ARTICLE

Analytical solution to the coupled evolution of multidimensional NMR data

Geoffrey A. Mueller

Abstract A substantial time savings in the collection of multidimensional NMR data can be achieved by coupling the evolution of nuclei in the indirect dimensions. In order to save time, the sampling of the indirect dimensions is inherently incomplete. Therefore, many algorithms and samplings schemes have been developed aimed at separating the coevolved frequencies into analyzable data with limited artifacts. This paper extends the use of circulant matrices to describe coupled evolution with convolutions. By understanding the data in terms of convolutions, there is an exact solution to the inversion problem of extracting the orthogonal vectors from the coupled dimensions. Previously, this inversion problem has been solved using peak coordinates extracted from spectra. In contrast, the method described here uses spectra directly. This solution suggests a simple sampling scheme of collecting N orthogonal spectra, and $N + 1$ projections at specific projection angles, however, the theory developed can be extended generally to arbitrary projection angles. The circulant matrix methodology is demonstrated for simulated and real data. Further, an algorithm for separating overlapped signals in the detected dimension is presented. The algorithm involves the forward calculation of the coupled spectra from the orthogonal spectra, followed by back calculation of the orthogonal spectra from the coupled spectra, thus permitting rigorous cross-validation. This algorithm is

shown to be robust in that erroneous solutions give rise to large artifacts.

Introduction

Multidimensional NMR is a powerful technique for analyzing molecules of various sizes. In order to gain more information about very complex molecules, such as proteins, spectra of increasing dimensions are utilized to ask increasingly sophisticated questions about molecular structure. However, the cost of more information is a substantial increase in sampling time for uniform sampling of the indirectly detected dimensions. As an alternative, many non-uniform sampling (NUS) techniques have emerged as time saving solutions (Kupce and Freeman 2003; Ding and Gronenborn 2002; Kim and Szyperski 2003). Some of these methods attempt to reconstruct the multidimensional spectra from NUS data (Kupce and Freeman 2004; Venters et al. 2005; Mobli et al. 2006) and many good solutions exist that approximate the uniform sampled data with few artifacts (Coggins and Zhou 2007; Marion 2006). Philosophically, these methods attempt to prevent artifacts by "filling in" the data points that were not sampled with various algorithms, or by judiciously choosing the data points sampled. Other methods do not attempt to reconstruct the N dimensional spectra, but instead use the peak coordinates to solve a series of linear equations that describe the coupled frequencies (Hiller et al. 2005; Malmodin and Billeter 2005b, Moseley et al. 2004; Eghbalnia et al. 2005; Fiorito et al. 2006). This is mathematically sound, however, it can be unsatisfying to

G. A. Mueller (✉)
National Institute of Environment Health Sciences,
111 T.W. Alexander Drive, MD-MR-01, Research Triangle
Park, NC 27709, USA
e-mail: Mueller3@niehs.nih.gov

spectroscopists accustomed to analyzing spectra and not tables of data. Additionally, accurate peak picking requires spectra with good signal to noise.

Fundamental to most of the NUS techniques is the coupling of the evolution periods for the indirectly detected dimensions. In the vocabulary of projection reconstruction techniques, the spectra *with* coupled evolution periods are the tilted projections, and the spectra *without* coupled evolution periods are the orthogonal spectra. The evolution of a peak in a tilted projection is the product of the corresponding orthogonal data, and the Fourier transform of the product of two functions is a convolution. Hence, the relation between the orthogonal spectra and tilted projection can be described with convolutions (Malmodin and Billeter 2005a).

The method PRODECOMP (Malmodin and Billeter 2005a, 2006) deviates from other NUS techniques in that PRODECOMP works with tilted projections but does not attempt to reconstruct the N dimensional spectra. Instead it optimizes a proposed model of the orthogonal spectra, which they call "shapes". These shapes form the basis for the assignments. The shape model is tested by calculating the tilted projections using the convolution relationship. Importantly, the convolution relationship is only true on a peak by peak basis. If there are overlapping peaks in the directly detected dimension, the convolution of the orthogonal spectra does not reproduce the projected spectra. Using a minimization procedure called multiway decomposition (Billeter and Orekhov 2003; Korzhnev et al. 2001; Orekhov et al. 2001, 2003), orthogonal spectra are determined that consistently match the tilted projections and consistently resolve the overlap. In essence, the orthogonal spectra are determined from the tilted projections, using a minimization procedure.

This paper demonstrates that instead of a minimization procedure, the orthogonal spectra can be used to exactly calculate the tilted projections *and* vice versa, thus permitting rigorous cross-validation. Similar to the PRODECOMP method of separating peaks, overlapping peaks can be differentiated by a failure of the algorithm to reproduce the data acquired. By testing various combinations of peaks against the data, the correct assignments can be determined. The elegance and novelty of this protocol is that it utilizes spectra at every step, and hence avoids reconstructions or minimizations. These features will be demonstrated with real data.

Methods

NMR data were collected using the protein NuiA (Kirby et al. 2002), a 138 residue protein, at a concentration of approximately 1 mM at 25°C. The data presented was acquired on a Varian INOVA spectrometer with an 11.7 T

magnet and cryogenically cooled probe. The pulse sequence used was the (4,2)D-HNCOCA sequence (Venters et al. 2005) provided with Varian's BIOPACK. All the orthogonal dimensions were acquired with a 4,000 Hz spectral window. While not a requirement, using a constant spectral window simplified the calculations because all of the data had the same Hz/point so that scaling for further manipulations was not required. The HN-CA orthogonal plane was acquired with 64 data points in the indirect dimension, the HN-CO plane with 33 points, and the HN-N plane with 94 points. Each projection was acquired with 64 data points in the indirect dimension. The effective spectral window of the projections can be calculated according to Venters et al. (Venters et al. 2005). All data were linear predicted to twice the number of acquired data points, and zero filled to 256 points using NMRPIPE (Delaglio et al. 1995). PRSP separates the hypercomplex data into complex data, resulting in four sub-spectra with opposite signs for the frequency of two of the indirect dimensions, analogous to quadrature detection in traditional sampling (Coggins and Zhou 2006). For simplicity these will be referred to as the quadrature detected spectra. Simulations and calculations with experimental data were carried out using the MATLAB platform (The Mathworks Inc.) using the equations described herein. The MATLAB code is available upon request.

In order to clarify the vocabulary, take for example the (4,2)D-HNCOCA experiment with three indirectly detected dimensions. Three 2D *orthogonal* planes can be acquired that collect (CA,H), (CO,H), and (N,H). An *orthogonal vector* refers to the strip taken at a given amide proton frequency (directly detected) of any of these planes. Note that in this paper, coupling refers exclusively to the coupling of acquisition times of the indirect dimensions, and not any other kind of coupling common to NMR. Additionally, all the manipulations between the orthogonal spectra and the projection spectra are in the frequency domain.

Where needed, spectra with multiple peaks were separated into spectra with single peaks manually using the following procedure. A 1D spectrum with two peaks was copied and then one of the peaks would be "covered up" by replacing the region of the peak with noise. The noise was a region with no peaks from the same 1D spectra, usually sampled from the far downfield region.

Results

Theory and simulations

In a coupled evolution NMR experiment, the time periods of the indirectly detected nuclei are incremented simultaneously. The relative size of the increments from one

evolution time to another defines a projection angle (Kupce and Freeman 2004), which for two indirectly detected evolution times t_1 and t_2 is defined by α:

$$\tan \alpha = \frac{\Delta t_2}{\Delta t_1} \tag{1}$$

For more dimensions it is convenient to write the projection angles in terms of direction cosines (Venters et al. 2005). A normalized vector pointed in any direction in space can be recast in terms of angles by simply taking the arccosine of the normalized size in each dimension. The frequency of a peak in a tilted projection (ω_{TILT}) can be determined by the relation:

$$\omega_{\mathrm{TILT}} = \sum_{i=1}^{n} \omega_i \cos \alpha_i \tag{2}$$

where i increments over n coevolving dimensions, α_i is the projection angle for a given dimension, and ω_i is the frequency in the orthogonal vector for that dimension. This equation works well for determining peak positions and these relations have been utilized to determine assignments with very good success (Malmodin and Billeter 2005b; Eghbalnia et al. 2005; Fiorito et al. 2006).

It is more difficult to work directly with spectra. The theory of how to do this was illustrated by Malmodin and Billeter; the relationship of coupled evolution periods can be expressed with a convolution (Malmodin and Billeter 2005a). For equal incrementing of t_1 and t_2, the projection angle from Eq. 1 would be 45°, and the projected vector can be calculated from the convolution of the two orthogonal vectors. Malmodin and Billeter (2006) further point out that these convolutions can be conveniently calculated with circulant matrices. A circulant matrix is made by taking a vector (or 1D spectra) and in each successive row of the matrix, shifting all the elements by 1 column until a square matrix is made. For example a vector a of N elements would make a circulant matrix as illustrated in Eq. 3.

$$\begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_N \\ a_N & a_1 & a_2 & \ldots & a_{N-1} \\ & & \ldots & & \\ & & & \ldots & \\ a_2 & a_3 & \ldots & a_N & a_1 \end{bmatrix} = A \tag{3}$$

In order to make a clear distinction between the vectors and the corresponding circulant matrices, the following convention will be used. Lower case letters will represent the vector or spectra, and capital letters will represent the corresponding circulant matrix. So the vector a can be formed into a circulant matrix A according to Eq. 3. Malmodin and Billeter calculate the convolution c, of a and b as

$$a * b = Ab = c \tag{4}$$

where the asterisk represents the convolution operation (Malmodin and Billeter 2006). This is convenient for 45° projections of two vectors. To extend this to more simultaneously evolved dimensions and other angles, it is helpful to use circulant matrices at all times and write:

$$AB = C \tag{5}$$

so that more matrix multiplications can be strung together. Except for a possible sign change to Eq. 4, these are mathematically equivalent, but when using Eq. 5 the vector c is extracted from the first row of the matrix C.

Circulant matrix multiplication for convolutions very nicely takes into account folding of peaks. However, in general, it impossible to determine whether a peak was folded or not. This discrepancy can be resolved by "zero padding" a and b, prior to convolution. The term zero padding will refer to adding zeros in the frequency domain, to avoid confusion with the more common "zero filling" of the time domain in NMR. Effectively, the zero padding simulates a larger spectral window for the convolution to guarantee peaks are not aliased. However, with padding, c becomes larger than it would be without padding. To get back to the units, in Hz per point, of a and b, one needs to interpolate or down-sample c to the number of points before a and b are "zero padded."

To illustrate the zero padding with an example, assume that there are two Lorentzian line shapes, a and b shown in blue in Fig. 1a, which represent the spectra from orthogonal vectors. The peak centers of a and b are at 50 and 150, on a scale with arbitrary units, 1–256. To calculate the spectra of a 45° projection of a and b, one multiplies AB. After convolution, the vector c has a peak at 200. Remember there are twofolding possibilities for c, which could be either a peak at 100 or 128 + 100, illustrated with black open circles. If we now zero pad a and b with 256 extra points out to 512, a peak appears at 200, which when downsampled to half as many points (the original units of A and B) results in a Lorentzian shaped peak at 100, illustrated with the cyan colored asterisks. To check the peak position from the convolutions with projection theory, Eq. 2 predicts for a 45° projection the peak position will be at 100, on a scale of 256 points. For details of this calculation, please see the Supplementary material. This result is illustrated in Fig. 1a with a Lorenztian lineshape drawn in red centered at 100, and it agrees with that predicted from circulant matrix multiplication with zero padding.

The zero padding of the vectors to prevent folding can be generalized. For a vector size N, and for M number of matrix multiplications, zeros should be added until the size is N times M. The vector a with zero padding will look like:
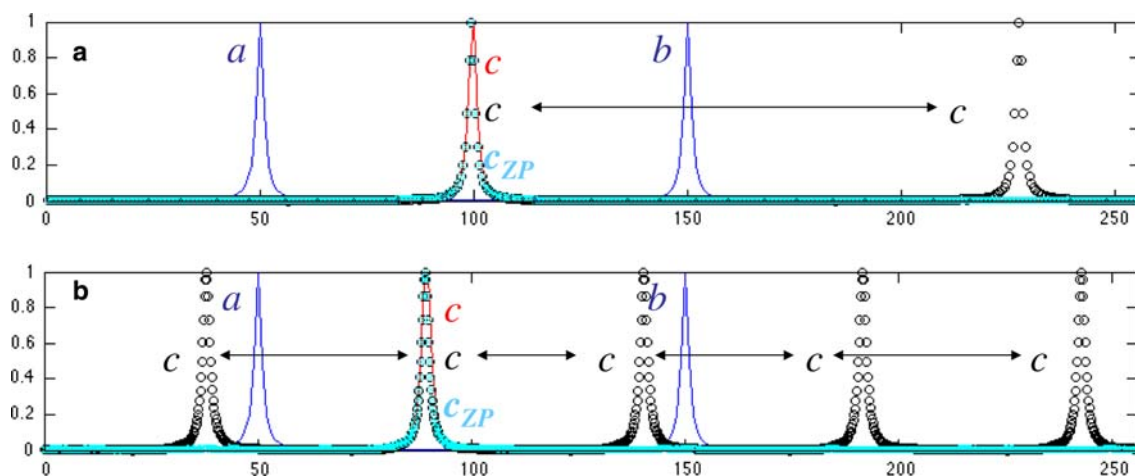
Fig. 1 Using circulant matrices to calculate convolutions and the relation to projection reconstruction. **a, b** *Blue lines* labeled *a* and *b* are simulated Lorentzian line shapes. Using circulant matrices to calculate the convolution of *a* and *b*, results in multiple possible frequency solutions shown with *open black circles* labeled with a *black c*. Zero padding *a* and *b* results in a unique solution to the convolution (*cyan asterisks labeled c_{ZP}*). This is equivalent to the projection reconstruction that would predict a Lorentzian given by the *red line*. **a, b** Simulate projection angles of (45, 45) and (33.6, 56.3)

$$a = \begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_N & 0_{N+1} & \ldots & 0_{MN} \end{bmatrix} \quad (6)$$

Additional matrix multiplications can be used to calculate the convolutions for other projection angles. The 45° projection described above can be viewed as a vector that points to the coordinates (1,1), and the convolution required to describe that projection requires the multiplication of one *A* matrix and one *B* matrix. For a projection vector that points to the coordinates (3,2), the convolution would require the multiplication of three *A* matrices and two *B* matrices or *AAABB = C* (Malmodin and Billeter 2006). This requires $M = 5$ matrix multiplications resulting in five folding possibilities, as shown with the black circles in Fig. 1b. If *a* and *b* are zero padded to 5 times their original size, and *c* is subsequently downsampled to the right number of points, the result is shown again with the cyan asterisks in Fig. 1b. To check that the convolution gives the same peak position as projection theory, one-first calculates that the (3,2) projection vector corresponds to projection angles of 33.6 and 56.3. The Supplementary material shows that on a scale of 256 points, the peak center should be at 90.0 which it is, as shown by the red line in Fig. 1b.

To generalize this to more dimensions and arbitrary angles, one simply writes the coordinates of any projection desired. For three indirect dimension this could be written with (*x*,*y*,*z*) or for *N* dimensions could be written more generally $(d_1, d_2, d_3, \ldots, d_N)$. The projection angles are the conversion of this vector into direction cosines. The corresponding matrix multiplications to get from the orthogonal vectors written for three indirect dimensions *a,*

b, c, or for *N* indirect dimensions $o_1, o_2, o_3, \ldots, o_N$ to the convolution or projection are

$$A^x B^y C^z = R$$
$$\prod_{i=1}^{N} O_i^{d_i} = R \quad (7)$$

where *R* is the resulting convolution, and the powers refer to the number of successive matrix multiplications. In order to get a unique peak position from the convolution, *M* zero pads will be required (with subsequent downsampling) where

$$M = x + y + z$$
$$M = \sum_{i=1}^{N} d_i \quad (8)$$

Equation 8 shows both the case for three indirect dimensions and more generally for *N* indirect dimensions. It is easiest to start with small integer values of d_i but note that this is not a requirement. Theoretically, any angle expressed in direction cosines can be approximated with integer coordinates up to the accuracy desired. The downside of starting with projection angles is that for large values of d_i the matrices become inconveniently large for computation if zero padding the vectors. In summary, projections at any angle, in any number of dimensions can be described by the multiplication of circulant matrices using Eqs. 7 and 8.

So far, the equations describe the calculation of the convolution from the orthogonal vectors. Is it possible to work backwards from the convolutions to the orthogonal

vectors? For Eq. 5 there exists an exact solution to the inversion of B, so that given B and C, one can solve for A (Davis 1994). For circulant matrices A, B, and C derived from the vectors a, b, and c, the vector a can be back calculated from the vectors b and c using the following relation:

$$a = F^{-1}\left(\frac{F(c)}{F(b)}\right) \tag{9}$$

where F represents the Fourier transform and F^{-1} represents the inverse Fourier transform. Equation 9 is a rearrangement of the familiar convolution theorem, which is well known in NMR, that states the Fourier transform of the convolution of a and b is equal to the product of the Fourier transform of a and the Fourier transform of b:

$$F(a*b) = F(a)F(b) = F(c) \tag{10}$$

Note that the variables a, b, and c are arrays so that the multiplication and division in Eqs. 9 and 10 refer to element by element operations. To return to the discussion of projections, this provides the needed theory to back calculate orthogonal vectors from projections.

To see how this might be helpful for multidimensional NMR experiments, we can write out the following series of projections for a hypothetical experiment with three indirectly detected dimensions. The corresponding projection coordinates and angles for the experiments would be:

$$(1,1,1) = (54.7, 54.7, 54.7) = R_1 \tag{11a}$$
$$(2,1,1) = (32.3, 65.9, 65.9) = R_2 \tag{11b}$$
$$(1,2,1) = (65.9, 32.3, 65.9) = R_3 \tag{11c}$$
$$(1,1,2) = (65.9, 65.9, 32.3) = R_4 \tag{11d}$$

These would correspond to the following in our matrix nomenclature using Eq. 10:

$$ABC = R_1 \tag{12a}$$
$$AABC = R_2 \tag{12b}$$
$$ABBC = R_3 \tag{12c}$$
$$ABCC = R_4 \tag{12d}$$

At this point it is important to state a few theorems about circulant matrices. The multiplication of circulant matrices always results in a circulant matrix, and the multiplication of circulant matrices commutes (Davis 1994). This allows a substitution of Eq. 12a into Eqs. 12b–12d, with the result:

$$AR_1 = R_2$$
$$BR_1 = R_3 \tag{13}$$
$$CR_1 = R_4$$

And from Eq. 9 we can rearrange Eq. 13 to see:

$$a = F^{-1}\left(\frac{F(r_2)}{F(r_1)}\right)$$
$$b = F^{-1}\left(\frac{F(r_3)}{F(r_1)}\right) \tag{14}$$
$$c = F^{-1}\left(\frac{F(r_4)}{F(r_1)}\right)$$

This suggests that by acquiring the four projections proposed in Eq. 11a, one could back calculate the orthogonal projections, exactly.

As a technical note, Eqs. 9–14, move freely back and forth between the circulant matrix and vector representations of the spectra. This is just a matter of preference, based on considerations of the vector size. Equation 13 could just as well be written, $a*r_1 = r_2$. However, an element by element convolution results in $2N-1$ points, whereas circulant matrix multiplication always results in another square matrix of the same size. Both are valid ways to calculate convolutions, but require different interpolations or zero filling to correctly match the units of the resulting spectra. Using circulant matrices for the forward calculation and vectors for the backward calculation was simply convenient in terms of programming.

To continue the discussion of vector size, the relative size of the vectors when performing the back calculation requires comment. Figure 2 illustrates what parts of the vectors are data versus zeros or noise. First recall that in order to forward calculate r_1 from a, b, and c, each vector would be zero padded to 3 times the original size, $M = 3$, see Fig. 2a. To calculate r_2 ($M = 4$) from a, b, and c each vector would be zero padded to 4 times the original size, see Fig. 2b. Therefore to multiply $AR_1 = R_2$, we should first zero pad both a and r_1 to the size of r_2, $M = 4$. The r_1 data is the convolution of three orthogonal vectors, so it should be interpolated to $M = 3$ size, to be appropriately scaled relative to r_2. Visually this is shown in Fig. 2c. Therefore for the backwards calculation, the opposite needs to be done. The r_2 vector will be interpolated to $M = 4$ size, and the r_1 vector interpolated to $M = 3$, and then padded to $M = 4$. This time the padding needs to be noise to prevent dividing by zero in Eq. 14. In general, the noise vector was extracted from a region of the 2D projection with no peaks, and no values near zero. Then, the result a is extracted from the first $M = 1$ data points after the inverse Fourier transform, as illustrated in Fig. 2d. All of the zero padding and scaling required for the calculations in this paper is tabulated in the Supplementary material.

Results from experimental data

To demonstrate this experimentally, the projections R_{1-4} from Eq. 11, and the three orthogonal planes of the (4,2)D-

Fig. 2 Zero padding and scaling to get unique solutions. *Solid lines* represent data, and *thin lines* represent zero padding. *Thin dashed lines* represent padding with noise instead of zeros to avoid dividing by zero. *Colors* have been added to help differentiate *a*, *b*, *c*, r_1 and r_2. **a**, **b** Depict the data and zero padding needed to execute Eqs. 12a and 12b respectively, **c** the needed zero padding for Eq. 13, where *a* could be *a*, *b*, or *c*. Similarly, **d** shows the needed noise padding for Eq. 14. The number of matrix multiplications, or *M*, as defined by Eq. 8 is displayed at the bottom as a "scale"
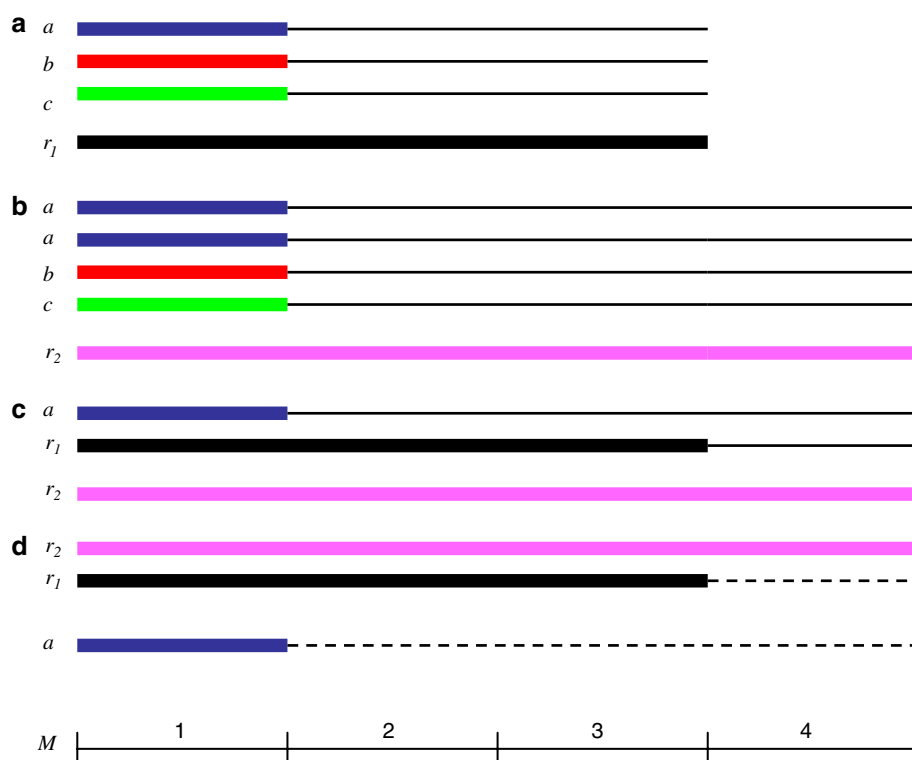
HNCACO experiment were acquired using the protein NuiA (Kirby et al. 2002; Venters et al. 2005). Figure 3a shows the HN strip of S137 from the r_2 projection in black. The line in red shows the r_2 vector extracted from the R_2 matrix calculated according to the formula above for AR_1. There is excellent agreement between the experimental data and the calculated spectra. Figure 3d demonstrates the backwards calculation of vector *a* using Eq. 14. Shown in black in Fig. 3d, is the HN strip of S137 from the CA plane (1,0,0). Calculating *a* from r_2 and r_1 according to Eq. 14 results in the black line in Fig. 3d. Again the results are in excellent agreement for the frequency, but this time the
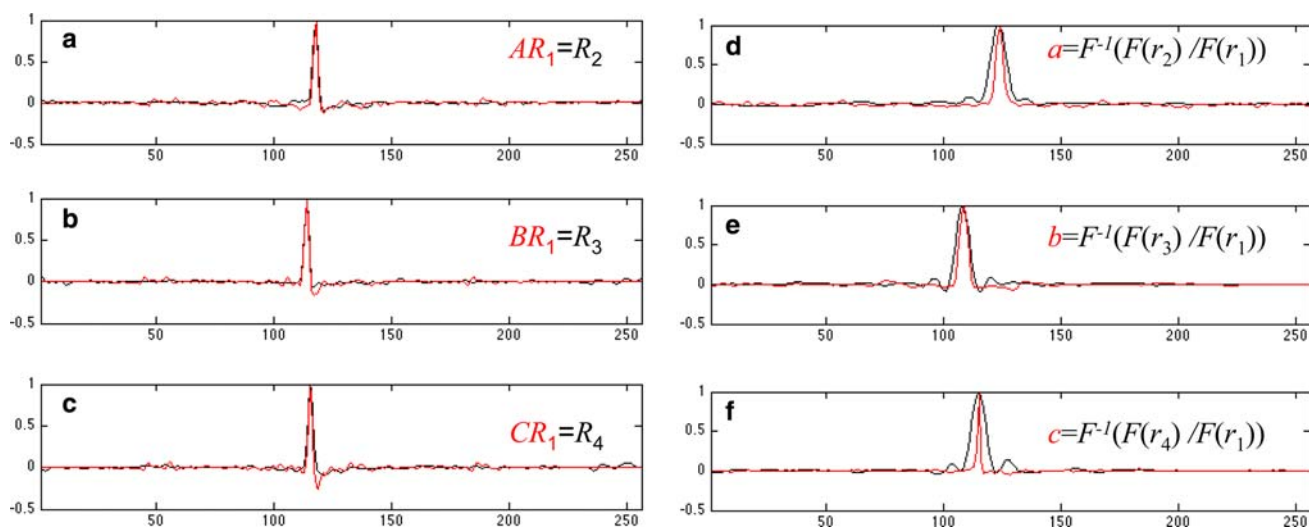
Fig. 3 Forward and backward calculations. **a–c** Experimental data from the amide of S137 in *black* from the projections $R_{2–4}$. The *red lines* are the calculated spectra according to Eq. 13. **d–f** the orthogonal vectors (CA, CO, and N respectively) of S137 are shown in *red*. The calculated spectra according to Eq. 14 is shown in *black*.

The *inset equations* are *color coded with the lines*. The *X*-axis scale is in points, which is convenient for matrix multiplication. The data in **a**, **b** and **c** were collected with a spectral window of 10.68 kHz, and the data in **d**, **e**, and **f** were collected with a spectral window of 4 kHz

linewidth is noticeably different due to the difference in resolution of the projections and the orthogonal vectors. For completeness, Fig. 3b and c show the forward calculations of $BR_1 = R_3$ and $CR_1 = R_4$ respectively, and Fig. 3e and f show the backwards calculation of the CO (b vector) and the N (c vector) according to Eq. 14 respectively. It should be noted that in these figures and calculations all the quadrature detected data of each projection was utilized, which will be discussed in greater detail later.

The mismatch of the linewidth in the backwards calculation is due to the difference in resolution of the projections and orthogonal spectra. The projection R_1 was acquired a 12.02 kHz effective spectral window, and R_{2-4} were acquired with a 10.68 kHz effective spectral window. Each projection was acquired with 64 data points in the indirect dimension, for a resolution of about 156–188 Hz/point. In contrast, the red lines in 3d–f were acquired at 62, 125, and 42 Hz/point, respectively, so the results are consistent with the calculated peaks retaining the resolution of the data from which they were derived. Potentially confusing, the linewidths in Fig. 3a–c appear narrower, but in fact are not. The scale in Hz instead of points for Fig. 3a–c is 10.68 kHz vs. 4,000 Hz in Fig. 3d–f, so the lines appear narrower in Fig. 3a–c but are in fact similar to Fig. 3d–f.

There is a relation between the coordinate representation of the projection and the forward and back calculations.

This becomes clear if we treat the coordinates like a vector. The coordinates (1,0,0) become [1 0 0]. If we add [1 0 0] + [0 1 0] + [0 0 1] = [1 1 1] this is equivalent to the forward calculation, $ABC = R_1$. In other words, multiplying the circulant matrices is the same as adding the vectors describing the projections. In contrast, the backward calculation is the same as subtracting the vectors. In other words [2 1 1] – [1 1 1] = [1 0 0]; that is a can be extracted from r_1 and r_2. This formula shows how more projections could be acquired to sort out degeneracies, and how the various projections are related. This formula is implicit in previous work using the forward calculation (Malmodin and Billeter 2006), but is explicitly stated for the forward and backward calculation here.

To test the predictions of this model of adding and subtracting the projection coordinates, we can utilize the data already collected in order to obtain quadrature detection that has not been dealt with up to this point. In order to collect a (4,2)D projection, four data sets are actually collected and in post processing added and subtracted such that the sign of the first two indirectly detected dimensions oscillates. The R_2 vectors after post processing can be represented as [2 1 1], [2 −1 1], [−2 1 1], and [−2 −1 1]. If we subtract, or backwards calculate, a from the respective vectors of R_1([±1 ±1 1]), we should get [1 0 0], [1 0 0], [−1 0 0], and [−1 0 0]. Figure 4 shows such calculations and how well each matches up with a, shown

Fig. 4 Examples of vectors with positive and negative frequencies. *Red lines* represent data, and *black lines* represent calculated spectra. **a** The backward calculation using the [2 1 1] and [1 1 1] vectors. Similarly, **b** uses the [2 −1 1] and [1 −1 1] vectors, **c** uses the [−2 1 1] and [−1 1 1] vectors, and **d** uses [−2 −1 1] and [−1 −1 1] vectors
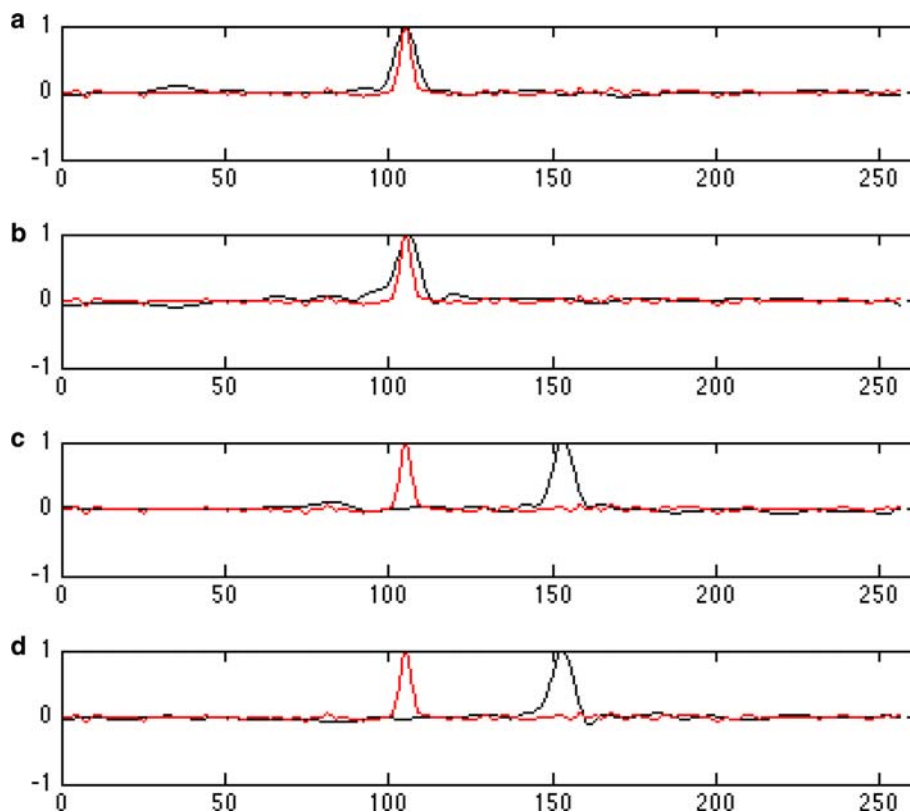
in red. Figure 4a, b highlight a good match, while Fig. 4c and d show that we indeed extract the negative frequency, shown in black. If we reverse the calculated vectors that have the opposite sign, and then sum all four calculated vectors in Fig. 4, we obtain the results shown in Fig. 3a. This demonstrates the utility of understanding the addition and subtraction of the vector model of the projections, and that all of the data in an experiment can be utilized for signal averaging.
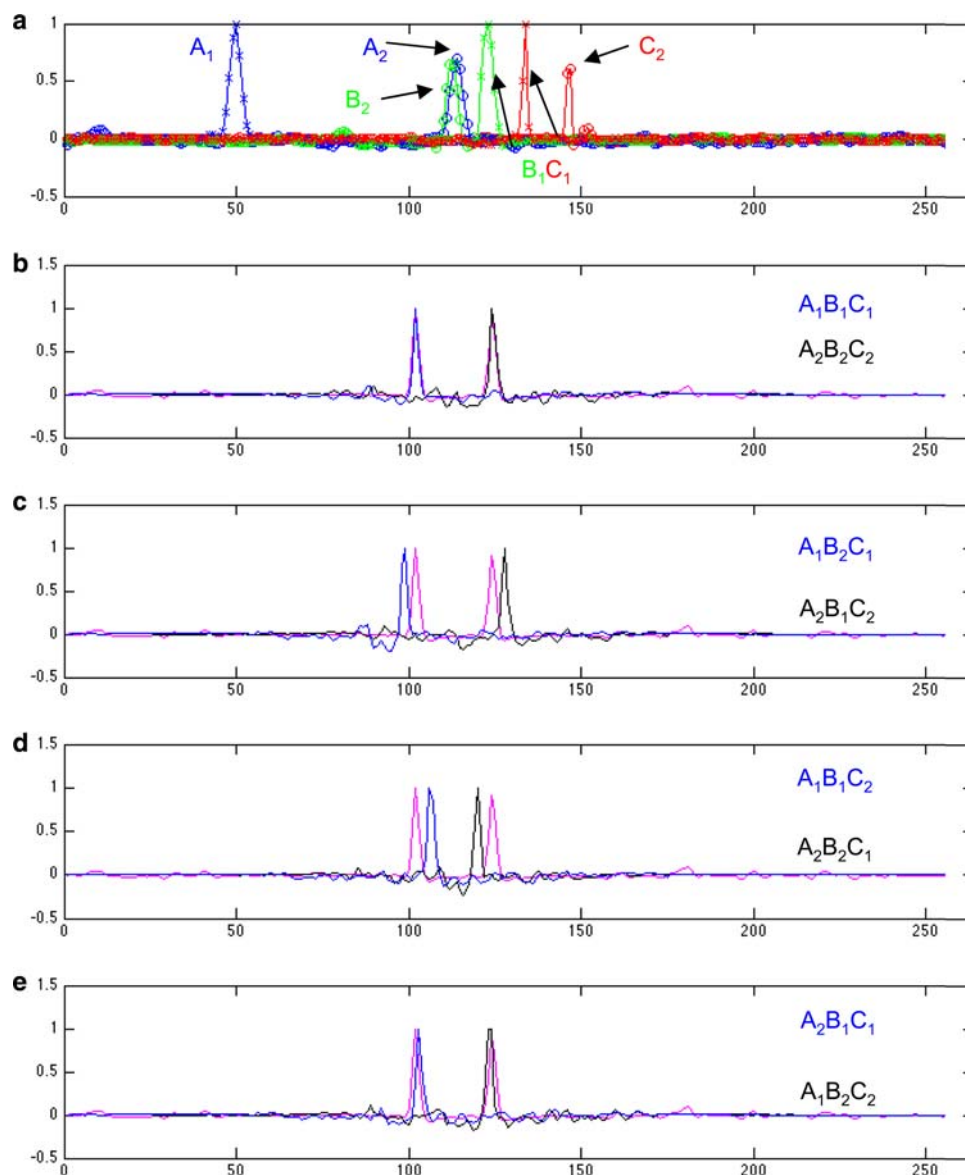
Separating overlap

The problem with describing coevolution with convolutions is that when there is overlap in the directly detected dimension, the convolution does not reproduce the data. This is because the relations are true only on a peak by peak basis. Therefore, this section will explore separating the spectra into data with only one peak and testing various combinations of peaks against the data using the relations above to generate assignments. The separation procedure relies on the fact that erroneous combinations of peaks will lead to large errors and not match the data. The following section demonstrates the feasibility of this separation procedure.

Figure 5a shows the spectra of two residues with overlapping amide proton frequencies, namely T50 and W84, in the (4,2)D-HNCACO data collected on the protein NuiA. There are two peaks in the CA strip shown in blue, two peaks in the CO strip shown in green, and two peaks in the N strip shown in red. Only one combination of the CA, CO, and N vectors should correctly reproduce the [1 1 1] projection. The two peaks in the CA strip were manually

Fig. 5 Separation of peaks using Eq. 12. **a** The amide vectors from the orthogonal planes of the CA (*labeled A, blue*), CO (*labeled B, green*), and N (*labeled C, red*) at 6.7 ppm using *solid lines*. There are peaks from two overlapping residues T50 and W84. To differentiate the peaks the vectors were split in two and labeled 1 (*asterisks*) or 2 (*circles*). Using Eq. 12a, four combinations were calculated according to the inset and shown in **b–e**. The results are colored coded with the *line color of the graph* and the inset equation text coordinated either *blue* or *black*. The *magenta line* in each is the experimental data for R_1
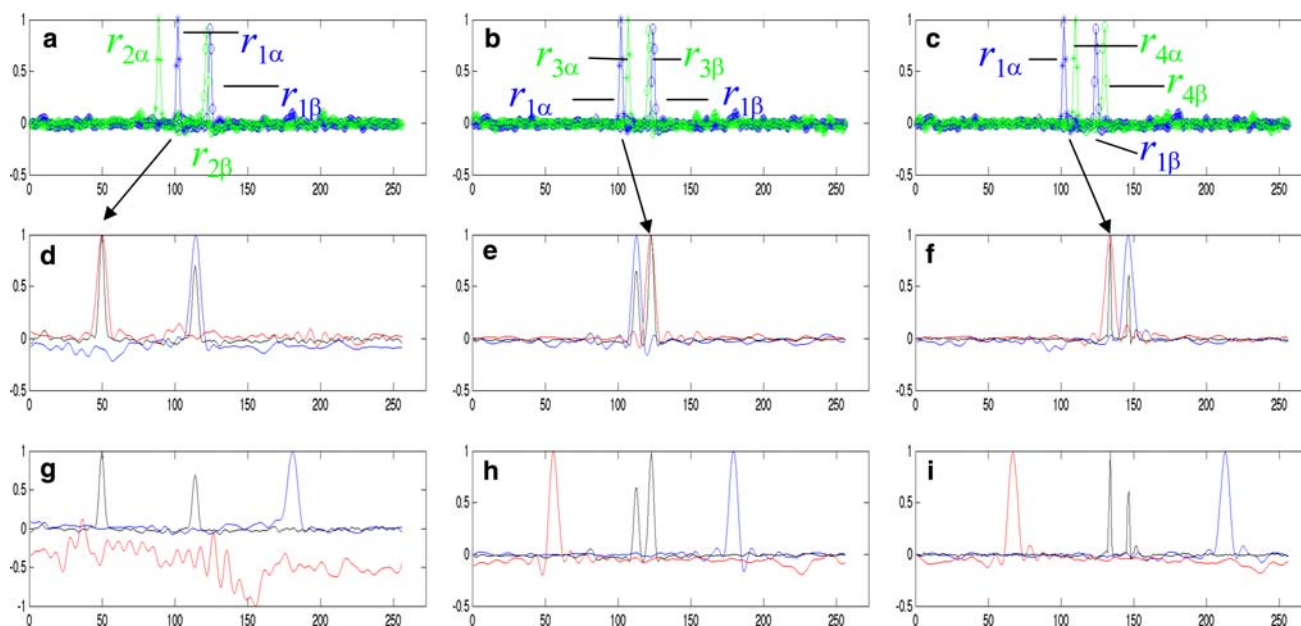
Fig. 6 Separation of peaks using Eq. 13. **a–c** The projections r_1 (*blue solid line each panel*) and r_{2-4} (*green solid line*, respectively) at 6.7 ppm. There are peaks from two overlapping residues T50 and W84. To differentiate the peaks the vectors were split in two vectors, shown with *asterisks* (*labeled* α) or *open circles* (*labeled* β) in panels **a**, **b**, and **c**. The r_2 r_3, r_4, and r_1 spectra were used to back calculate the orthogonal vectors of CA (**d**, **g**), CO, (**e**, **h**), and N (**f**, **i**) shown with *black lines*. The calculated spectra with alternate pairing of the manually separated data are shown with *red* and *blue lines* in **d–i**. *Red lines* in **d–f** indicate pairing of $r_{1\alpha}$ with $r_{2\alpha}$, $r_{3\alpha}$, and $r_{4\alpha}$, respectively, while *blue lines* indicate pairing of $r_{1\beta}$ with $r_{2\beta}$, $r_{3\beta}$, and $r_{4\beta}$, respectively. *Red lines* in **g–i** indicate pairing of $r_{1\alpha}$ with $r_{2\beta}$, $r_{3\beta}$, and $r_{4\beta}$, respectively, while *blue lines* indicate pairing of $r_{1\beta}$ with $r_{2\alpha}$, $r_{3\alpha}$, and $r_{4\alpha}$, respectively

separated into 2 vectors, a_1 and a_2, and similarly for CO, and N. Then the four exclusive convolutions of these six vectors ($A_1B_1C_1$, $A_2B_2C_2$, or $A_1B_2C_1$, $A_2B_1C_2$, or $A_1B_1C_2$, $A_2B_2C_1$, or $A_2B_1C_1$, $A_1B_2C_2$) were compared for their similarity to the r_1 projection. Figure 5b–e show the calculated vectors in blue and black versus the experimental data in magenta. Figure 5b and 5e are very good matches with 5b visually appearing slightly better. A simple way to mathematically compare the match of 1D vectors is to take the dot product; for parallel vectors (a good match) the dot product will be 1.0, and for antiparallel vectors the dot product will be −1.0. The dot product of the calculated versus experimental vectors in Fig. 5b is 0.89 and in Fig. 5e it is 0.68. Indeed, examining the data from the previous assignments of NuiA confirms that the correct assignment is T50 peaks a_1,b_1,c_1 and W84 peaks a_2,b_2,c_2.

To gain more confidence that this is the correct separation of these peaks, we can also back calculate the orthogonal vectors from the projections. Figure 6a–c shows again the same two residues, with the [1 1 1] projection in blue and the [2 1 1], [1 2 1], and [1 1 2] projections in green, respectively. With only two peaks in each spectra there are now only two exclusive combinations of these peaks that using Eq. 14 will either reproduce the data acquired for the orthogonal vector or not. The peaks were manually separated, for example into $r_{1\alpha}$ and $r_{1\beta}$, and $r_{2\alpha}$ and $r_{2\beta}$ for Fig. 6a and similarly for Fig. 6b–c. The

orthogonal vectors were calculated according to Eq. 14 for the separated vectors in the two exclusive combinations. Figures 6d–f vs. Fig. 6g–i show these comparisons as indicated in the caption. In each case, only one calculation makes a good match with the experimental data. One can assign the spectra by keeping track of which match is correct, and from which peak in the [1 1 1] projection the orthogonal peak came from. For example, Fig. 6 was rigged to consistently show that the peak drawn with the blue asterisks ($r_{1\alpha}$) in Fig. 6a–c, always combines with the other peak drawn with the green asterisks ($r_{2\alpha}$, $r_{3\alpha}$, and $r_{4\alpha}$) to give the red line that matches in the orthogonal spectra shown in Fig. 6d–f. By following the arrow in Fig. 6 from the same peak in the [1 1 1] projection to the calculated orthogonal spectra, the correct assignment of peaks to T50 and W84 can be determined. This confirms the previous assignments obtained with the forward calculations and again shows that only the correct pairing of peaks gives rise to the projected spectra.

Discussion

The basic manipulations of circulant matrices to perform convolutions and the relations to co-elvoved spectroscopy were described by Malmodin and Billeter (Malmodin and Billeter 2005a, 2006). The implementation of their theory,

PRODECOMP, looked for self-consistent solutions to Eq. 4 over many projections using a minimization technique called multiway decomposition (Malmodin and Billeter 2006; Staykova et al. 2008a, b). The two factors that differentiate the work presented here from the previous work are, first, the manipulation of the projections to back calculate the orthogonal spectra, and second, no multidimensional minimizations. Equation 9 shows that Eq. 5 can be inverted and orthogonal vectors calculated exactly from projections with many coupled evolution periods. Additionally, this suggests a new strategy for what projection angles to acquire that would "deconvolute" the convoluted frequencies. For example, this paper demonstrated how the [2 1 1] projection can be used in concert with the [1 1 1] project to extract the [1 0 0] orthogonal vector. Using the theory described here one can image how many other projections and integers will work as well. To pick a random example, the [2 3 1] projection could be used with the [2 2 1] projection to extract the [0 1 0] orthogonal vector. For any given angle, the vector model of the projections can be used to suggest a complimentary angle to extract an orthogonal vector.

While there are differences, many of the benefits of the PRODECOMP method apply here as well: There are no reconstructions from incomplete data, no sparse sampling artifacts, and no cleaning or correcting procedure. There are no angle restrictions. And, there are no requirements that angles be restricted to being 'on a grid', where angles are limited by the resolution of the number of data points.

The appeal of using this procedure to extract the frequencies that coevolve is that one works with spectra at every step thus permitting rigorous cross validation to insure that frequencies are matched correctly. The theory shows how to manipulate the orthogonal spectra to match the projected spectra and vice versa. Figures 5 and 6 demonstrated how the spectra can be "decomposed" manually to generate the assignments for overlapped frequencies in the directly detected dimension. The main assumption is that incorrectly separated peaks will generate results that do not match real spectra. In the future, hopefully this can be implemented in a more automated way.

Again, convolutions correctly describe coevolution only for single peaks in the 1D spectrum. This will therefore define the signal to noise limit for this protocol. As noise peaks get larger they begin to be convoluted with the "real peaks" and the assumptions break down. In data not shown random vectors with spectral noise were added to the calculations shown here in Figs. 1, 3 and 4. Up to ten times the noise could be added before the results became uninterpretable showing that this procedure should work robustly for data significantly worse than that presented here.

One current detraction of what is presented here is that this proves to be difficult, but not theoretically impossible, to work with varying spectral windows for the indirect dimensions. By setting all the spectral windows to 4,000 Hz, and zero filling each to 256 points, there was no need to scale the spectra beyond that already mentioned. This presents a disadvantage for any dimension, like ^{15}N, which could benefit from a much narrower spectral window. Another difficulty with varying spectral windows is how to fill with noise when a vector needs to be scaled to a smaller Hz/point, so as to avoid dividing by zero when using Eq. 12. In data not shown, it proved easier to work with constant spectral windows.

The zero padding and "noise padding" introduced here to prevent folding of peaks may appear to unnecessarily complicate the units. In addition, it may appear to eliminate the usefulness of circulant matrices, in that these matrices were introduced to correctly calculate the folding of peaks, but now peaks are prevented from being folded. The advantage, discussed above, is that there is no ambiguity in the coordinates of a peak. Another useful feature is that the zero padding keeps the linewidths narrow. In a traditional point by point discrete convolution of two vectors of the same size N, the size of the resulting vector is $2N-1$. The linewidth of the new peak is roughly the sum of the linewidths of the original peaks. When the resulting vector is down-sampled to N points, it appears that the linewidth does not change. In contrast, when using circulant matrices for convolutions, the size of the resulting vector is only N, but the linewidth is the sum of the linewidths of the original two peaks, making it appear as if the linewidth increased. For only two or three multiplications, like that used in PRODECOMP, this is generally not a problem. However, to reach more angles that require M matrix multiplications the line broadening can become excessive (data not shown). Hence, the zero and noise padding helps to keep the line widths narrow, padding the vectors to a size similar to that done in point by point convolutions.

A final fascinating point, is that the projection vectors [2 1 1] and [1 1 1] are only separated by 19°, in Euclidean space. This paper demonstrated that this separation is enough to extract the frequency of one of the orthogonal components, using convolutions. In contrast, a projection reconstruction utilizing only the same two vectors would not be expected to give very good frequency resolution. This comment is not intended to disparage projection reconstruction. Instead it emphasizes that the method proposed in this paper for extracting assignments is significantly different from existing reconstruction protocols.

In summary, this paper shows an exact method to back calculate the orthogonal vectors from projections by connecting the vector model to describe the projections, and the circulant matrix theory. The method is similar, but

deviates from previous methodologies (Malmodin and Billeter 2005b, 2006; Staykova et al. 2008a). Hopefully the generalizations presented will stimulate further development of algorithms for analyzing data from non-uniform sampling techniques.

References

Billeter M, Orekhov V (2003) Three-way decomposition and nuclear magnetic resonance. In: Computational Science—Iccs 2003, Pt I, Proceedings, vol 2657 pp 15–24

Coggins BE, Zhou P (2006) PR-CALC: a program for the reconstruction of NMR spectra from projections. J Biomol NMR 34:179–195

Coggins BE, Zhou P (2007) Sampling of the NMR time domain along concentric rings. J Magn Reson 184:207–221

Davis PJ (1994) Circulant matrices. Chelsea Publishing, New York, NY

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

Ding K, Gronenborn AM (2002) Novel 2D triple-resonance NMR experiments for sequential resonance assignments of proteins. J Magn Reson 156:262–268

Eghbalnia HR, Bahrami A, Tonelli M, Hallenga K, Markley JL (2005) High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. J Am Chem Soc 127:12528–12536

Fiorito F, Hiller S, Wider G, Wuthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. J Biomol NMR 35:27–37

Hiller S, Fiorito F, Wuthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102:10876–10881

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

Kirby TW, Mueller GA, Derose EF, Lebetkin MS, Meiss G, Pingoud A, London RE (2002) The nuclease A inhibitor represents a new variation of the rare PR-1 fold. J Mol Biol 320:771–782

Korzhnev DM, Ibraghimov IV, Billeter M, Orekhov VY (2001) MUNIN: application of three-way decomposition to the analysis of heteronuclear NMR relaxation data. J Biomol NMR 21:263–268

Kupce E, Freeman R (2003) Projection-reconstruction of three-dimensional NMR spectra. J Am Chem Soc 125:13958–13959

Kupce E, Freeman R (2004) The Radon transform: a new scheme for fast multidimensional NMR. Concepts Magn Reson Part A 22A:4–11

Malmodin D, Billeter M (2005a) Multiway decomposition of NMR spectra with coupled evolution periods. J Am Chem Soc 127:13486–13487

Malmodin D, Billeter M (2005b) Signal identification in NMR spectra with coupled evolution periods. J Magn Reson 176:47–53

Malmodin D, Billeter M (2006) Robust and versatile interpretation of spectra with coupled evolution periods using multi-way decomposition. Mag Res Chem 44:S185–S195

Marion D (2006) Processing of ND NMR spectra sampled in polar coordinates: a simple Fourier transform instead of a reconstruction. J Biomol NMR 36:45–54

Mobli M, Stern AS, Hoch JC (2006) Spectral reconstruction methods in fast NMR: reduced dimensionality, random sampling and maximum entropy. J Magn Reson 182:96–105

Moseley HNB, Riaz N, Aramini JM, Szyperski T, Montelione GT (2004) A generalized approach to automated NMR peak list editing: application to reduced dimensionality triple resonance spectra. J Magn Reson 170:263–277

Orekhov VY, Ibraghimov IV, Billeter M (2001) MUNIN: a new approach to multi-dimensional NMR spectra interpretation. J Biomol NMR 20:49–60

Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. J Biomol NMR 27:165–173

Staykova DK, Fredriksson J, Bermel W, Billeter M (2008a) Assignment of protein NMR spectra based on projections, multi-way decomposition and a fast correlation approach. J Biomol NMR 42:87–97

Staykova DK, Fredriksson J, Billeter M (2008b) PRODECOMPv3: decompositions of NMR projections for protein backbone and side-chain assignments and structural studies. Bioinformatics 24:2258–2259

Venters RA, Coggins BE, Kojetin D, Cavanagh J, Zhou P (2005) (4, 2)D Projection–reconstruction experiments for protein backbone assignment: application to human carbonic anhydrase II and calbindin D(28 K). J Am Chem Soc 127:8785–8795