independent of small variations of various FDM2K parameters such as window size, basis density, regularization parameter and signal size (in the proton dimension). An auxiliary validation is to estimate the finite DFT from the computed FDM2K spectral parameters, $d_{kk'}$, u_{1k}, and $u_{2k'}$, and compare to the actual finite DFT results. (Note that Figs. 2–4 show the FDM estimations of the infinite-time DFT spectra.) To this end we replace Eq. (12) by a double-absorption form which is directly comparable to that obtained by the finite DFT over a rectangular time domain of size $N_1 \times N_2$

$$A^{N_1,N_2}(\omega_1,\omega_2) = \sum_{k,k'} \mathrm{Re}\,\{d_{kk'}\}\mathrm{Re}\left\{\frac{1-(u_{1k}/z_1)^{N_1}}{1-u_{1k}/z_1}-\frac{1}{2}\right\} \times \mathrm{Re}\left\{\frac{1-(u_{2k}/z_2)^{N_2}}{1-u_{2k}/z_2}-\frac{1}{2}\right\}. \quad (15)$$

The other 2D spectral projections are obtained by changing the indices $(1, 2)$ appropriately. Apodization can also be included. In particular, exponential or trigonometric apodization functions give simple analytic forms that can be included in the lineshape function without any difficulty. All the following results were obtained with cosine window functions in all dimensions, implemented numerically in the actual DFT calculations and analytically in the FDM2K estimations. Fig. 6 compares the FDM2K estimation of the finite DFT projection in the ($^{13}C'$, ^{1}H) plane with $(N_1, N_3) = 6 \times 1024$ to the actual FT projection obtained from the data of the same size. Here the FDM2K spectral parameters were the same as those used to produce Figs. 2–4. The two spectra match well, showing that FDM2K can provide a consistent fit of the experimental data. Similar agreement was obtained for the other two projections (data not shown). Further verification of the FDM2K fit, using the same set of spectral parameters, can be obtained by estimating the finite DFT spectrum of a (hypothetical) larger signal and comparing it to the actual DFT result of a larger data set. While FDM2K results are excellent for each of the projections (see Figs. 2–4), the ($^{13}C'$, ^{15}N) plane is the most difficult to resolve by DFT. Therefore, it is most interesting to validate the corresponding FDM2K result shown in Fig. 4. In Fig. 7, we compare the FDM2K estimation (using the same short signal) of the finite DFT expected with $(N_1, N_2) = 12 \times 32$ to an actual DFT of data set of this size. Such a comparison necessitated another CT-HNCO experiment with a longer CT period $2T = 24$ ms in the $^{13}C'$ dimension as the number of increments would exceed the $2T = 8$ ms constant time period. Transverse relaxation during the longer constant time period, caused the intensities in the actual FT projection to be lower, but otherwise the two projections match well. This comparison provides further strong evidence that the fit computed by FDM2K for a small data set is reliable and can be used for substantial resolution enhancement.

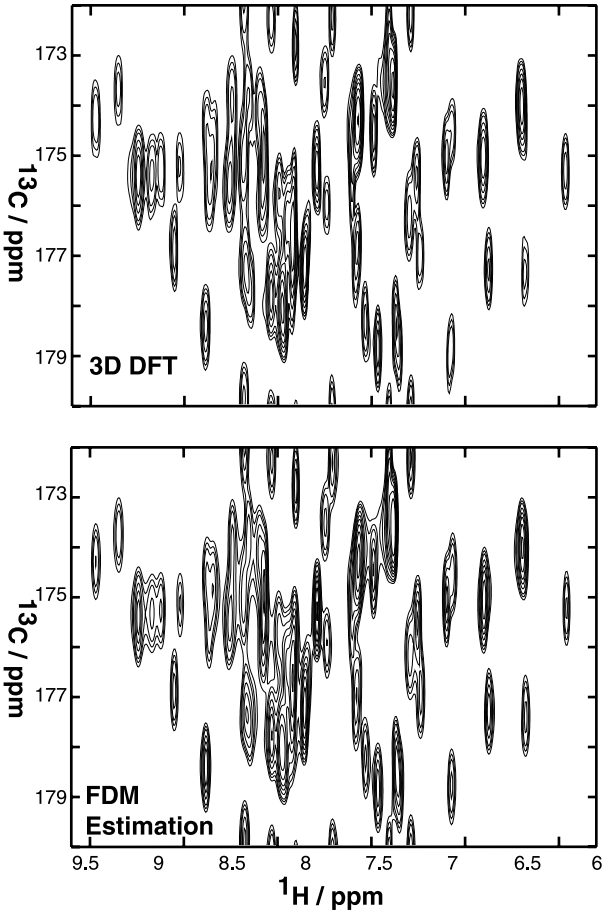

Fig. 6. Comparison of the FDM2K estimation of the finite DFT with $(N_1, N_3) = (6, 1024)$ of the ($^{13}C'$, ^{1}H) projection with the actual DFT obtained from the original data. Both methods used the same 3D data set. The agreement between the two spectra validates the consistency of the FDM2K fit.

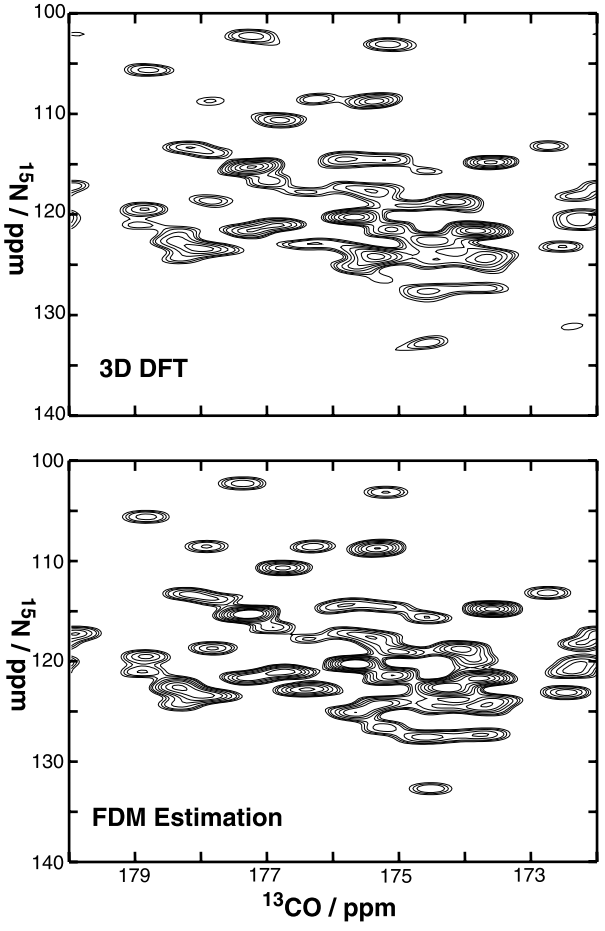

Fig. 7. Comparison of the FDM estimation of the DFT ($^{13}C'$, ^{15}N) projection with the actual DFT corresponding to $(N_1, N_2) = 12 \times 32$. The FDM spectrum is constructed using the spectral parameters obtained from the short data set of size $(N_1, N_2, N_3) = 6 \times 8 \times 1024$. The DFT spectrum was computed using a different and larger data set of size $(N_1, N_2, N_3) = 12 \times 32 \times 1024$.

4. Conclusions

In this article, FDM2K was applied to a slightly modified 3D CT-HNCO experiment, as an example of a member of the suite of backbone correlation experiments, that encodes the $^{13}C'$ and ^{15}N chemical shifts in a CT manner. The time-reversal symmetry of the CT signal makes it particularly suitable for FDM2K. Unprecedented resolution was obtained using extremely short 3D data sets with a small number of increments, recorded in only 25 min on a small double-labeled protein (human ubiquitin) at a realistic concentration with a conventional NMR probe at 500 MHz. The FDM2K calculation was fast and provided reliable results that were confirmed through an independent consistency test.

Multi-dimensional triple-resonance experiments of biological molecules are typically recorded under conditions requiring a compromise between adequate spectral resolution and sufficient overall sensitivity. Typically, therefore, the interferogram sampled in the indirect dimensions is truncated, resulting in poor frequency resolution of the FT spectra in these corresponding dimensions. Although an extension of the signal by linear prediction may improve the resolution, generally the improvement is not significant. In such cases, FDM can offer substantial advantages over the conventional techniques. The results presented apply also generally to other triple-resonance experiments where chemical shift evolution can be implemented during CT periods.

If the overall sensitivity is not limiting, e.g., when dealing with concentrated samples, more sensitive cryoprobe technology, or higher magnetic field strengths, then high resolution 3D spectra can be obtained very rapidly, by recording just a few increments. In this high-throughput mode, FDM2K permits drastic reduction of the total experiment time, while simultaneously providing adequate resolution.

Alternatively, if fast transverse relaxation prohibits the use of long CT periods, FDM2K allows higher resolution to be obtained from a briefer CT interval, without sacrificing sensitivity. This issue becomes particularly important for larger proteins with faster relaxing nuclei, e.g., in HNCA [21] experiments. Here the improved total resolution can be obtained simply by sampling additional points in dimensions where the chemical shifts of slower relaxing nuclei evolve, e.g., during TROSY relaxation-compensated periods.

We have shown that well-resolved 3D spectra can be obtained very rapidly, and that the chemical shifts that are extracted are accurate. Of course, the HNCO experiment is not the most important among the usual suite of 3D experiments that are employed for assignment purposes. The HNCA experiment is more challenging because twice as many carbon resonances are usually obtained. In addition, the shifts must be accurate and consistent so that sequential assignments can be made. In non-deuterated medium-sized proteins, the potential to obtain high resolution along the C-α dimension without a corresponding long evolution time will be an important development. We will describe this and related results in future publications.

Acknowledgments

This work was supported by the National Institutes of Health, R01 GM066763-01. V.A.M. is an Alfred P. Sloan research fellow.

References

[1] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, J. Am. Chem. Soc. 125 (2003) 1385–1393.
[2] E. Kupce, R. Freeman, Projection-reconstruction of three-dimensional NMR spectra, J. Am. Chem. Soc. 125 (2003) 13958–13959.
[3] J.C. Hoch, A.S. Stern, NMR Data Processing, Wiley-Liss, New York, 1996.
[4] M.R. Wall, D. Neuhauser, Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode frequencies from a small number of residues or a short-time segment of a signal. I. Theory and application to a quantum-dynamics model, J. Chem. Phys. 102 (1995) 8011–8022.
[5] V.A. Mandelshtam, H.S. Taylor, Harmonic inversion of time signals and its applications, J. Chem. Phys 107 (1997) 6756–6769.
[6] V. Mandelshtam, H. Hu, N.D. Taylor, M. Smith, A.J. Shaka, Highly resolved double absorption 2D NMR spectra from complex severely truncated 2D phase modulated signals by filter-diagonalization-averaging method, Chem. Phys. Lett. 305 (1999) 209–216.
[7] J. Chen, V.A. Mandelshtam, A.J. Shaka, Regularization of the two-dimensional filter diagonalization method: FDM2K, J. Magn. Reson. 146 (2000) 363–368.
[8] V.A. Mandelshtam, FDM: the filter diagonalization method for data processing in NMR experiments, Prog. NMR Spectrosc. 38 (2001) 159–196.
[9] A. Bax, A.F. Mehlkopf, J. Smidt, Homonuclear broadband-decoupled absorption spectra, with linewidths which are independent of the transverse relaxation rate, J. Magn. Reson. 35 (1979) 373–377.
[10] A. Bax, R. Freeman, Investigation of complex networks of spin–spin coupling by 2D NMR, J. Magn. Reson. 44 (1981) 542–561.
[11] P. Powers, A.M. Gronenborn, G.M. Clore, A. Bax, 3-dimensional triple-resonance NMR of C-13/N-15-enriched proteins using constant-time evolution, J. Magn. Reson. 94 (1991) 209–213.
[12] J. Chen, A.A. De Angelis, V.A. Mandelshtam, A.J. Shaka, Progress on two-dimensional filter diagonalization method, J. Magn. Reson. 162 (2003) 74–89.
[13] E. Kupce, J. Boyd, I.D. Campbell, Short selective pulses for biochemical applications, J. Magn. Reson. Ser. B 106 (1995) 300–303.
[14] A.J. Shaka, C.J. Lee, A. Pines, Iterative schemes for bilinear operators—application to spin decoupling, J. Magn. Reson. 77 (1988) 274–293.

[15] L. Emsley, G. Bodenhausen, Experimental aspects of chirped NMR-spectroscopy, J. Magn. Reson. Ser. A 102 (1993) 293–301.
[16] A.J. Shaka, J. Keeler, R. Freeman, Evaluation of a new broadband decoupling sequence—WALTZ-16, J. Magn. Reson. 53 (1983) 313–340.
[17] M.A. McCoy, L. Mueller, Selective shaped pulse decoupling in NMR—homonuclear [C-13]carbonyl decoupling, J. Am. Chem. Soc. 114 (1992) 2108–2112.
[18] D. Marion, M. Ikura, R. Tschudin, A. Bax, Rapid recording of 2D NMR-spectra without phase cycling—application to the study of hydrogen-exchange in proteins, J. Magn. Reson. 85 (1989) 393–399.
[19] L.E. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, J. Am. Chem. Soc. 114 (1992) 10663–10665.
[20] J. Chen, Nolinear Methods for High Resolution Spectral Analysis and Their Applications in Nuclear Magnetic Resonance Experiments, PhD thesis, University of California at Irvine, 2002. Available from <http://www.physics.uci.edu/~jianhanc/DOC/PhD.pdf>.
[21] L.E. Kay, M. Ikura, R. Tschudin, A. Bax, Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins, J. Magn. Reson. 89 (1990) 496–514.