

High-resolution pyrimidine- and ribose-specific 4D HCCH-COSY spectra of RNA using the filter diagonalization method

Justin T. Douglas · Michael P. Latham ·
Geoffrey S. Armstrong · Brad Bendiak ·
Arthur Pardi

Received: 9 May 2008 / Accepted: 24 June 2008 / Published online: 15 July 2008
© Springer Science+Business Media B.V. 2008

Abstract The NMR spectra of nucleic acids suffer from severe peak overlap, which complicates resonance assignments. 4D NMR experiments can overcome much of the degeneracy in 2D and 3D spectra; however, the linear increase in acquisition time with each new dimension makes it impractical to acquire high-resolution 4D spectra using standard Fourier transform (FT) techniques. The filter diagonalization method (FDM) is a numerically efficient algorithm that fits the entire multi-dimensional time-domain data to a set of multi-dimensional oscillators. Selective 4D constant-time HCCH-COSY experiments that correlate the H5–C5–C6–H6 base spin systems of pyrimidines or the H1'–C1'–C2'–H2' spin systems of ribose sugars were acquired on the ¹³C-labeled iron responsive element (IRE) RNA. FDM-processing of these 4D experiments recorded with only 8 complex points in the indirect dimensions showed superior spectral resolution than FT-processed spectra. Practical aspects of obtaining optimal FDM-processed spectra are discussed. The results here demonstrate that FDM-processing can be used to obtain

high-resolution 4D spectra on a medium sized RNA in a fraction of the acquisition time normally required for high-resolution, high-dimensional spectra.

Keywords Filter diagonalization method · Multi-dimensional NMR · RNA · HCCH-COSY · High dimensionality

Introduction

The high degree of spectral overlap of the base and sugar resonances represents a major challenge for NMR studies of nucleic acids (Wijmenga and van Buuren 1998). Thus, studies of even moderately-sized DNA or RNA oligomers normally require isotopic labeling and 3D NMR spectra for complete resonance assignment (Pardi 1995). Although 4D (or higher-dimensional) experiments have the potential to greatly simplify NMR studies of nucleic acids (Nikonowicz and Pardi 1992), each additional dimension results in a multiplicative increase in experimental time using standard FT techniques. Thus, 4D NMR experiments are generally collected with only a limited number of points in each indirect dimension, which leads to broad lines that can limit the advantage of the additional dimension(s). Recently, various methods have been developed with the goal of obtaining high-resolution, high-dimensional NMR experiments including: G-matrix Fourier transform (GFT) (Kim and Szyperski 2003; Kim and Szyperski 2004), projection reconstruction (Kupce and Freeman 2003, 2004), covariance (Bruschweiler 2004; Bruschweiler and Zhang 2004), multi-dimensional decomposition (Orekhov et al. 2003), maximum entropy reconstruction (Stephenson 1988), non-linear sampling with maximum entropy reconstruction (Schmieder et al. 1993; Hyberts et al.

Justin T. Douglas and Michael P. Latham contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9253-3) contains supplementary material, which is available to authorized users.

J. T. Douglas · M. P. Latham · G. S. Armstrong · A. Pardi (✉)
Department of Chemistry and Biochemistry, 215 UCB,
University of Colorado, Boulder, CO 80309-0215, USA
e-mail: arthur.pardi@colorado.edu

B. Bendiak
Department of Cell and Developmental Biology and
Biomolecular Structure Program, University of Colorado Health
Sciences Center, Mail Stop 8108, P.O. Box 6511, Aurora, CO
80045, USA

2007), polar Fourier transform (Coggins and Zhou 2006), and the filter diagonalization method (FDM) (Hu et al. 1998; Mandelshtam et al. 1998).

The FDM differs from most of these other techniques because it employs true multi-dimensional processing of the time-domain data. FDM utilizes linear algebra techniques to generate the multi-dimensional spectrum from the entire multi-dimensional time-domain data set (Mandelshtam 2000). Because the FDM fits the entire time-domain data simultaneously each dimension is not processed independently, as is the case in FT processing. Thus, the large number of data points usually acquired in the direct dimension can improve the spectral resolution in the indirect dimensions that have been sampled with only a small number of data points (Mandelshtam et al. 1998). Since NMR data processed with FDM do not require as large numbers of time-domain points in the indirect dimensions, high-resolution, high-dimensional NMR spectra can be acquired in shorter times compared to what would be required for high-resolution spectra processed by the FT. For example, Shaka et al. have reported high quality ^1H , ^{15}N HSQC spectra recorded with as few as 4 points in the indirect dimension (Hu et al. 2000). The FDM methodology has been successfully applied to proteins (Hu et al. 2000; Chen et al. 2003, 2004) and oligosaccharides (Armstrong et al. 2004, 2005) to process 3D and 4D NMR experiments.

Previous applications of FDM-processing to NMR spectroscopy have shown that incorporating several simple pulse sequence modifications increases the efficiency and robustness of the approach (Mandelshtam et al. 1998; Chen et al. 2000; Hu et al. 2000; Mandelshtam 2000). These modifications still allow standard FT processing of the data and include: selective excitation, constant-time evolution of the indirect dimensions and pure in-phase signals in the indirect dimensions. FDM-processing of the experimental data is conceptually different than FT processing because the entire time-domain data is used for the fitting of multi-dimensional time-domain sinusoids (Hu et al. 1998). Analogous to empirical optimization of window functions or linear prediction in FT processing, the FDM algorithm contains parameters that need to be empirically optimized to obtain the highest quality spectrum.

FDM-processing was applied here to overcome spectral overlap in the crowded aromatic and sugar regions of a ^{13}C -labeled RNA. A selective HCCH-COSY experiment (Bax et al. 1990; Kay et al. 1990) was used to produce high-resolution 4D spectra for resonance assignments in a model 29 nucleotide ^{13}C -labeled RNA, the iron responsive element (IRE) (Fig. 1a, Address et al. 1997; McCallum and Pardi 2003). This experiment was used to correlate the H5–C5–C6–H6 spins in C and U bases, taking advantage of large $^1J_{\text{CH}}$ (180 Hz) and $^1J_{\text{CC}}$ (65 Hz) coupling constants to

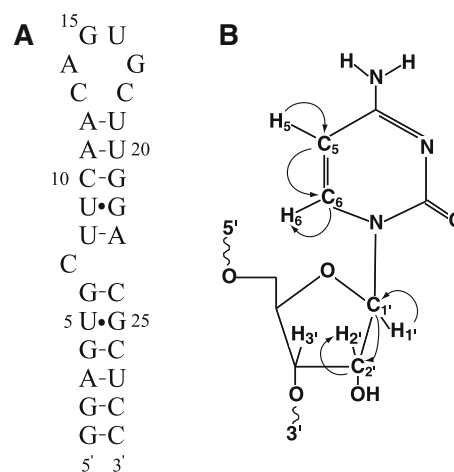


Fig. 1 (a) The sequence and secondary structure of the iron responsive element RNA is shown. (b) The magnetization transfer pathways for the pyrimidine- and sugar-specific HCCH-COSY experiments are illustrated for cytidine

achieve efficient magnetization transfer (Fig. 1b, Wijmenga and van Buuren 1998). Comparisons of FT- and FDM-processing of a 4D data set recorded with only 8 complex points in each of the indirect dimensions show that the FDM-processed spectra have much narrower lines and superior spectral resolution. A sugar-specific 4D HCCH-COSY experiment that correlates the H1'–C1'–C2'–H2' spin systems of the ribose sugars was also collected on the IRE. This application represents a greater challenge due to the higher degree of spectral overlap of the sugar resonances in RNAs (Wijmenga and van Buuren 1998), and again FDM-processed spectra show far superior spectral resolution. The practical effects of signal-to-noise, total number of points sampled in the indirect dimensions, phase errors and processing parameters on the FDM-processed spectra are discussed.

Materials and methods

NMR data collection and analysis

All experiments were recorded using a previously prepared 1.0 mM sample of uniformly ^{13}C , ^{15}N -labeled IRE RNA in 99% D_2O , 10 mM sodium phosphate, 10 mM sodium chloride, 0.1 mM EDTA at pH 6.5 (McCallum and Pardi 2003). All NMR data were recorded at 25°C on a Varian Inova 500 MHz spectrometer equipped with a triple resonance z -axis pulsed-field gradient probe. The nmrPipe software package was used for the FT processing (Delaglio et al. 1995). 4D FDM processing was implemented using software written in-house and described previously (Armstrong et al. 2005) based on 3D FDM software provided by Jianhan Chen and Vladimir Mandelshtam (UC Irvine). The

FORTRAN source code can be downloaded from <http://cunmr800.colorado.edu/software.html>. The spectra were analyzed using CCPN Analysis (Vranken et al. 2005) and Sparky (Goddard and Kneller 1998). Data were processed on a 3 GHz Dual-Core Intel Xeon MacPro with 4 GB of RAM.

A pyrimidine-specific 4D constant-time HCCH-COSY experiment was performed on the IRE with 8 transients per FID, an interscan delay of 2.0 s, 8 complex points in each indirect dimension, and 1024 complex points in the directly acquired dimension for a total acquisition time of ~20 h. Spectral widths of 2000, 4000, 2000 and 6000 Hz were employed in the H5, C5, C6 and H6 dimensions, respectively and the proton and carbon carriers were set to 4.78 ppm in H5 and H6, 100.0 ppm in C5 and 139.8 ppm in C6. A 4D spectrum was generated using standard FT processing techniques. For the three indirect dimensions, the size of the time-domain was doubled using mirror-image linear prediction (Zhu and Bax 1990), a squared cosine-bell window function was applied to the linear predicted data and then zero-filled to 128, 64, 32 and 1024 complex points in H5, C5, C6 and H6, respectively, prior to FT. Linear prediction was employed in an indirect dimension only after the other dimensions had been transformed, as previously described (Kay et al. 1991). Only the left half of the spectrum was saved for the H6 dimension, giving a total matrix size of $128 \times 64 \times 32 \times 512$. The FT processing of the 4D data set required ~13 min of CPU time.

The optimal FDM-processed spectrum was obtained by using all 8 complex points in each indirect dimension, but only the first 256 complex points in the direct dimension. The FDM calculation was performed on only part of the full spectrum, consisting of the full sweep width in the three indirect dimensions, but only the 1000 Hz region from ~6.8 to 8.8 ppm in the directly acquired dimension. In practice the FDM calculation is divided into a number of equally spaced and overlapping spectral regions (Hu et al. 1998; Mandelshtam et al. 1998). Dividing the data set into sub-sets simplifies the FDM fitting procedure and reduces the overall time of the calculation (Armstrong et al. 2005). For the constant-time HCCH-COSY experiments, the FDM calculation was divided into 11 windows in the direct dimension and 3 windows in the each of the indirect dimensions, resulting in 297 overlapping windows each with an size of ~1000 × 2000 × 1000 × 175 Hz in the H5, C5, C6 and H6 dimensions, respectively. The optimal regularization parameter, q^2 , was empirically found to be 0.0005 and the optimal smoothing parameter, Γ , was found to be 10 Hz in H5, 15 Hz in C5 and C6 and 4 Hz in H6. The reconstructed FDM matrix size is $128 \times 64 \times 32 \times 256$. FDM-processing of the 4D data required ~36 min. A -15° and -10° zero-order phase correction

was required in the C5 and C6 dimensions, respectively for both the FT and FDM processed spectra. These phase corrections were performed on the time-domain data prior to the FDM calculations using the nmrPipe software program (Delaglio et al. 1995).

The sugar-specific 4D constant-time HCCH-COSY was acquired as described above except with an interscan delay of 1.8 s for a total acquisition time of ~18.5 h. Spectral widths of 2000, 1500, 1500 and 6000 Hz were employed in the H1', C1', C2' and H2' dimensions, respectively, and the proton and carbon carrier frequencies were set to 4.79 ppm in H1' and H2', 90.1 ppm in C1' and 74.2 ppm in C2'. The same FT- and FDM-processing schemes as described above were employed except the FDM-processing was performed on a 1025 Hz region from ~3.4 to 5.4 ppm in the directly acquired dimension and the smoothing parameter was set to 15 Hz for H1'. No phase corrections were required for the sugar-specific HCCH-COSY data. Processing parameters for all FDM experiments are given in Supplementary Table S1.

Results

Constant-time HCCH-COSY pulse sequence

The constant-time HCCH-COSY experiment employed here (Fig. 2) incorporates minor modifications from the standard pulse sequence (Bax et al. 1990; Kay et al. 1990). The modifications introduced for subsequent FDM-processing are: (1) selective EBURP1 shaped 90° pulses (Geen and Freeman 1991), which are used to initially excite the H5 (H1') resonances and then to further direct the magnetization to the C5 (C1') resonances. The use of selective excitation allows for narrower sweep widths in the indirect dimensions. A soft square 180° pulse that has an excitation maximum for the C6 (C1') resonance and a null on the C5 (C2') resonances is used during the final INEPT transfer to emphasize the selected magnetization transfer pathway. (2) Frequency labeling in all of the indirect dimensions employs constant-time evolution periods. This is because an efficient procedure has been developed for FDM processing that uses the quadrature pairs in the constant-time evolution period to effectively double the number of data points in the indirect dimension, in a manner directly analogous to mirror-image linear prediction (Zhu and Bax 1990; Chen et al. 2003). (3) The use of gradients is minimized to help reduce phase distortions, which can lead to frequency shifts and artifacts in FDM-processed spectra (see below). The pyrimidine-specific HCCH-COSY experiment selectively transfers magnetization through the H5–C5–C6–H6 spin system of the U and C bases (Fig. 1b). Scalar coupling to the C4 is refocused via selective ^{13}C I-

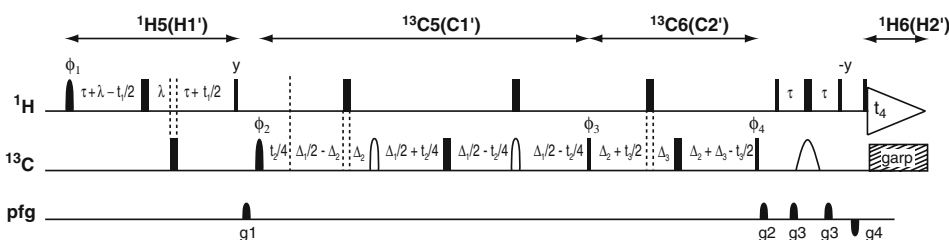


Fig. 2 Modified pulse sequence for the selective 4D constant-time HCCH-COSY (Bax et al. 1990; Kay et al. 1990) experiments used to correlate the H5–C5–C6–H6 spin systems of pyrimidines or the H1'–C1'–C2'–H2' spin systems of ribose. The nucleus that is evolving during the t_1 , t_2 and t_3 constant-time periods, and the t_4 acquisition period, are indicated above. Narrow and wide bars represent full power square 90° and 180° pulses, respectively. Filled bell shapes represent band selective EBURP1 90° pulses (Geen and Freeman 1991). The narrow open bell shapes represent off-resonance $^{13}\text{C}4$ selective IBURP2 180° pulses (Geen and Freeman 1991) and the wide open bell shape represents a soft-square 180° pulse with an excitation maximum on-resonance for C6 (C2') and the first null of the excitation profile on C5 (C1'). For the 4D pyrimidine-specific experiment the delay times were: $\tau = 1.32$ ms, $\lambda = 800$ μs , $\Delta_1 = 3.62$ ms and $\Delta_2 = 1.32$ ms and $\Delta_3 = 2.30$ ms. For the 4D sugar-specific experiment the delay times were: $\tau = 1.47$ ms, $\lambda = 1.54$ ms, $\Delta_1 = 5.55$ ms, $\Delta_2 = 1.47$ ms and $\Delta_3 = 1.31$ ms. For both experiments, gradient powers and durations have the following values: $g_1 = 43.1$ G/cm, 600 μs , $g_2 = 23.7$ G/cm,

1.04 ms, $g_3 = 23.7$ G/cm, 800 μs and $g_4 = -43.1$ G/cm, 500 μs . ^{13}C GARP decoupling was applied during the acquisition period. Presaturation prior to the first proton pulse was used to suppress the residual water signal. The ^1H EBURP1 pulse was applied with a bandwidth of 1050 Hz and an offset of 320 Hz and the ^{13}C EBURP1 pulse was applied on-resonance with a bandwidth of 2000 Hz and 1625 Hz for the pyrimidine- and sugar-specific experiments, respectively. For the pyrimidine-specific experiment, the IBURP2 pulses were applied 8900 Hz off-resonance with a bandwidth of 5000 Hz and were not used in the sugar-specific experiment. The delays t_1 , t_2 and t_3 represent the chemical shift evolution time for H5 (H1'), C5 (C1') and C6 (C2'), respectively. Phase cycling is $\phi_1 = (x, -x)$, $\phi_2 = (2(x), 2(-x))$, $\phi_3 = (x, -x)$, $\phi_4 = (4(x), 4(-x))$, receiver = $(x, -x, -x, x, -x, x, x, -x)$. All pulses were applied along the x-axis unless otherwise noted. Quadrature detection in each of the indirect dimensions was achieved by incrementing ϕ_1 for t_1 , ϕ_2 for t_2 and ϕ_3 for t_3 along with the receiver according to the States-TPPI protocol (Marion et al. 1989)

BURP 180° pulses (Geen and Freeman 1991) applied off-resonance at the C4 frequency. The sugar-specific HCCH-COSY experiment selectively transfers magnetization through the H1'–C1'–C2'–H2' ribose spin system (Fig. 1b). Scalar coupling between C2' and C3' cannot be refocused due to their similar chemical shifts; thus, the C2' coherences evolve under scalar coupling to both the C1' and C3'. This $^1\text{J}_{\text{C}2'-\text{C}3'}$ coupling reduces the sensitivity of the sugar-specific experiment compared to the pyrimidine-specific experiment.

Acquisition and processing of the pyrimidine-specific 4D HCCH-COSY spectrum

The pyrimidine-specific 4D HCCH-COSY was performed on the ^{13}C , ^{15}N -labeled IRE RNA (Fig. 1a) with 8 complex points in each indirect dimension. Standard FT processing yields a 4D spectrum with large linewidths in all indirect dimensions, leading to three peaks in the 2D plane shown in Fig. 3a, due to leakage of peaks from adjacent planes. FDM-processing of the same 4D data set shows only one peak in this 2D plane (Fig. 3b) where the other two peaks are now resolved into individual planes (Fig. 3c and d). Thus, the U9, C10 and C29 peaks are resolved in the 4D FDM-processed spectrum even though they have similar H5 and C6 chemical shifts. Overall, 13 of the 15 pyrimidine H5, C5, C6, H6 spin systems in the IRE have distinct chemical shifts and 9 of these are resolved into individual planes with FDM-processing, whereas only 1 spin system

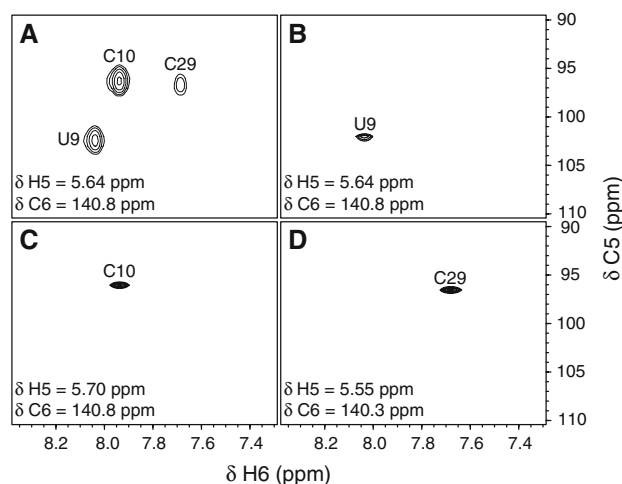


Fig. 3 Individual 2D H6–C5 planes of the pyrimidine-specific 4D HCCH-COSY spectrum of the IRE RNA generated with (a) FT processing or (b), (c) and (d) FDM-processing. The 4D experiment was acquired in 20 h with $8 \times 8 \times 8$ complex points in the indirect dimensions with 8 scans per FID. The chemical shifts in the other dimensions for the 2D plane are denoted in the lower left of each panel and assignments are from Address et al. (1997)

is resolved into an individual plane in the FT-processed spectrum. This increased resolution arises from the reduced apparent linewidths in the indirect dimensions for the FDM spectrum (on average 28, 67 and 66 Hz for the H5, C5 and C6 resonances, respectively) compared to the FT spectrum (on average 134, 241 and 125 Hz for the H5, C5 and C6 resonances, respectively). The FDM algorithm produces

high-resolution estimates of frequencies and amplitudes of resonances, which are then visualized in the reconstructed spectrum (Mandelsham 2000). To keep the 4D matrix at a manageable size, the spectrum was reconstructed with a digital resolution of 3.9 Hz/pt (256 pts), 15.6 Hz/pt (128 pts), 63 Hz/pt (64 pts) and 63 Hz/pt (32 pts) in the H6, H5, C5 and C6 dimensions, respectively. Hence, for this data set, the apparent linewidths in the C5 and C6 dimensions of the FDM-processed spectrum are determined by the digital resolution of the matrix. The observed linewidth of the H5 dimension (28 Hz) demonstrates the increased resolution offered by FDM-processing. The digital resolution of the FDM-processed spectrum can be varied except that the maximum number of points for the current software is limited to a file size of 256 MB. This restriction limits the digital resolution achievable in the 4D spectrum; however, the FDM software can calculate all 6 high resolution 2D projections of the 4D data, which can be used to achieve higher digital resolution if needed.

Acquisition and processing of a sugar-specific 4D HCCH-COSY spectrum

A sugar-specific 4D HCCH-COSY experiment was also recorded with 8 complex points in each indirect dimension. A 2D plane (δ H1' = 6.07, δ C2' = 74.3 ppm) from the FT-processed spectrum (Supplementary Fig. S1A) contains parts of six peaks because of leakage from adjacent planes due to large linewidths in all indirect dimensions. The same 2D plane in the optimized FDM-processed spectrum shows only one peak, and the other peaks are resolved into other planes (Supplementary Fig. S1B). In total, 26 of the 29 H1', C1', C2' and H2' spin systems in the IRE have distinct chemical shifts and 13 of these are fully resolved into individual planes with FDM-processing, whereas only 4 spin systems are resolved into an individual planes in the FT-processed spectrum. Figure 4 compares the 2D C1'–H1' projections of the 4D FT-processed spectrum with the FDM-processed spectrum. As observed for pyrimidine-specific spectra, the average linewidths of the FT spectrum (137, 96 and 92 Hz for the H1', C1' and C2' resonances, respectively) are larger than those observed the FDM spectrum (43, 48 and 54 Hz for the H1', C1' and C2' resonances, respectively). Due to smaller spectral widths for this sugar-specific data set, not all the observed linewidths in the indirect dimensions are limited by the digital resolution of the reconstructed spectrum (15.6, 23 and 47 Hz/pt in the H1', C1' and C2' dimensions, respectively). Overall, the FDM-processed spectra show dramatically higher resolution and reduced linewidths in the indirect dimensions compared to the FT spectra.

resonances, respectively). Due to smaller spectral widths for this sugar-specific data set, not all the observed linewidths in the indirect dimensions are limited by the digital resolution of the reconstructed spectrum (15.6, 23 and 47 Hz/pt in the H1', C1' and C2' dimensions, respectively). Overall, the FDM-processed spectra show dramatically higher resolution and reduced linewidths in the indirect dimensions compared to the FT spectra.

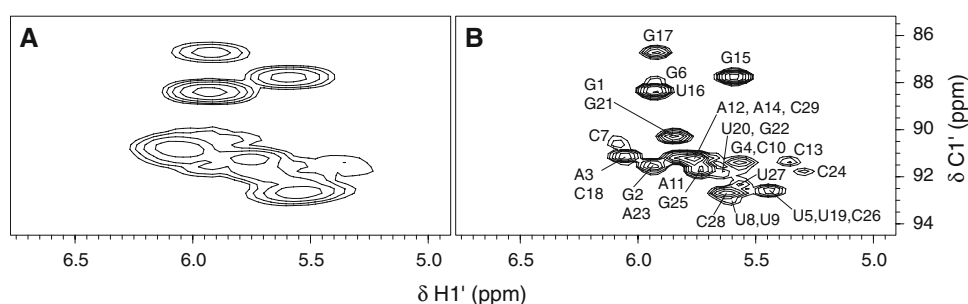
Data acquisition strategies for FDM-processing

To examine the effects of the signal-to-noise and the number of points collected in the indirect dimensions on FDM-processed spectra, pyrimidine-specific 4D HCCH-COSY experiments were performed where the number of scans per FID was reduced or where the total number of scans was not changed but fewer points were collected in the indirect dimensions. When 8 complex points were collected in each indirect dimension, the 2D C6–H6 projections were quite similar for spectra acquired in 20 and 4.5 h (Fig. 5a and b). However, acquiring fewer scans per FID reduces the signal-to-noise so two peaks were not observed or resolved in the 4.5 h spectrum (C13 and C24). In contrast, when the total number of scans in the spectra was kept constant but only 4 complex points were collected in each indirect dimension, the resolution and quality of the spectrum was dramatically reduced (see Fig. 5a and c). Thus, for the 4D HCCH-COSY spectra on this 1.0 mM 29 nucleotide IRE RNA reducing the number of scans by factor of 4 only marginally effects the quality of the spectrum; whereas, reducing the number of complex points in the indirect dimension severely broadens resonances and substantially reduces the quality of the spectrum.

Optimization of FDM-processing parameters

The FDM has several processing parameters that are empirically varied to optimize the NMR spectrum (Hu et al. 1998; Jeschke et al. 1999; Chen et al. 2000, 2003). First, the number of data points used for processing the acquisition dimension is adjusted. Since the time-domain signal in the acquisition dimension is decaying due to T₂ relaxation, the total number of points can substantially affect the overall

Fig. 4 2D H1'–C1' projections of the sugar-specific 4D HCCH-COSY spectrum of the IRE RNA generated with (a) FT processing and (b) FDM processing. The 4D experiment was acquired in 18.5 h with $8 \times 8 \times 8$ complex points in the indirect dimensions with 8 scans per FID. Assignments are from Address et al. (1997)



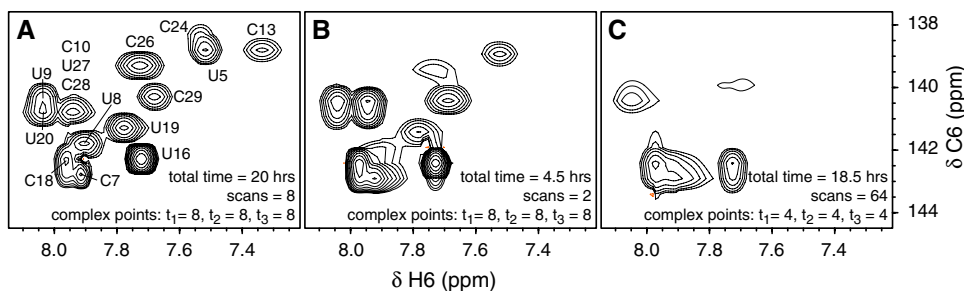


Fig. 5 2D H6–C6 projections showing the effect of acquisition parameters on the FDM processed pyrimidine-specific 4D HCCH-COSY spectra. **(a)** The data were acquired with the same parameters as in Fig. 3. **(b)** The data were acquired with $8 \times 8 \times 8$ complex points

in the indirect dimensions and 2 scans per FID for a total acquisition time of 4.5 h. **(c)** The data were acquired with $4 \times 4 \times 4$ complex points in the indirect dimensions and 64 scans per FID for a total acquisition time of 18.5 h. Assignments are from Address et al. (1997)

signal-to-noise of the FDM-processed spectrum. For example, although 1024 complex points were collected in the direct dimension of the pyrimidine-specific 4D HCCH-COSY experiment, only 256 complex points were used for the FDM-processing of the spectra shown in Figs. 3–7. Inclusion of only the initial part of the data for FDM-processing serves a similar purpose as applying a decaying exponential apodization function in FT processing.

Figure 6 illustrates the effects of two parameters that need to be optimized for FDM-processing, Γ , a smoothing parameter, and q^2 , a regularization parameter. The former represents a threshold for the linewidths calculated by the

FDM. If the magnitude of the calculated linewidth is less than Γ , then the calculated linewidth is replaced by Γ . Hence, Γ usually represents the minimum linewidth attainable in FDM processed spectra. In this way, the FDM is able to effectively deal with oscillators that are fit to increasing exponentials instead of decaying exponentials (Chen et al. 2000). Figure 6e shows that values of $\Gamma_{H6} = 4$ Hz, $\Gamma_{C6} = 15$ Hz and $q^2 = 5 \times 10^{-4}$ give the highest quality spectrum for the 4D pyrimidine-specific HCCH-COSY data. Increasing both Γ_{H6} and Γ_{C6} by a factor of two and four, respectively, leads to broader peaks and poorer resolution (Fig. 6f), whereas decreasing Γ_{H6}

Fig. 6 2D H6–C6 projections of the 4D pyrimidine-specific HCCH-COSY spectrum showing the effects of varying the FDM-processing parameters q^2 and Γ . The regularization parameter, q^2 , is set to **(a, b, c)** $q^2 = 0.05$, **(d, e, f)** $q^2 = 0.0005$ and **(g, h, i)** $q^2 = 0.000005$. Smoothing parameters in the H6, H5, C5 and C6 dimensions (Γ_{H6} , Γ_{H5} , Γ_{C5} and Γ_{C6} , respectively) are set to **(a, d, g)** $\Gamma_{H6} = 0$, $\Gamma_{H5} = 0$, $\Gamma_{C5} = 0$ and $\Gamma_{C6} = 0$; **(b, e, h)** $\Gamma_{H6} = 4$, $\Gamma_{H5} = 10$, $\Gamma_{C5} = 15$ and $\Gamma_{C6} = 15$; and **(c, f, i)** $\Gamma_{H6} = 8$, $\Gamma_{H5} = 40$, $\Gamma_{C5} = 60$ and $\Gamma_{C6} = 60$. **(e)** Optimal values of q^2 , Γ_{H6} , Γ_{H5} , Γ_{C5} and Γ_{C6} were found to be 0.0005, 4, 10, 15 and 15 Hz, respectively. The number of basis functions per window was set to 256. Red contours denote peaks with opposite sign

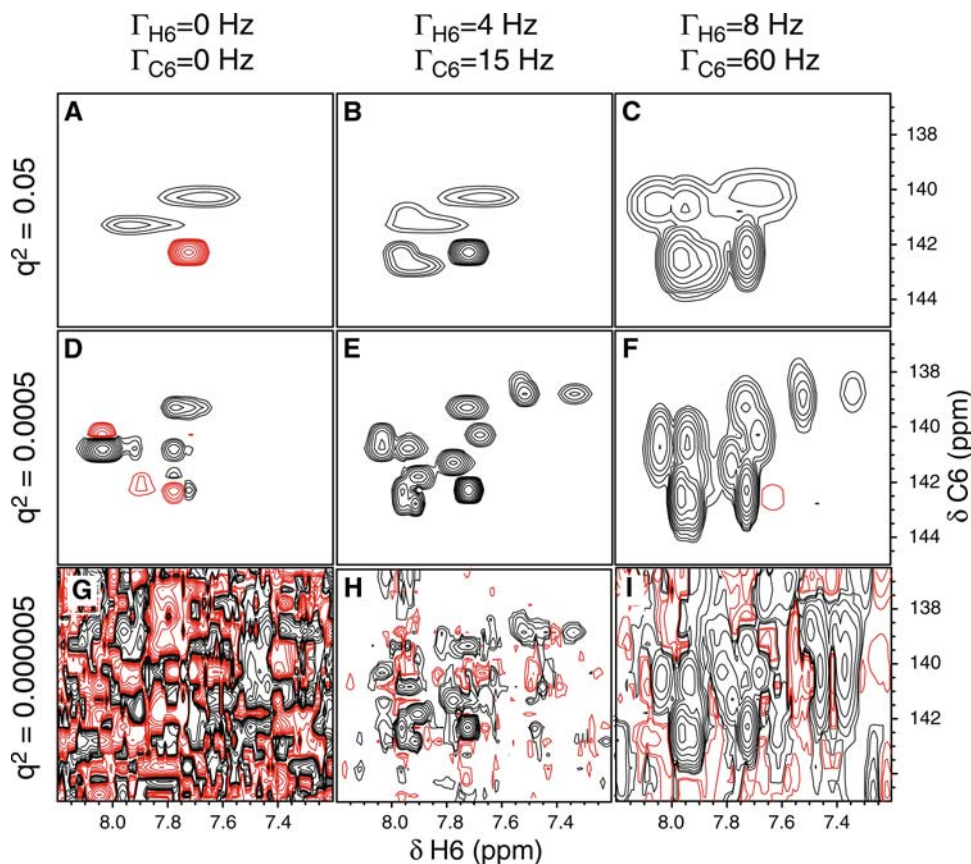
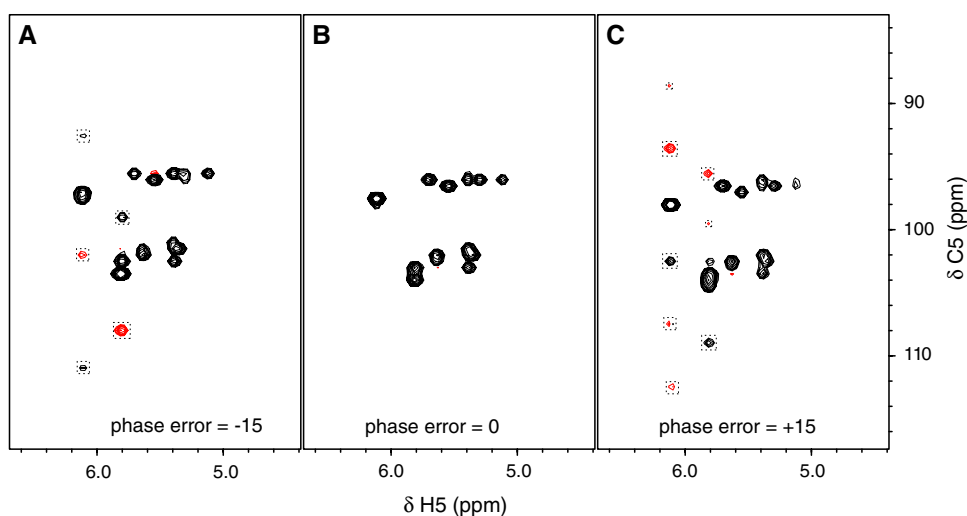


Fig. 7 2D H5–C5 projections of the 4D pyrimidine-specific HCCH-COSY spectrum showing the effects of introducing phase errors prior to FDM-processing. Zero-order phase errors of (a) -15° , (b) 0° and (c) $+15^\circ$ were applied in the time-domain for the C5 dimension prior to FDM-processing. Spurious symmetrical peaks induced by the phase errors in (a) and (c) are highlighted with dashed boxes and red contours denote peaks with opposite sign



and Γ_{C6} to 0 Hz (Fig. 6d) severely distorts the spectrum causing some peaks to change position, sign, amplitude and/or linewidth.

The FDM involves diagonalizing a large complex matrix, which may have eigenvalues and eigenvectors that are close to zero. This is an inherently ill-conditioned problem, which if not accounted for can lead to artifacts in the final spectrum. Therefore, the parameter q^2 was introduced to regularize the resulting FDM solution (Chen et al. 2000). Decreasing the regularization parameter q^2 by a factor of 100 introduces substantial artifacts in the spectrum as seen by comparing Fig. 6h and e. Increasing q^2 by a factor of 100 (Fig. 6b) distorts the spectrum by broadening peaks or altering peak positions, especially for peaks with lower signal-to-noise.

In practice, the q^2 and Γ parameters are empirically optimized. This procedure increases the time required to process data using the FDM, and as seen in Fig. 6, if incorrect values for parameters are initially used, the spectra can be severely distorted. One of the challenges when using FDM to process NMR data, even with high signal-to-noise, is that improper values for processing parameters can lead to large spectral artifacts. We have found that initially setting Γ to a value somewhat larger than the estimated natural linewidth and calculating 2D projections with a range of q^2 helps pinpoint the appropriate value. Then Γ is incrementally decreased until peaks change position, sign, amplitude or linewidth, which normally occurs when Γ is smaller than the natural linewidth (compare Fig. 6d and e).

Effects of phase distortion on FDM-processed spectra

The current implementation of the FDM-processing software can only apply zero-order phase corrections in the directly detected dimension (Mandelshtam 2001). Thus, the

pulse sequences used for FDM data collection are designed to eliminate any phase corrections in the indirect dimensions. However, phase shifts can still arise in the indirect dimensions from imperfections in pulses, gradients or off-resonance effects (Emsley and Bodenhausen 1990). Since the algorithm employed in the FDM-processing assumes perfectly in-phase signals, the effects of phase errors on spectra obtained from FDM-processing are very different from those obtained with FT processing. Figure 7 shows FDM-processed 4D pyrimidine-specific HCCH-COSY spectra where $\pm 15^\circ$ zero-order phase errors for the C5 dimension were introduced into the time-domain data. The phase errors in the time-domain signals lead to frequency shifts in the FDM-processed spectra. For example, the resonance at 6.12 and 97.5 ppm experiences apparent frequency shifts of -51 and 62 Hz for phase errors of -15° and 15° , respectively. In addition, spurious peaks are symmetrically observed around the correct signal (dashed boxes in Fig. 7a and c). These artifacts can complicate spectral analysis, so it is important that all phase errors are corrected prior to FDM-processing.

Discussion

One of the unique features of FDM compared to other high-dimensional NMR processing methods is that FDM simultaneously fits the entire time-domain NMR data set. This is in contrast to methods such as the FT and linear prediction that independently fit individual dimensions or covariance NMR, projection reconstruction and GFT that simultaneously analyze only a subset of the dimensions (Kim and Szyperski 2003; Kupce and Freeman 2003; Bruschiweiler and Zhang 2004). This leads to a key property of FDM, that additional data points in one dimension can help improve the resolution of other dimensions

(Mandelshtam et al. 1998). Thus high-resolution, high-dimensional NMR spectra can be acquired with a smaller number of time-domain data points in the indirect dimensions leading to much shorter total acquisition times than a high-dimensional conventional FT data set. The FDM methodology has been previously used with proteins (Hu et al. 2000; Chen et al. 2003, 2004) and oligosaccharides (Armstrong et al. 2004, 2005), and here we report the application of FDM-processing to studies of RNA.

The theory behind FDM has been described in detail elsewhere (Mandelshtam et al. 1998; Chen et al. 2000; Hu et al. 2000; Mandelshtam 2000); thus, only the important features for applications to the studies of RNA will be discussed here. Minor modifications were implemented in the HCCH-COSY experiment to facilitate FDM-processing. Several selective pulses were employed in the HCCH-COSY experiments, which allowed smaller sweep widths in the indirect dimensions and helped eliminate unwanted coherences. Frequency labeling in the indirect dimensions employed constant-time evolution periods, which makes it possible to double the number of indirect data points used in the FDM calculation (Zhu and Bax 1990; Chen et al. 2003). These simple modifications are readily introduced into most pulse sequences and allow for both FT- and FDM-processing.

FDM-processing of NMR data does not require apodization functions, zero filling or linear prediction techniques that are normally employed in standard FT processing. Parameters for these techniques are empirically varied to optimize the signal-to-noise, resolution and overall quality of FT-NMR spectra. Similarly, the FDM requires empirical optimization of various parameters for each data set processed. For the FDM the primary adjustable parameters are: the number of basis functions employed in a calculation window, and the regularization (q^2) and smoothing parameters (Γ). The FDM can be thought of as describing the time-domain data as a sum of exponentially damped sinusoids where the fitting is performed in the frequency domain (Mandelshtam and Taylor 1997). The total number of sinusoids that can be used in an individual dimension is determined by the basis functions defined in the frequency domain for that dimension (Hu et al. 1998). The number of basis functions is related to the number of experimental data points in that dimension; thus, different dimensions will generally have a different number of basis functions. For example, if 8 complex data points are acquired in a constant-time indirect dimension with a sweep width of 2000 Hz, constant-time doubling results in 15 total data points and 7 basis functions that can be used in the FDM calculation (Chen et al. 2003). In multi-dimensional FDM calculations, the total number of basis functions that can be used is the product of the basis functions from each dimension. For a hypothetical 2D data set consisting of two

indirect-type dimensions with 8×8 complex constant-time points, 49 total basis functions can be used in the 2D FDM calculation (Mandelshtam et al. 1998). As discussed below, the minimum number of basis functions required to fully describe the data depends upon the number of unique frequencies in a 1D trace in the nD spectrum.

Figure 5 shows the results of varying the signal-to-noise or varying the number of data points, and therefore the number of basis functions, on the FDM-processed 4D pyrimidine-specific HCCH-COSY spectra. In this case, reducing the number of scans, but keeping the number of data points in each dimension constant, had a relatively minor effect on the spectrum. The major difference is that one peak with low amplitude (C13) is missing and several other partially resolved peaks (C24 and U8) are no longer resolved in the experiment collected with 1/4 of the total number of scans. However, keeping the total scans constant, but reducing the number of data points by a factor of two in all three indirect dimensions, has a dramatic effect on the quality of the spectrum. Both the number of peaks and the resolution of the peaks in this spectrum are adversely affected by the more limited sampling. The reduction in spectral quality arises from having an insufficient number of local basis functions to define this region of the spectrum (Chen et al. 2003). This is because reducing the number of data points from 8 to 4 leads to only 3 basis functions (instead of 7) for each of the indirect dimensions (in Fig. 5c). This means that only 3 sinusoids (peaks and/or noise) can be determined along any 1D trace at a particular position in the 4D matrix. For example, if the plot in Fig. 5c were a 2D spectrum that had many peaks with degenerate chemical shifts in the H6 dimension the largest number of peaks that could be obtained by the FDM at that H6 frequency is less than or equal to the number of basis functions in the C6 dimension. Thus if there were 4 separate peaks with degenerate H6 frequencies but different C6 frequencies an FDM calculation with only 3 basis functions along C6 could give at most 3 peaks. This would lead to merging of several peaks or disappearance of one or more peaks, as has been previously illustrated for FDM-processing of NMR data (Chen et al. 2003). The disappearance or merging of peaks is seen in Fig. 5c and indicates that only 4 time-domain data points in all the indirect dimensions is below the threshold required for this experiment on the IRE and does not provide enough basis functions to describe all the frequencies in this local spectral region. A similar phenomenon is observed when using linear prediction in 4D FT spectra because there are usually a relatively small number of time-domain data points in each indirect dimension. Thus, in FT processing of 4D data, the linear prediction is often only performed in an indirect dimension after at least two other dimensions have been transformed (Kay et al. 1991). The rationale for

this is simply to minimize the number of frequencies present when performing the linear prediction calculation, because, similar to the FDM, the total number of frequencies that can be determined by linear prediction is dictated by the number of time-domain data points in that dimension. Although this principle still applies for the FDM calculations, the FDM differs because all dimensions are analyzed simultaneously, which has the distinct advantage that fewer sinusoids are generally required to fit any 1D trace in the nD data set.

Reducing the number complex data points in only one dimension from 8 to 4 had relatively little effect on the results (Supplementary Fig. S2), because there are sufficient basis functions to describe all of the frequencies. Somewhat surprisingly, the quality of the spectra was not significantly influenced whether fewer points are collected in a dimension with a high degree of spectral overlap (for example the C5 dimension which is poorly resolved as seen in Fig. 7) or in a well-resolved dimension (such as the H5). To further test the effect of number of data points on the quality of the FDM spectrum, the FDM calculation was performed on the pyrimidine-specific 4D HCCH-COSY data where all 8 complex data points were employed in the 3 indirect dimensions, but 512 complex points were used in the direct (H6) dimension instead of 256 (Supplementary Fig. S3). As expected, since there are already a large number of basis functions in the H6 dimensions, doubling the number of points (and therefore basis functions) did not significantly influence the final spectrum. This illustrates the threshold property of FDM-processing: once there are enough basis functions to accurately reproduce the spectral features, additional data points will generally not significantly improve the spectral resolution. The results here further illustrate the property that FDM-processing is progressively more powerful as the number of dimensions increases (Chen et al. 2003).

The q^2 regularization parameter in the FDM calculation was introduced by Chen et al. (2000) as a means to deemphasize small amplitude signals such as noise over larger amplitude signals. Previous studies showed that noise in the experimental time-domain data leads to spurious frequencies in the FDM calculations, which are very sensitive to the input parameters used to process the data (Chen et al. 2000). The effect of introducing this regularization procedure is that a peak will be effectively broadened away by the algorithm if the normalized amplitude of the peak is similar in magnitude to q^2 . This is observed in the bottom three panels in Fig. 6 where q^2 is set too low and thus many artifacts are observed. However, if q^2 is set too high then true peaks with lower amplitudes are not observed because they are now effectively broadened by the regularization procedure. Thus, q^2 needs to be iteratively adjusted for each data set processed to a value that does not broaden away

true signals while also eliminating artifactual, low amplitude signals. In data with high enough signal-to-noise ratio, the distinction between peaks and artifacts is clear, making optimizing q^2 straightforward.

The other critical parameter for optimization of FDM is Γ , the smoothing parameter. The FDM-spectrum is generated from the calculated frequencies and amplitudes using Gaussian line shapes which are smoothed by Γ (Chen et al. 2000; Armstrong et al. 2005). The effect of Γ is seen in Fig. 6 where a larger Γ value increases the linewidth of all the resonances, thereby reducing the effective resolution in the spectrum (compare the middle panels to the right panels in Fig. 6). In contrast to the broadening caused by q^2 , Γ has a similar effect on all the peaks, regardless of their relative amplitudes. This is because Γ is applied after the FDM calculation of the frequencies and amplitudes, whereas q^2 is a direct part of the FDM calculation (Chen et al. 2000). If the Γ is set too low, artifacts can be introduced such as incorrect signs for peaks (Fig. 6) and a reasonable value for Γ is the natural linewidth of the resonance in that dimension (Chen et al. 2000; Armstrong et al. 2005).

Figure 7 shows the effects of phase distortions on 2D C5–H5 projections of the pyrimidine-specific 4D HCCH-COSY spectrum. Since the current implementation of the algorithm expects perfectly in-phase time-domain signals, the effects of phase errors on spectra obtained from FDM-processing are very different from those obtained with FT processing. As shown in Fig. 7, phase distortions in the time-domain signal produces shifts in the frequency of the peaks and also introduces spurious peaks of opposite sign in symmetric positions around the most intense peaks. Frequency shifts were also observed in simulations where one in-phase sinusoid was used to fit an out-of-phase target sinusoid (data not shown). Time-domain simulations using the frequencies and amplitudes of the spurious peaks in Fig. 7 showed that they produce a beat pattern that also helped fit the out-of-phase character of the main peak (data not shown). However, the origin of the specific frequency for the symmetric peaks observed in Fig. 7 is not understood. Hence, 2D projections should be examined for symmetric peaks of opposite sign around a large peak, signifying a small phase error that should be corrected prior to FDM-processing.

Conclusions

The results here demonstrate the applicability of high-resolution, high-dimensional FDM-processing to 4D NMR studies on RNA. These studies represent the first application of the alternative processing techniques to RNA. The severe overlap for many of the resonances in RNAs (e.g., ribose ^1H and ^{13}C) makes them a logical target for these

methodologies. The ability to acquire high-resolution high-dimensional NMR spectra that correlate multiple nuclei should accelerate resonance assignments in RNA. The HCCH-COSY experiment utilized here represents a favorable case to test the applicability of the FDM-processing algorithm, because it has a high signal-to-noise ratio afforded by the efficient transfer of magnetization through the H5–C5–C6–H6 or H1'–C1'–C2'–H2' spin systems.

Analogous to conventional FT processing of NMR data, FDM-processing requires empirical variation of several parameters to obtain optimal spectra. Some of these parameters have non-linear effects on the spectrum and thus incorrect values can lead to huge spectral artifacts (see Fig. 6). This has become less of an issue with recent implementations of the FDM-software because the calculation is split into many windows and a 4D data set can be processed in under 40 min. The results here for the HCCH-COSY experiment on RNA illustrate that the optimal number of data points for an individual dimension will depend on the spectral overlap in that dimension in the full nD spectrum, consistent with the multi-dimensional properties of the FDM. Thus, the effect of the number of data points in a particular indirect dimension on the overall spectrum is usually very different for FDM processing compared to conventional FT processing. Once there are enough data points to provide enough local basis functions to fit the most crowded region of a spectrum, increasing the number of data points (and therefore basis functions) can have a relatively small effect on the spectrum, especially compared to what is observed for FT processing of the same data set. This point needs to be explored for individual applications to make the most efficient use of FDM processing. The results here demonstrate that FDM-processing is very valuable for obtaining high-resolution 4D HCCH-COSY spectra on a medium-sized RNA. It is expected that the methods employed here can be readily applied to other similarly sized RNAs. It will be interesting to test the ability of FDM-processing techniques to produce high-resolution high-dimensional data sets for larger RNAs, which suffer from more severe overlap and will benefit even more from the use of high-dimensional NMR techniques.

Acknowledgements This work was supported in part by NIH grants AI33098 (Arthur Pardi), GM68928 (Geoffrey S. Armstrong), NSF MCB-0236103 (Brad Bendiak) and Michael P. Latham was supported in part by an NIH Training Grant T32 GM65103.

References

- Address KJ, Basilion JP, Klausner RD, Rouault TA, Pardi A (1997) Structure and dynamics of the iron responsive element RNA: Implications for binding of the RNA by iron regulatory binding proteins. *J Mol Biol* 274:72–83
- Armstrong GS, Cano KE, Mandelshtam VA, Shaka AJ, Bendiak B (2004) Rapid 3D NMR using the filter diagonalization method: application to oligosaccharides derivatized with ¹³C-labeled acetyl groups. *J Magn Reson* 170:156–163
- Armstrong GS, Mandelshtam VA, Shaka AJ, Bendiak B (2005) Rapid high-resolution four-dimensional NMR spectroscopy using the filter diagonalization method and its advantages for detailed structural elucidation of oligosaccharides. *J Magn Reson* 173:160–168
- Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M, Kay LE (1990) Practical aspects of proton carbon carbon proton 3-dimensional correlation spectroscopy of C-13-labeled proteins. *J Magn Reson* 87:620–627
- Bruschweiler R (2004) Theory of covariance nuclear magnetic resonance spectroscopy. *J Chem Phys* 121:409–414
- Bruschweiler R, Zhang FL (2004) Covariance nuclear magnetic resonance spectroscopy. *J Chem Phys* 120:5253–5260
- Chen J, Mandelshtam VA, Shaka AJ (2000) Regularization of the two-dimensional filter diagonalization method: FDM2K. *J Magn Reson* 146:363–368
- Chen J, De Angelis AA, Mandelshtam VA, Shaka AJ (2003) Progress on the two-dimensional filter diagonalization method. An efficient doubling scheme for two-dimensional constant-time NMR. *J Magn Reson* 162:74–89
- Chen J, Nietlisbach D, Shaka AJ, Mandelshtam VA (2004) Ultra-high resolution 3D NMR spectra from limited-size data sets. *J Magn Reson* 169:215–224
- Coggins BE, Zhou P (2006) Polar Fourier transforms of radially sampled NMR data. *J Magn Reson* 182:84–95
- 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
- Emsley L, Bodenhausen G (1990) Phase-shifts induced by transient Bloch-Siegert effects in NMR. *Chem Phys Lett* 168:297–303
- Geen H, Freeman R (1991) Band-selective radiofrequency pulses. *J Magn Reson* 93:93–141
- Goddard TD, Kneller DG (1998) SPARKY 3. University of California, San Francisco
- Hu H, Van QN, Mandelshtam VA, Shaka AJ (1998) Reference deconvolution, phase correction, and line listing of NMR spectra by the 1D filter diagonalization method. *J Magn Reson* 134:76–87
- Hu H, De Angelis AA, Mandelshtam VA, Shaka AJ (2000) The multidimensional filter diagonalization method. *J Magn Reson* 144:357–366
- Hyberts SG, Heffron GJ, Tarragona NG, Solanky K, Edmonds KA, Luithardt H, Fejzo J, Chorev M, Aktas H, Colson K, Falchuk KH, Halperin JA, Wagner G (2007) Ultrahigh-resolution (1)H-(13)C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. *J Am Chem Soc* 129:5108–5116
- Jeschke G, Mandelshtam VA, Shaka AJ (1999) Pure absorption electron spin echo envelope modulation spectra by using the filter-diagonalization method for harmonic inversion. *J Magn Reson* 137:221–230
- Kay LE, Ikura M, Bax A (1990) Proton proton correlation via carbon-carbon couplings—a 3-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with C-13. *J Am Chem Soc* 112:888–889
- Kay LE, Ikura M, Zhu G, Bax A (1991) 4-Dimensional heteronuclear triple-resonance NMR of isotopically enriched proteins for sequential assignment of backbone atoms. *J Magn Reson* 91:422–428
- Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc* 125:1385–1393

- Kim S, Szyperski T (2004) GFT NMR experiments for polypeptide backbone and $^{13}\text{C}\beta$ chemical shift assignment. *J Biomol NMR* 28:117–130
- Kupce E, Freeman R (2003) Projection-reconstruction of three-dimensional NMR spectra. *J Am Chem Soc* 125:13958–13959
- Kupce E, Freeman R (2004) Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *J Am Chem Soc* 126:6429–6440
- Mandelshtam VA (2000) The multidimensional filter diagonalization method—I. Theory and numerical implementation. *J Magn Reson* 144:343–356
- Mandelshtam VA (2001) FDM: the filter diagonalization method for data processing in NMR experiments. *Prog NMR Spec* 38:159–196
- Mandelshtam VA, Taylor HS (1997) Harmonic inversion of time signals and its applications. *J Chem Phys* 107:6756–6769
- Mandelshtam VA, Taylor HS, Shaka AJ (1998) Application of the filter diagonalization method to one- and two-dimensional NMR spectra. *J Magn Reson* 133:304–312
- Marion D, Ikura M, Tschudin R, Bax A (1989) Rapid recording of 2D NMR spectra without phase cycling—application to the study of hydrogen-exchange in proteins. *J Magn Reson* 85:393–399
- McCallum SA, Pardi A (2003) Refined solution structure of the iron-responsive element RNA using residual dipolar couplings. *J Mol Biol* 326:1037–1050
- Nikonowicz EP, Pardi A (1992) Application of 4-dimensional heteronuclear NMR to the structure determination of a uniformly C-13 labeled RNA. *J Am Chem Soc* 114:1082–1083
- Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. *J Biomol NMR* 27:165–173
- Pardi A (1995) Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA. *Methods Enzymol* 261:350–380
- Schmieder P, Stern AS, Wagner G, Hoch JC (1993) Application of nonlinear sampling schemes to COSY-type spectra. *J Biomol NMR* 3:569–576
- Stephenson DS (1988) Linear prediction and maximum entropy methods in NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* 20:515–526
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696
- Wijmenga SS, van Buuren BNM (1998) The use of NMR methods for conformational studies of nucleic acids. *Prog Nucl Magn Reson Spectrosc* 32:287–387
- Zhu G, Bax A (1990) Improved linear prediction for truncated signals of known phase. *J Magn Reson* 90:405–410