



## Optimized angle selection for radial sampled NMR experiments

John M. Gledhill Jr., A. Joshua Wand\*

Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania, 905 Stellar-Chance Laboratories, 422 Curie Boulevard, Philadelphia, PA 19104-6059, USA

### ARTICLE INFO

#### Article history:

Received 21 May 2008

Revised 3 September 2008

Available online 11 September 2008

#### Keywords:

NMR spectroscopy

Multidimensional NMR

Multidimensional Fourier transform

Radial sampling

Sampling angle selection

### ABSTRACT

Sparse sampling offers tremendous potential for overcoming the time limitations imposed by traditional Cartesian sampling of indirectly detected dimensions of multidimensional NMR data. Unfortunately, several otherwise appealing implementations are accompanied by spectral artifacts that have the potential to contaminate the spectrum with false peak intensity. In radial sampling of linked time evolution periods, the artifacts are easily identified and removed from the spectrum if a sufficient set of radial sampling angles is employed. Robust implementation of the radial sampling approach therefore requires optimization of the set of radial sampling angles collected. Here we describe several methods for such optimization. The approaches described take advantage of various aspects of the general simultaneous multidimensional Fourier transform in the analysis of multidimensional NMR data. Radially sampled data are primarily contaminated by ridges extending from authentic peaks. Numerical methods are described that definitively identify artifactual intensity and the optimal set of sampling angles necessary to eliminate it under a variety of scenarios. The algorithms are tested with both simulated and experimentally obtained triple resonance data.

© 2008 Elsevier Inc. All rights reserved.

### 1. Introduction

With the advent of cold probe technology, time rather than sensitivity is often a limiting factor in multidimensional NMR experiments of macromolecules such as proteins. This is particularly true in the case of experiments of high dimensionality. For example, the collection of a traditional four-dimensional experiment with high resolution, because of the strict requirement of Cartesian sampling, would generally require weeks if not months of measurement time. Multiple approaches have recently been introduced in an effort to overcome the measurement time requirements presented by sequential and equi-spaced sampling of time domain data. These include analysis of non-linearly sampled time domain data [1,2], filter diagonalization [3], the GFT-based approach [4], projection reconstruction [5] and the direct multidimensional Fourier transform [6–8]. Although all of the methods listed have proven utility, the basis for selecting one method over another has yet to be established. This uncertainty arises from the various sparse sampling schemes employed by each of the methods.

Of the sparse sampling methods radial sampling of the indirect evolution domain is perhaps most appealing under the appropriate conditions because of its suitability to processing with both deterministic and statistical methods [9–13]. Widely implemented in

the context of projection reconstruction and the equivalent multidimensional Fourier transform, this sampling scheme is desirable in many cases because of the flat baseline outside of the ridges that extend from the peaks. Additionally, it has been shown that relatively few data points are needed to resolve specific spectral information [12]. In the context of (3,2) projection reconstruction and its related techniques, radial sampled data are first processed into two-dimensional tilt planes where the tilt angle is dependent upon the radial sampling angle selected during data collection. Subsequently, various methods can be used to either generate a final spectrum or peak list. In the context of direct multidimensional Fourier transform, the radial sampled data are processed either into single angle multidimensional spectra, with ridges extending from the peak chemical shifts at a vector dependent upon the radial sampling angle. Then the single angle spectra are compared to generate a final spectrum. Alternatively, multiple angle data sets can be combined and Fourier transformed simultaneously to produce a final multidimensional spectrum with ridges extending at all of the sampling angles included. Regardless of the processing method applied, the quality of the final data is directly dependent upon the radial sampling angles chosen during data collection. It is the issue of angle selection that is the focus here.

Two methods have been implemented for angle selection. The first, implemented in the context of HIFI-NMR, uses a probability distribution to determine subsequent angles from an initial data set [14]. The second, implemented in the context of projection reconstruction, selects subsequent angles by choosing the angle

\* Corresponding author. Fax: +1 215 573 7290.

E-mail address: [wand@mail.med.upenn.edu](mailto:wand@mail.med.upenn.edu) (A. Joshua Wand).

that resolves the most peaks from a provisional spectrum [15,16]. In brief, this approach first generates a provisional spectrum from data that has already been collected. Subsequent sampling angles are assessed by generating a skyline projection spectrum for each angle and scoring the skyline projections for the number of resolved peaks it contains. A skyline projection spectrum is generated for each potential sampling angle by taking the maximum value along a vector extending through the provisional spectrum perpendicular to the angle of interest. If two peaks in the provisional matrix are on the same vector only one response is shown in the skyline projection. The next angle is selected by comparing the number of resolved peaks in each skyline projection spectrum and avoiding angles that have relatively few resolved peaks in their skyline projections. As pointed out by the authors, this algorithm functions optimally when a spectrum is not too complex. When the complexity of the spectrum increases the algorithm could falter, primarily from limited resolution in the skyline spectra. Additionally, this method does not present the ability to determine when a sufficient number of radial sampled angle spectra have been collected. This inability arises from not treating the peaks in the provisional spectrum individually, but rather looking at the total number of resolved peaks in the skyline projection. Only in the case where the total number of peaks resolved in the skyline spectrum is equal to the number of peaks in the provisional spectrum is one able to assess that ‘enough’ data have been collected. In the absence of such a condition it is impossible to decipher an authentic peak from an artifact peak. In the absence of a deterministic angle selection algorithm, data collection becomes inefficient and time is potentially wasted by collecting too much data or by collecting data of poor quality (i.e. data with a large number of artifact peaks).

In order to increase the utility of the radial sampling approach we present methods to optimize the set of sampling angles employed. The approaches can be classified into two general situations. The first is when the peak resonance frequencies are known and need to be resolved from artifact and the second is when the peak resonance frequencies are not known and need to be resolved and assigned. The former case corresponds to a need to measure variation in intensity such as in a hydrogen exchange or classical relaxation experiment. For this two algorithms have been developed. One determines the *minimum* set of angles necessary to distinguish authentic peak intensity from artifactual intensity introduced by the Fourier analysis of radially sampled data (i.e. the ridges). The second algorithm determines the fewest angles needed to produce an artifact free spectrum when a lower value comparison is performed. Alternatively, for situation where the peak resonance frequencies are not known, an algorithm is developed to provide for iterative post-acquisition determination of the optimal sampling angles to collect and to provide a definitive conclusion regarding the separation of authentic peak intensity from ridge artifacts. This type of algorithm is essential for the optimized application of radial sampling of data to be employed for *de novo* resonance assignment. Both algorithms are tested in the context of a radial sampled HNC0 processed with the direct multidimensional Fourier transform combined with lower value comparison but are applicable, with minor modifications to the selection criteria, to more sophisticated artifact removal schemes.

## 2. Theory

In a three-dimensional spectrum, radial sampling is accomplished by linking the evolution of the two indirect dimensions by setting  $t_1 = \tau \cos(\alpha)$  and  $t_2 = \tau \sin(\alpha)$ , where  $\tau$  is the incremented time domain and  $\alpha$  is the sampling angle, while continuing to collect the traditional quadrature pairs for both indirect dimensions

[15]. To generate a frequency domain spectrum the data can be processed with a direct single step, 2D Fourier transform [11]

$$S(\omega_1, \omega_2) = \sum_{t_1=0}^{t_1^{\max}} \sum_{t_2=0}^{t_2^{\max}} \exp(-i\omega_1 t_1) \exp(-j\omega_2 t_2) f(t_1, t_2) w(t_1, t_2)$$

where  $i$  and  $j$  are quaternion numbers;  $t_1, t_2$  are the incremented times,  $\omega_1$  and  $\omega_2$  comprise the frequency pair being determined,  $f(t_1, t_2) = \exp(-i\Omega_1 t_1) \exp(-j\Omega_2 t_2)$  is the data being transformed,  $\Omega_1$  and  $\Omega_2$  are the chemical shifts for time domains 1 and 2, respectively, and  $w(t_1, t_2)$  is a weighting factor to account for the non-equidistant sampling of the time domain.

When  $t_1$  and  $t_2$  are incremented independently, in traditional Cartesian fashion, the direct 2D Fourier transform produces the same results as the traditional sequential one-dimensional Fourier transforms. In the case of radial sampling the Fourier transform is effectively underdetermined and produces ridges that extend through the spectrum where Eq. (2) is satisfied

$$\frac{\omega_1 - \Omega_1}{\omega_2 - \Omega_2} = \tan(\alpha) \quad (2)$$

This relationship is true when  $\alpha$  is either positive or negative, leading to two ridges extending from the each peak in the spectrum, one with a positive slope and the other with a negative slope.

We define an ordered triple with the directly detected dimension,  $\omega_3$ , in the first position and the two linked indirect dimensions,  $\omega_1$  and  $\omega_2$ , in the second and third positions, respectively. The following linear equation describes the ridge extending from a peak located at point  $P_1$  in the so-called (3,2) radially sampled experiment, where we employ the nomenclature of Szyperski and Atreya [17]

$$P = P_1 + n(0, \cos(\pm(90 - \alpha)), \sin(\pm(90 - \alpha))) \quad (3)$$

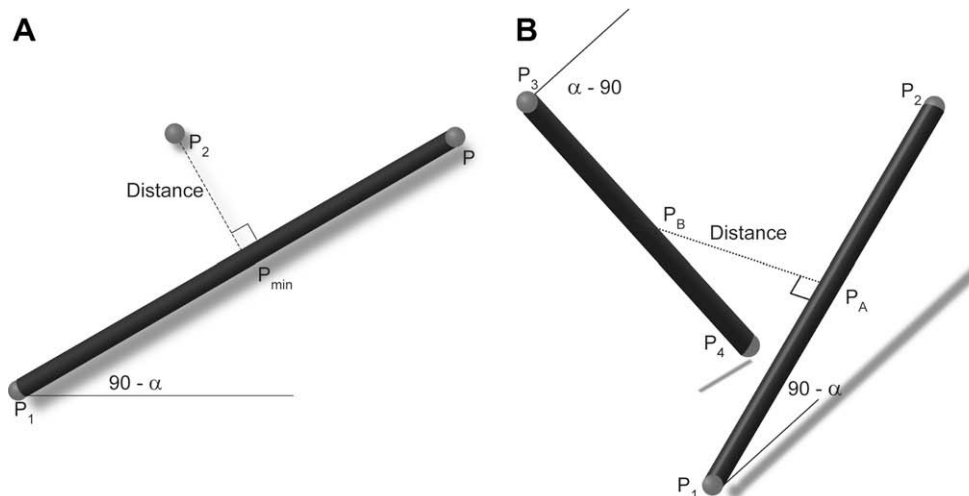
where  $P$  represents a point on the ridge,  $\alpha$  is the sampling angle and  $n$  is a scalar. As before, the  $\pm$  sign is included because two ridges extend, one with a positive slope and another with a negative slope. In the case of a (4,2) radially sampled experiment four ridges would extend from each peak. In this case, Eq. (3) is expanded to account for two sampling angles,  $\alpha$  and  $\beta$ , as described by the following equation:

$$P = P_1 + n(0, \cos(\pm(90 - \alpha)) \cos(\pm(90 - \beta)), \sin(\pm(90 - \alpha)) \times \cos(\pm(90 - \beta)), \sin(\pm(90 - \beta))) \quad (4)$$

These basic descriptions allow the determination of whether two peaks are resolved at a given sampling angle and where all of the potential artifact positions are located. Further, this description allows all peaks to be analyzed simultaneously, regardless whether they are resolved in the directly detected dimension.

### 2.1. Peak–peak resolution

Two peaks in a radially sampled experiment are not resolved if the ridge from one of the peaks intersects the second. To determine if two peaks are resolved the distance from one of the peaks to the closest points on the positive and negative ridge components of the other peak is determined. If both distances are greater than a specified cutoff (chosen to reflect a finite line width), the peak is considered resolved. The distance measurement is illustrated in Fig. 1A, where the peaks are represented by points  $P_1$  and  $P_2$ . For clarity only one of the ridge components is shown in the figure. The distance between  $P_2$  and the ridge from  $P_1$  is determined by applying the point to line distance algorithm commonly encountered in computer graphics [18]. Here we generalize this approach. The first step is to define an equation in order to solve for point  $P_{\min}$ , the closest point on the ridge to the peak located at  $P_2$



**Fig. 1.** Illustration of the peak to ridge distance in 3D space (A). Given two peaks, located at  $P_1$  and  $P_3$ , the shortest distance is calculated between the peak located at position  $P_3$  and the nearest point on the ridge, extending from the peak at position  $P_1$ , located at point  $P$ . See text for details regarding this distance calculation. If the distance is greater than a specified cutoff the two peaks are resolved at the given sampling angle  $\alpha$ . Illustration of the ridge to ridge distance (B). The ridge to ridge distance is used to determine if an artifact is generated from the intersection of ridges extending from peaks at locations  $P_1$  and  $P_3$ . Additionally, if the distance between the two ridges is less than a specified cutoff, the artifact position is determined as the average of the closest points  $P_A$  and  $P_B$ . See text for details regarding this calculation.

$$P_{\min} = P_1 + u(P - P_1) \quad (5)$$

To determine  $P$  an arbitrary non-zero scalar  $u$  is used. When the distance between point  $P_2$  and  $P_{\min}$  is minimized the vector from  $P_{\min}$  to  $P_2$  is perpendicular to the ridge. Therefore, the dot product of the two vectors is zero

$$(P_2 - P_{\min}) \cdot (P - P_1) = 0 \quad (6)$$

In order to solve for the point  $P_{\min}$ , Eq. (5) is substituted into the dot product relationship and the scalar  $u$  is determined

$$u = \frac{P_2 \cdot (P - P_1) - P_1 \cdot (P - P_1)}{(P - P_1) \cdot (P - P_1)} \quad (7)$$

Finally, the expression for  $u$  is used to determine the point  $P_{\min}$

$$P_{\min} = P_1 + \frac{P_2 \cdot (P - P_1) - P_1 \cdot (P - P_1)}{(P - P_1) \cdot (P - P_1)} (P - P_1) \quad (8)$$

The distance between  $P_{\min}$  and  $P_2$  is then

$$\text{distance} = \sqrt{(P_{\min} - P_2) \cdot (P - P_2)} \quad (9)$$

The distance defined above corresponds to an infinitely narrow line. To accommodate consideration of a finite linewidth the effective width along the line connecting the two points of interest must be determined. This is accomplished by setting the origin of the Cartesian basis at point  $P$  then defining two angles between points  $P_{\min}$  and  $P_2$  that would be used to describe the latter's position with respect to the former in a polar basis. These two angles are defined as:

$$\alpha_1 = \arctan \left[ \frac{P_{\min}(\omega_1) - P_2(\omega_1)}{P_{\min}(\omega_3) - P_2(\omega_3)} \right] \quad (10a)$$

$$\alpha_2 = \arctan \frac{|P_{\min}(\omega_2) - P_2(\omega_2)|}{\sqrt{(P_{\min}(\omega_1) - P_2(\omega_1))^2 + (P_{\min}(\omega_3) - P_2(\omega_3))^2}} \quad (10b)$$

Here the subscript defines the Cartesian chemical shift components of vector  $P_{\min}$  and  $P_2$ . Note that the denominator is zero if the angle is  $90^\circ$ . The linewidths along the specified distance line can be determined using the above defined angles as follows:

$$\text{linewidth}_{\omega_1} = (\text{cartesian linewidth}_{\omega_1})(\sin \alpha_1)(\cos \alpha_2) \quad (11a)$$

$$\text{linewidth}_{\omega_2} = (\text{cartesian linewidth}_{\omega_2})(\sin \alpha_2) \quad (11b)$$

$$\text{linewidth}_{\omega_3} = (\text{cartesian linewidth}_{\omega_3})(\cos \alpha_1)(\cos \alpha_2) \quad (11c)$$

The effective linewidth for a peak along a given vector is then the Euclidean distance of the scaled components

$$\text{linewidth} = \left( \sum_{q=\omega_1, \omega_2, \omega_3} \text{linewidth}_q^2 \right)^{1/2} \quad (12)$$

The same scaling components can be used for both the peak at  $P_2$  and the ridge at  $P_{\min}$  because of the mirror symmetry between them. Note that the linewidth at point  $P_{\min}$  is the same as the linewidth at peak  $P_1$ . Finally, a measure of resolution is obtained by subtracting the two linewidths from the distance measured above

$$\text{resolution distance} = \text{distance} - \text{linewidth}_{P_1} - \text{linewidth}_{P_2} \quad (13)$$

The correction for finite line width in a higher dimensional experiment expands accordingly while the minimum distance algorithm remains unchanged. In the case of a (4,2) experiment four linewidths will be scaled using three angles defined in the same manner as Eqs. (10a) and (10b). The four scaling components are:  $(\sin \alpha_1)(\cos \alpha_2)(\cos \alpha_3)$ ,  $(\sin \alpha_2)(\cos \alpha_3)$ ,  $(\sin \alpha_3)$  and  $(\cos \alpha_1)(\cos \alpha_2)(\cos \alpha_3)$  for the  $\omega_1$ ,  $\omega_2$ ,  $\omega_3$  and  $\omega_4$ , respectively. Where  $\omega_1$ ,  $\omega_2$  and  $\omega_3$  are the indirectly detected dimensions and  $\omega_4$  is the directly detected dimension. Additionally, this treatment also provides a mechanism for filtering a peak list to allow for authentic peaks that will never be resolved to be treated as one peak. For example, one peak that encompasses two non-resolved peaks can be set to have a peak position equal to the average of the two peaks and a broader linewidth to account for both peaks.

## 2.2. Potential artifact positions

The lower value algorithm efficiently removes ridge artifacts if an appropriate combination of angle spectra are compared [15]. In instances where an inappropriate (insufficient) number of angle spectra are employed, ridge intensity may be present at positions in the spectrum not corresponding to authentic peak intensity and represents an artifact in the spectrum. The artifacts occur at locations when multiple ridges intersect. Therefore,

determining all possible ridge intersection points can lead to the identification of artifact peaks. If the ridge linewidths were infinitely narrow the potential artifact locations would be the solution to the set of linear equations describing the ridges. In order to accommodate finite linewidths a ridge to ridge distance algorithm is used. The algorithm is an application of the general line to line distance algorithm also often used in computer graphics [18]. If the distance between two ridges is less than a specified cutoff, the average of the two closest points on each ridge is marked as a potential artifact. This procedure is illustrated in Fig. 1B. Here two ridges extend from points  $P_1$  and  $P_2$  and points  $P_A$  and  $P_B$  represent the closest points between the two ridges. This figure illustrates only one of the ridge components from each peak. The total number of ridges is defined by the sampling scheme as discussed above.

The two ridges of Fig. 1B are represented by the linear equations (14a) and (14b):

$$P_A = P_1 + \alpha(P_2 - P_1) \quad (14a)$$

$$P_B = P_3 + \alpha(P_4 - P_3) \quad (14b)$$

Here  $P_1$  and  $P_3$  are the positions of authentic peaks. Points  $P_2$  and  $P_4$  are defined as a function of the sampling angle, as presented in Eqs. (3) and (4) for a non-zero scalar  $n$ .  $a$  and  $b$  are the scalars used to define points the closest points  $P_A$  and  $P_B$ . A vector  $W$  can be defined between the two closest points as:

$$W = P_A - P_B \quad (15)$$

As before, in order to solve for  $P_A$  and  $P_B$  the scalars  $a$  and  $b$  must be determined.  $W$  becomes

$$W = P_1 + a(P_2 - P_1) - P_3 - b(P_4 - P_3) \quad (16)$$

For clarity we can define a vector from the peak at  $P_1$  to the peak at  $P_3$

$$W_0 = P_1 - P_3 \quad (17)$$

The simplified expression for  $W$  is presented by substituting Eq. (17) into Eq. (16)

$$W = W_0 + a(P_2 - P_1) - P_3 - b(P_4 - P_3) \quad (18)$$

From the definition of two skew lines we know the vector describing the line between the closest points is the only line uniquely perpendicular to both the lines describing the ridges. Therefore, the dot product and unit vectors  $(P_2 - P_1)$  and  $(P_4 - P_3)$  that describe the two ridges is zero

$$W \cdot (P_2 - P_1) = 0 \quad (19a)$$

$$W \cdot (P_4 - P_3) = 0 \quad (19b)$$

Substituting the expression for  $W$  into the dot product definitions puts the equations in terms of the scalars  $a$  and  $b$ :

$$W_0 \cdot (P_2 - P_1) = b(P_2 - P_1) \cdot (P_4 - P_3) - a(P_2 - P_1) \cdot (P_2 - P_1) \quad (20a)$$

$$W_0 \cdot (P_4 - P_3) = b(P_4 - P_3) \cdot (P_4 - P_3) - a(P_4 - P_3) \cdot (P_2 - P_1) \quad (20b)$$

For clarity the following scalars are defined:  $h = W_0 \cdot (P_2 - P_1)$ ,  $i = (P_2 - P_1) \cdot (P_2 - P_1)$ ,  $j = (P_2 - P_1) \cdot (P_4 - P_3)$ ,  $k = W_0 \cdot (P_4 - P_3)$  and  $l = (P_4 - P_3) \cdot (P_4 - P_3)$ . Substituting into Eqs. (20a) and (20b) gives:

$$a = \frac{ki - hl}{jl - i^2} \quad (21a)$$

$$b = \frac{jk - hi}{jl - i^2} \quad (21b)$$

Upon substitution of Eqs. (21a) and (21b) into Eqs. (14a) and (14b), the points  $P_A$  and  $P_B$  become:

$$P_A = P_1 + \frac{ki - hl}{jl - i^2} (P_2 - P_1) \quad (22a)$$

$$P_B = P_3 + \frac{jk - hi}{jl - i^2} (P_4 - P_3) \quad (22b)$$

The Euclidean distance between points  $P_A$  and  $P_B$  is defined as:

$$\text{distance} = \sqrt{(P_A - P_B) \cdot (P_A - P_B)} \quad (23)$$

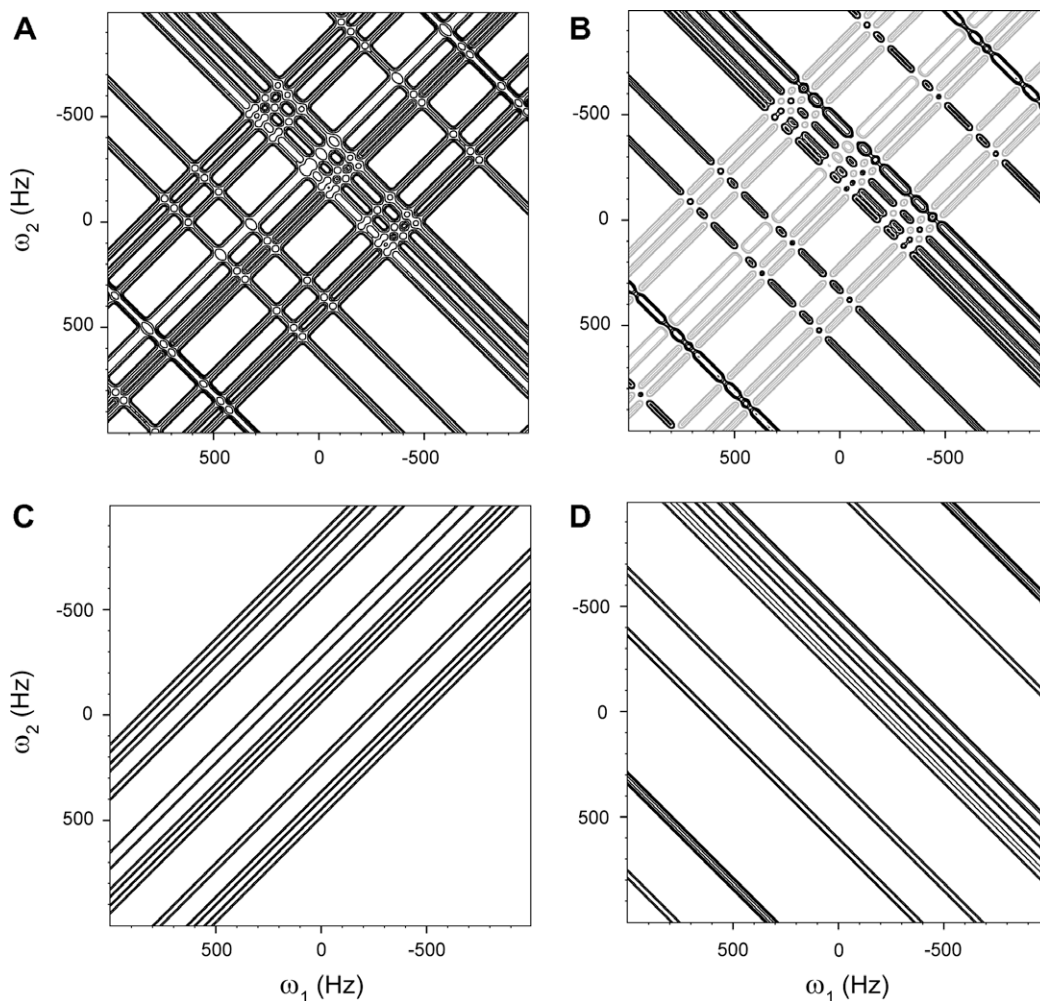
In order to determine if an artifact is present the distance is scaled for the line widths in the same manner as in the peak to ridge distance. If the scaled distance is less than a specified cutoff a potential artifact is located at the average of points  $P_A$  and  $P_B$ . Again, this algorithm is also applicable to higher dimensional experiments.

We now have the tools necessary to optimize the collection of radially sampled data.

### 2.3. Minimum angles to resolve peak intensities

The first case that we consider is the situation where the positions of authentic peaks are known. This would be encountered during the collection of three-dimensionally resolved relaxation or hydrogen exchange data, for example. Here the goal is to collect data as efficiently as possible such that all authentic peaks are free from contaminating artifact intensity. Importantly, in this situation, artifact intensity that is resolved from known authentic peak intensity is of no consequence.

Relaxation rates vary as a function of angle for radial sampled experiments because the data are a product of two relaxation components, one from each of the two indirectly evolved dimensions (spins). The variation in relaxation can be eliminated by using a single sampling angle for a series of experiments. In turn, treating the angles independently, allows for the ridge intensity to be left in the spectrum, skipping any lower value comparison. To increase the number of peaks resolved at a sampling angle the positive and negative ridge components are isolated and analyzed separately. The ridge components are isolated in the same manner as presented in our recent description of how to phase correct radially sampled data [19]. In brief, two spectra are generated, one using the matching two-dimensional Fourier transform and the other using a non-matching Fourier transform (sin-sin Fourier transform for cos-cos modulated data). The difference of these two spectra generates the positive slope ridge component and the sum generates the negative slope ridge component as demonstrated in Fig. 2. If all authentic peaks are not resolved from ridge intensity with a single sampling angle, multiple sampling angles can be used but the resulting data should be treated independently. Treating the angle spectra independently, allows for a simple algorithm to determine the optimal sampling angles. The first step in the algorithm is to edit the peak list grouping authentic peaks that are not resolved from each other as opposed to resolved from artifactual peak intensity. Unresolved authentic peaks will not be resolved by any sampling angle and are therefore treated as a single peak with a linewidth that spans the group of peaks. After the peak list has been edited every combination of peaks is tested for resolution from artifact intensity using the peak to ridge distance algorithm for a selected series of angles. The peak to ridge distance accounts for a difference of the chemical shifts in the directly detected dimension avoiding the need to sort the peaks to assure they are in the same indirect plane of the spectrum. This step generates two lists of peaks for the sampling angle tested, one for the peaks resolved in the positive slope component spectrum,



**Fig. 2.** Demonstration of how to separate angle spectra into ridge component spectra. For this example a 10 peak quaternion data set was generated using a  $45^\circ$  sampling angle. Spectrum A shows the absorptive spectra generated by applying the matching direct 2D Fourier transforms to the four data components. Spectrum B shows the spectrum generated by applying the non-matching direct 2D Fourier transforms, for example the  $\sin$ - $\sin$  2D Fourier transform was applied to the  $\cos$ - $\cos$  modulated data component. Transforming with the non-matching 2D Fourier transform produces a spectrum with negative intensity (grey) for the positive sloped ridge and positive intensity (black) for the negative sloped ridge. Spectrum C shows the isolated positive sloped ridge component spectrum, generated by taking the difference of spectrum B from spectrum A. Spectrum D shows the negative sloped ridge component spectrum generated by summing spectra A and B.

and another for the peaks resolved in the negative slope component spectrum.

The results of a series of sampling angles can be sorted to determine the minimum number and identity of the sampling angles needed to resolve the intensity of all of the authentic peaks in a spectrum. The two lists of resolved peaks at each angle are combined and any redundancy removed; some peaks will be resolved in both of the component spectra. The angle that resolves the most peaks is then found. If the selected angle fails to resolve all of the peaks additional angles are selected on the basis of resolving the most peaks.

This procedure was tested with a simulated data set consisting of 10 peaks, all located in the same plane. The peaks were randomly distributed in two dimensions, with the criteria that they would not be resolved by only one of the two dimensions. The results are shown in Fig. 3. For comparison the same peak frequencies and linewidths were used to generate a Cartesian sampled data set resulting in the spectrum shown in Fig. 3A. Analysis of the peak list concluded that an  $85^\circ$  sampling angle would resolve all of the peaks. The positive slope ridge component of the  $85^\circ$  spectrum is shown in Fig. 3B. For clarity the Cartesian sampled spectrum is overlaid in gray.

#### 2.4. Minimum sampling angles to generate an artifact free spectrum

In the second scenario that is likely to be encountered, an artifact free spectrum is desired. To produce such a spectrum the lower value algorithm is used to remove the artifacts, the success of which is dependent upon the collection of an appropriate set of sampling angles. If suboptimal sampling angles are used intensity can remain at non-authentic peak locations. In a manner similar to above, we apply the ridge to ridge distance algorithm to determine all of the potential artifact positions, and subsequently apply the peak to ridge distance algorithm to determine if potential artifact positions are resolved in at least one of the selected sampling angles and will consequently be removed by the lower value algorithm.

As before, the first step is to edit the peak list to combine truly unresolved peaks. Sets of unresolved peaks are replaced by a single peak with an adjusted linewidth to account for their unresolved components. Unlike for the previous case described above, the sampling angles are no longer independent, which requires sets of angles to be selected. The number of angles and which angles selected can be definitively determined. If some angles must necessarily be collected, such as the  $0^\circ$  and  $90^\circ$  used to determine phase corrections [19], they can be included in every set of angles tested.

Typically, initial tests use a small number of angles and increase the number if all of the artifacts are not removed after a given number of trials.

For a given set of test angles, all of the artifact positions are determined through application of the ridge to ridge distance algorithm. This is accomplished by iterating over each set of angles for each peak against every other ridge. For example, in the case where there are two peaks,  $P_1$  and  $P_2$ , and two sampling angles,  $\alpha_1$  and  $\alpha_2$ , the ridge extending from peak  $P_1$  at  $+\alpha_1$  would be tested against the  $-\alpha_1$  ridge of  $P_2$  and both the  $+\alpha_2$  and  $-\alpha_2$  ridge of  $P_2$ . The other three ridge components of  $P_1$  are tested in the same manner. To speed the analysis, only those peaks that are not resolved in the directly detected dimension are tested. The list of all potential

artifacts is then edited to remove peaks that are in the same location as the authentic peaks. If the primary concern is to determine the chemical shifts, removing the overlapping artifacts from the list will not affect the final spectrum, it can only affect the peak intensities. If the intensities are a concern the minimum angle to resolve peak intensity algorithm can be run first to select the angles that resolve intensities. The intensities can then be extracted from the appropriate angle spectra.

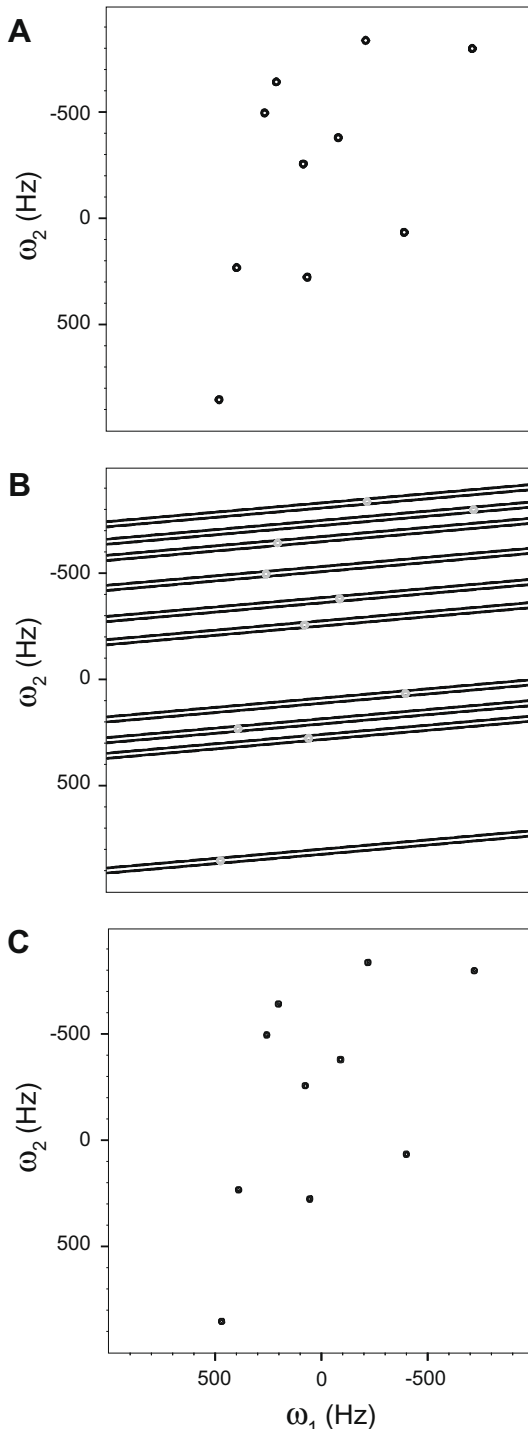
The final step is to determine if the potential artifacts will be removed during lower value comparison. This is accomplished by applying the peak to ridge distance algorithm. The ridges will only extend from the authentic peaks, and not the potential artifact positions. Accordingly, the potential artifact positions are tested against all of the ridges extending from the authentic peaks. If an artifact position is resolved in one of the angles, the potential artifact will be removed during the lower value comparison. A list of unresolved artifacts is thereby compiled. If the number or location of the remaining artifacts is unsatisfactory a new set of angles is then tested.

This algorithm was test against same 10 peak generated data test case used in the previous algorithm. Analysis of the peak list concluded that two sampling angles,  $6^\circ$  and  $85^\circ$  are needed to remove all of the artifacts. The results are shown in Fig. 3C. Comparison with the Cartesian sampled spectrum indicates that only authentic peak intensity remains after the lower value comparison.

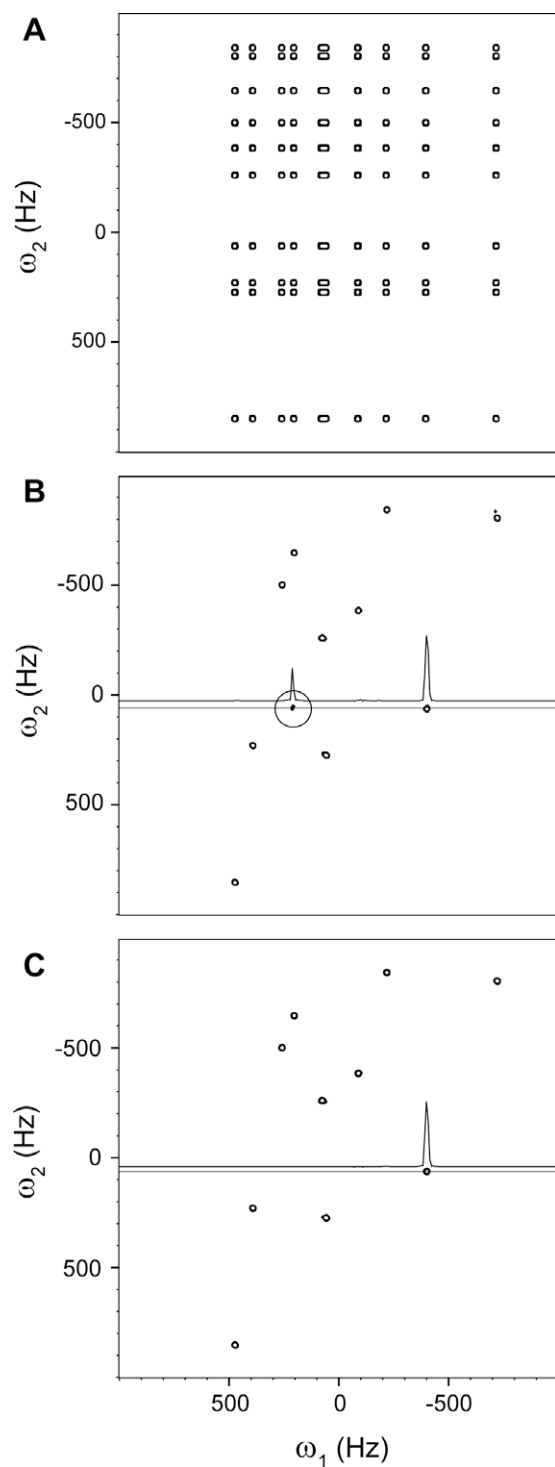
### 2.5. Spectrum analysis and iterative data collection

In the two previous scenarios, the peak positions are known and the appropriate angles to either resolve peak intensities or resolve all of the artifacts can be determined unambiguously. In situations where the position of the authentic peaks are not known rather than testing for sampling angles that resolve the potential artifacts, the remaining peak intensity in a lower value comparison spectrum needs to be tested. Without knowing the location of the authentic peaks, all intensity in a lower value spectrum must be treated as a potential peak until it is determined to be authentic. If the intensity in a lower value spectrum is resolved in at least one angle, the peak must be authentic. If two peaks are not resolved they are marked as potential artifacts until additional data resolves them.

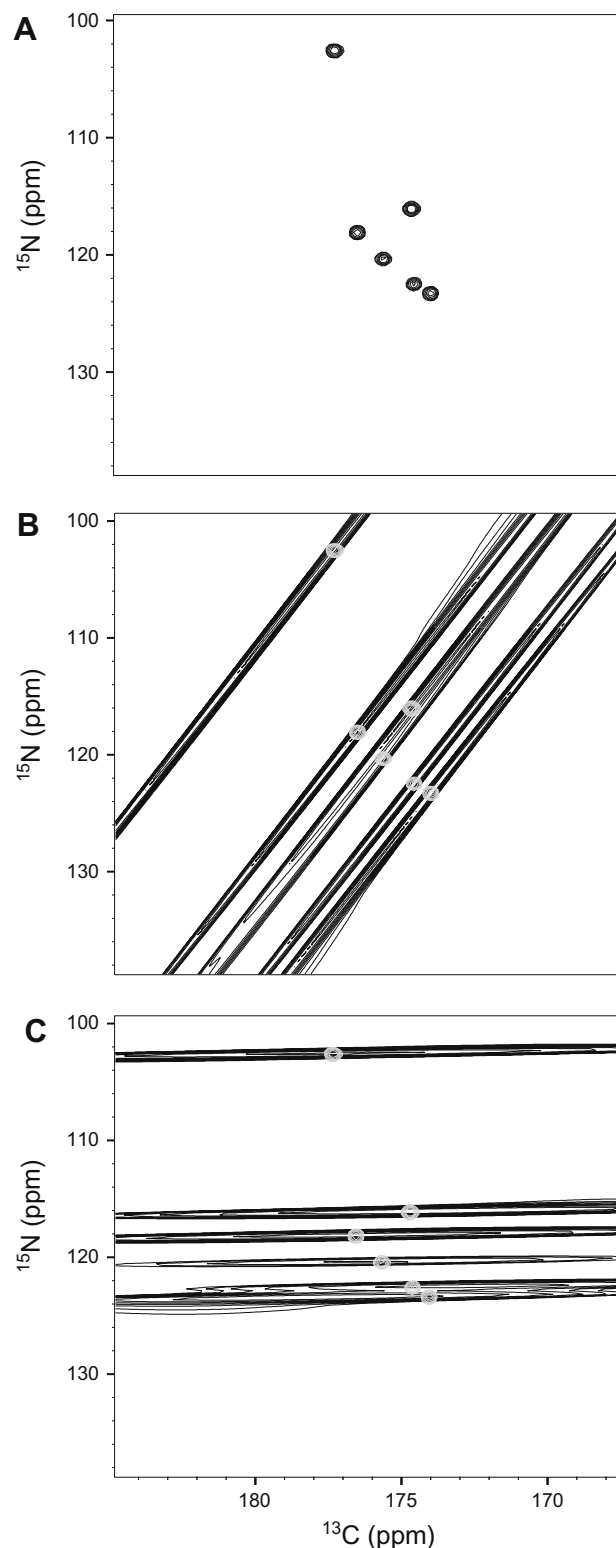
The first step in this analysis is to collect an initial data set, process the angles separately, compare them with the lower value algorithm and generate a peak list. All such peaks are considered potentially authentic at this point. The peak to ridge distance algorithm is applied to test if a potential peak is resolved from the ridges from all of the other potential peaks. The ridges are generated at each of the sampling angles used. If a peak is not resolved at any of the sampling angles, it is marked as a potential artifact. After a list of potential artifacts is generated, the set of minimum angles to resolve all of the potential peak intensities is determined as described above. Additional data are then collected at the



**Fig. 3.** Comparison of Cartesian and radial sampled spectra illustrating how appropriate angle selection can speed data collection. Spectrum A shows the Cartesian sampling spectrum. Spectrum B demonstrates the selection of the minimum angles needed to resolve all of the peak intensities. For this set of peaks the positive slope component of  $85^\circ$  sampling angle resolves all of the peaks. Data were processed with the matching and non-matching direct two-dimensional Fourier transform to isolate the positive ridge component. For clarity the Cartesian sampled spectra was overlaid in gray. Spectrum C demonstrates selecting the minimum angles to produce a spectrum with no artifact peaks. Here data were generated with two radial sampling angles,  $6^\circ$  and  $85^\circ$ . Data were processed with the matching and non-matching direct two-dimensional Fourier transforms to isolate the positive and negative ridge components of the two angles generating four spectra. The four spectra were compared with the lower value algorithm to produce an artifact free spectrum. A slice take at 60 Hz is overlaid, with slight offset for clarity, demonstrates an artifact free baseline.



**Fig. 4.** Demonstration of iterative angle selection and spectrum analysis. Spectrum A shows the lower value comparison of the face,  $0^\circ$  and  $90^\circ$  sampling angle, spectra. The peak list of the spectrum A was analyzed and a  $35^\circ$  sampling angle was determined to remove the most artifacts. Spectrum B shows the lower value comparison of the  $0^\circ$ ,  $35^\circ$  and  $90^\circ$  sample angle spectra. The circled peak indicates a peak in the spectrum that was determined to be a possible artifact. The overlaid slice demonstrates the relative intensity of the possible artifact peaks. Analysis of the peak list from this spectrum determined that a sampling angle of  $73^\circ$  would remove any remaining artifacts. Spectrum C shows the lower value comparison of the resulting spectra from sampling angles at  $0^\circ$ ,  $35^\circ$ ,  $73^\circ$  and  $90^\circ$ . Analysis of the peak list from this spectrum determines that all peaks are resolved at least one of the sampling angles and therefore must be authentic peaks and not artifacts. The overlaid slice demonstrates the removal of the artifact peak. The slices in B and C are slightly offset for clarity.



**Fig. 5.** Demonstration of the minimum angles needed to determine the peak intensities for ubiquitin using an HNCO. Peak list analysis determined that sampling angles of  $36^\circ$  and  $90^\circ$  would resolve all of the peak intensities. Shown here are the  $^{13}\text{C}$ - $^{15}\text{N}$  indirect planes of three HNCO spectra at  $^1\text{H}$  shift of 8.15 ppm. Spectrum A shows Cartesian sampled experiment as a reference. Spectrum B shows the  $90^\circ$  sampling angle spectra while spectrum C shows the positive ridge component of the  $36^\circ$  sampling angle. Note that the peaks that are not resolved at  $90^\circ$  are resolved at  $36^\circ$ .

suggested angles. Data are processed and compared, using the lower value algorithm, to the previous spectrum. A new peak list is

created and analyzed and the process is repeated until all of the intensity is resolved, or the remaining potential artifacts do not complicate further analysis.

Fig. 4 demonstrates this method of iterative analysis and data collection. Here the same 10 peak test case was used as before. For the first round of data collection, data were generated at 0° and 90°, processed independently and compared with the lower value algorithm. The resulting lower value spectrum is shown in Fig. 4A. As anticipated, it is impossible to determine the 10 authentic peaks from the 100 potentially authentic peaks. Analysis of the peak list generated from the spectrum in Fig. 4A led to the conclusion that a sampling angle of 35° was optimal. An additional data set was generated at 35°, processed and compared using the lower value algorithm to the 0° and 90° lower value spectrum. As shown in Fig. 4B, inclusion of the additional sampling angle removed a large number of the additional of the potential artifact peaks, leaving 11 peaks in the spectrum. The one artifact in the spectrum is circled to highlight that without the analysis described here it is impossible to distinguish it from an authentic peak. A subsequent iteration determined that a sampling angle of 73° would resolve all of the remaining peaks, removing any potential artifacts. Generating the additional data set, processing and comparison to the 0°, 90° and 35° lower value spectrum produces the spectrum shown in Fig. 4C. Analysis of this spectrum's peak list determines that all of the peaks must be authentic because they are all resolved from ridge artifacts.

### 3. Results

We have essentially described three algorithms: finding a minimum set of sampling angles to resolve authentic intensities from ridge artifact intensity; finding a minimum set of sampling angles to remove all ridge artifacts from the spectrum; and an iterative analysis and data collection procedure for obtaining an artifact free spectrum when the positions of authentic peaks are not known *a priori*. Each procedure was tested in the context of the HNC0 spectrum of recombinant human ubiquitin. The results are illustrated in Figs. 5–7. To establish the minimum set of sampling angles to resolve authentic intensity, an initial peak list was derived from the conventional Cartesian sampled HNC0 spectrum (Fig. 5A) though such a <sup>15</sup>N, <sup>13</sup>C list could be taken from any reliable source. This spectrum also served as a comparison with equivalent resolution to the radial sampled spectrum. High resolution

was achieved by collecting 64 increments in both indirect dimensions. Accordingly, the data collection time was approximately 36 h. Analysis of the peak list suggested sampling angles of 36° and 90° would resolve all authentic peak intensity from artifactual intensity. Indirect dimension slices of the single step two-dimensional FT of the positive slope component of 36° and 90° sampling angle data sets are shown in Fig. 5B and C. Total data collection time for the two angle planes was 51 min corresponding to a 43-fold time advantage over Cartesian sampling. These slices are all taken at 8.15 ppm in the <sup>1</sup>H acquisition dimension ( $\omega_3$ ). Importantly, when measuring peak intensities it is clearly advantageous to separate the spectrum into the individual ridge components. This aids in determining the intensity because distracting artifact peaks are not present. Additionally, by separating the spectrum into its ridge components fewer sampling angles need to be collected. The spectra are not symmetrical so each ridge component contains unique data. Separating the ridges also aids in removing artifact peaks if the lower value algorithm is applied. In summary, these spectra demonstrate the successful resolution of authentic peak intensity from artifactual ridge intensity in a deterministic manner. Analysis of the entire 3D spectrum derived from the two sampling angles confirmed that all of the peak intensities are resolved (data not shown).

The two angles selected to resolve peak intensities are not a unique solution. However, the combination of 36° and 90° was selected for multiple reasons. First, it is advantageous to include the 0° or 90° projections or “faces” as they are needed to determine the phase corrections [19]. Additionally, the 0° and 90° faces can be collected with two quadrature components as compared to four needed for other angles. Finally, the faces are only affected by relaxation arising from spins associated with only one incremented time domain.

In cases where the spectrum is especially complex, it might not be possible to choose a set of angles that resolve all peak intensities. In this circumstance either a subset of peaks must be focused on or an alternate experiment must be chosen. When a subset of the peaks is focused on, the sorting routine can be modified to include a weighting term to favor the angles that resolve the peaks of interest. While the time savings occurred by radial sampling make it appealing, the algorithm described here allows a definitive mechanism for deciding whether radial sampling is applicable.

The same approach was used to test the procedure for defining the minimum set of angles necessary to remove all artifacts from the spectrum. Analysis of the peak list concluded that three

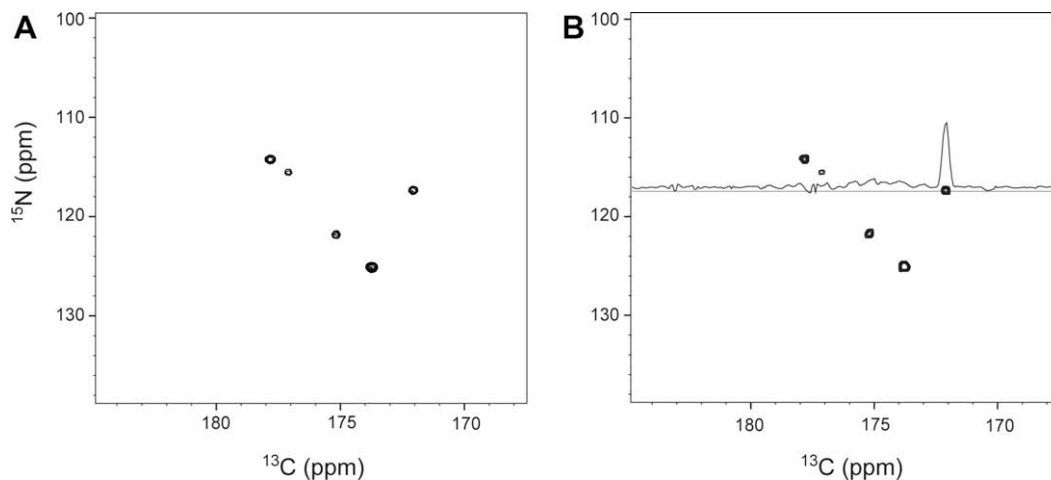
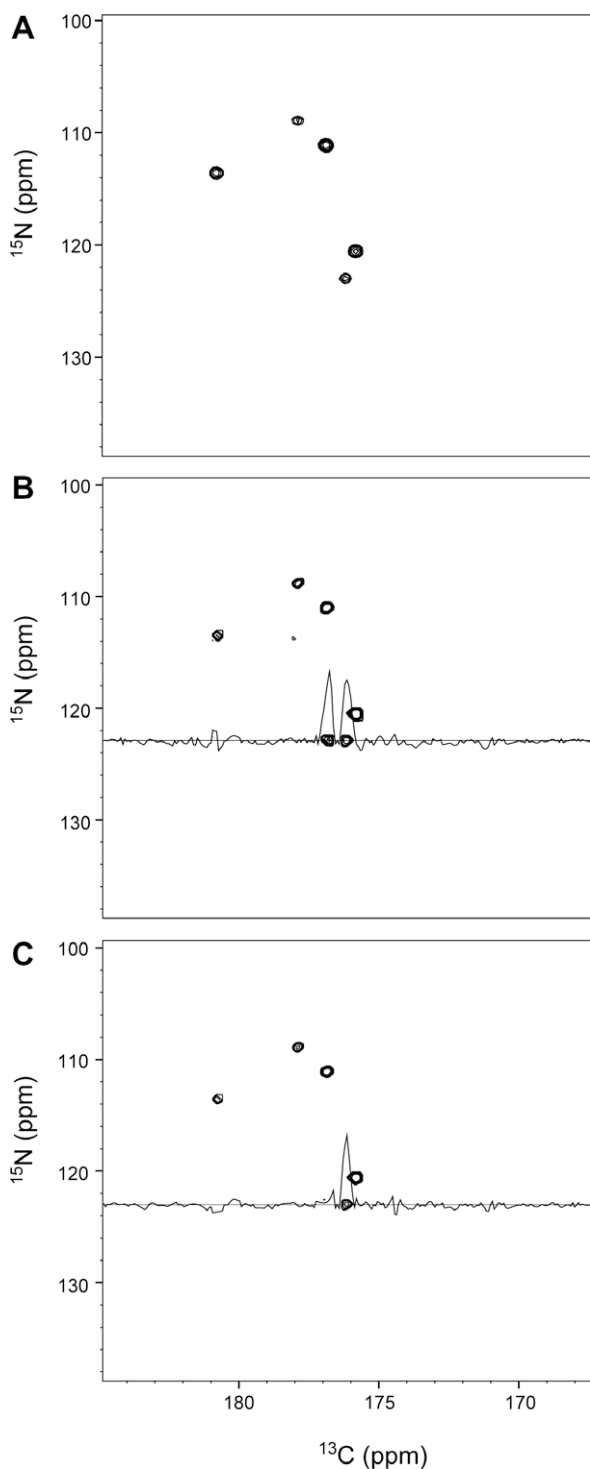


Fig. 6. An example of calculating the fewest angles needed to generate an artifact free HNC0 spectrum of ubiquitin is shown here. Spectrum A shows the comparison Cartesian sampled spectrum at <sup>1</sup>H 8.71 ppm and Spectrum B shows the same indirect slice of the radial sampled experiment using the calculated sampling angle of 0°, 36° and 90°. The overlaid slice demonstrates the typical baseline quality for the entire 3D spectrum.





**Fig. 7.** Demonstration of the use of iterative angle selection to generate an artifact free HNCOSY spectrum of ubiquitin with radial sampling. For comparison spectrum A shows an indirect slice of the Cartesian sampled HNCOSY at  $^1\text{H}$  8.49 ppm. Spectrum B shows the same indirect plane as A for the radial sampled data generated from the lower value comparison of  $0^\circ$ ,  $45^\circ$  and  $90^\circ$  sampling angle spectra. Analysis of the peak list for the entire 3D experiment concludes that a sampling angle of  $64^\circ$  will remove the most remaining artifacts, if any. Spectrum B shows the lower value comparison of  $0^\circ$ ,  $45^\circ$  and  $90^\circ$  with a newly collected  $64^\circ$  spectrum. Notice the removal of one artifact peak, as demonstrated by the overlaid slices. Analysis of this peak list concludes that all peaks in the spectra are resolved and therefore authentic.

sampling angles ( $0^\circ$ ,  $35^\circ$  and  $90^\circ$ ) would suffice. The total measurement time for the three angles is 68 min corresponding

to a 32-fold time advantage over equivalent resolution Cartesian sampling. A representative slice of the HNCOSY spectrum illustrates that only the desired authentic intensity is present (Fig. 6). Again, the three angles selected by the algorithm to remove all of the artifacts are not unique; other combinations of angles would produce the equivalent results. In this case, the  $0^\circ$  and  $90^\circ$  sampling angles were required to be in the angle set in order to determine the necessary phase corrections [19]. Other additional angles could be included in the angle set. Time is the only disadvantage to including more angles if the minimum angles are known. Including additional angles will not produce ridge artifacts.

The number of angles needed to remove all of the artifacts can be decreased by reducing the linewidth of the peaks. The algorithm is based on a distance measurement; therefore the effective distance between the peaks is optimized by reducing the linewidths. Standard methods can be used to decrease the linewidths. Increasing the number of increments if relaxation is not limiting or using constant time approaches where the line widths is adjusted by the convolution and apodization functions are two obvious options. Effective use of linear prediction adapted to radially sampled experiments would also decrease the linewidths with a concomitant reduction in the minimum number of sampling angles required.

The final example illustrates the iterative data analysis and collection procedure used to faithfully reveal authentic peaks while suppressing artifactual intensity without prior knowledge of the peak positions. Fig. 7 shows a representative indirect plane of the HNCOSY generated from lower value comparison of three sampling angles ( $0^\circ$ ,  $45^\circ$  and  $90^\circ$ ). This was used as a starting point. Analysis of the peak list determined that a sampling angle of  $67^\circ$  would resolve the most additional peaks in the spectrum. The  $67^\circ$  sampling angle spectrum was collected and processed and compared to the ( $0^\circ$ ,  $45^\circ$  and  $90^\circ$ ) lower value comparison spectrum. A representative slice is shown in Fig. 7. Analysis of the resulting peak list concluded that all of the peaks were resolved and therefore authentic.

#### 4. Discussion

The three algorithms described here provide a means to confidently collect radially sampled multidimensional NMR data that such that the integrity of peak intensity is maintained (algorithms 1 and 2) or the spectrum entirely free of artifacts arising from ridge intensity inadvertently surviving the lower value data reduction (algorithm 3). The retrospective spectrum analysis described here removes all uncertainty as to whether a peak is authentic or artifact through a quantitative measure of resolution. Furthermore, the approach optimizes the data collection by reducing if not eliminating the collection of unnecessary data and identifying when sufficient data have been collected to produce a suitable spectrum. From a practical point of view, any inefficiency that is introduced by the analysis during data collection can be overcome by collecting angles of other experiments while it is being performed. Typically, assignment experiments are run as pairs so a second experiment is collected concurrently. Regardless, the analysis is rapid and not computationally intensive. Additionally, once the first peak list from the initial data set is generated, additional rounds or analysis are much faster. Automation could be applied to this step very easily. Though only a (3,2) radially sampled HNCOSY spectrum was used to illustrate the potential of the three algorithms described here, all of the methods presented are directly amenable to higher dimensional experiments. Iterative data analysis and collection is particularly appealing in high order  $n\text{D}$  experiments where sensitivity and resolution are generally limiting. While radial sampling affords an easy method to increase the resolution of such experiments, optimal data collection allows for less

angles to be collected and more time to be used for signal averaging with the attendant gain in signal-to-noise. Future work will assess optimized radial sampling in sensitivity limiting cases and in the case of four-dimensional spectra.

## 5. Methods

All simulated data were created using a set of 10 peaks distributed in two dimensions to simulated the two linked indirect dimensions of a (3,2) sampled experiment. The randomly assigned resonance frequencies of the 10 peaks are: (248.9, -503.4); (-97.7, -387.2); (-226.5, -844.5); (67.6, -263.5); (462.7, 845.5); (-407.8, 58.3); (194.1, -649.1); (380.9, 224.9); (47.9, 269.3) and (-727.8, -806.1) Hz. The linewidths of all of the peaks was set to 50 Hz in both dimensions. The spectral width was set to 2000 Hz in both dimensions. Each sampling angle used was the result of four quadrature data components collected with 128 increments.

NMR data were collected on a 900  $\mu\text{M}$   $^{13}\text{C}$ ,  $^{15}\text{N}$  uniformly labeled sample of human ubiquitin at 25 °C on a Varian INOVA 500 MHz spectrometer. The sample conditions consisted of 50 mM phosphate buffer pH 5.5, 50 mM NaCl and 0.04% azide. Recombinant ubiquitin was prepared as described [20]. NMR data were collected using a standard HNC0 [21] or a modified version for radial sampling, such that  $t_1 = t_1 \cos(\alpha)$  and  $t_2 = t_1 (sw_1/sw_2) \sin(\alpha)$  with the following experimental conditions. For radial sampled data each sampling angle, other than 0° and 90°, was collected with four quadrature components at 64 increments composing 256 FIDs. The 0° and 90° sampling angles were collected with two quadrature components at 64 increments composing 128 FIDs. Maximum  $t_1$  ( $^{13}\text{C}$ ) and  $t_2$  ( $^{15}\text{N}$ ) values were 29.14 and 31.64 ms, respectively. Cartesian sampled data were collected with equivalent resolution using four quadrature components at 64 increments in both indirect dimensions composing 16,384 FIDs. In both sampling schemes each FID contained 512 complex points and was the average of eight scans, the minimum number of phase cycling steps stated in the original reference. Using a 1.0-s inter-scan delay the measurement times for 0° and 90° sampling angles was 17 min. The measurement time for all other angles was 34 min. The measurement time for the Cartesian sampled spectrum was 36.4 h. The spectral width was set to 12 ppm in the proton dimension. The spectral widths for the indirect dimensions were chosen to assure no peaks were folded and set to 17.5 and 40 ppm in the carbon and nitrogen dimensions, respectively. With the corresponding carrier frequencies set at 176 and 119 ppm. The angle spectra were processed independently using a direct 2D Fourier transform. Prior to Fourier transforming the data was apodized and zero filled. A cosine squared apodization function was applied to remove truncation artifacts and to approximate the correction for unequal spaced data. Subsequently, the angle spectra were compared using the lower value (magnitude) algorithm to remove the ridge artifacts. The Cartesian sampled data were processed with corresponding techniques in one dimension. All

processing was done using an in-house program and visualized using Sparky [22].

## Acknowledgments

This work was supported by a grant from the Mathers foundation and NIH Grant GM 35940. J.M.G. is an NIH predoctoral trainee (GM 008275).

## References

- [1] M. Mobli, M.W. Maciejewski, M.R. Gryk, J.C. Hoch, An automated tool for maximum entropy reconstruction of biomolecular NMR spectra, *Nat. Meth.* 4 (2007) 467–468.
- [2] V. Jaravine, I. Ibraghimov, V.Y. Orekhov, Removal of a time barrier for high-resolution multidimensional NMR spectroscopy, *Nat. Meth.* 3 (2006) 605–607.
- [3] J.H. Chen, D. Nietispach, A.J. Shaka, V.A. Mandelshtam, Ultra-high resolution 3D NMR spectra from limited-size data sets, *J. Magn. Reson.* 169 (2004) 215–224.
- [4] S. Kim, T. Szyperki, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, *J. Am. Chem. Soc.* 125 (2003) 1385–1393.
- [5] R. Freeman, E. Kupce, New methods for fast multidimensional NMR, *J. Biomol. NMR* 27 (2003) 101–113.
- [6] B.E. Coggins, P. Zhou, Polar Fourier transforms of radially sampled NMR data, *J. Magn. Reson.* 182 (2006) 84–95.
- [7] K. Kazimierczuk, W. Kozminski, I. Zhukov, Two-dimensional Fourier transform of arbitrarily sampled NMR data sets, *J. Magn. Reson.* 179 (2006) 323–328.
- [8] D. Marion, Processing of ND NMR spectra sampled in polar coordinates: a simple Fourier transform instead of a reconstruction, *J. Biomol. NMR* 36 (2006) 45–54.
- [9] E. Kupce, R. Freeman, Fast multi-dimensional NMR by minimal sampling, *J. Magn. Reson.* 191 (2008) 164–168.
- [10] K. Kazimierczuk, A. Zawadzka, W. Kozminski, Optimization of random time domain sampling in multidimensional NMR, *J. Magn. Reson.* 192 (2008) 123–130.
- [11] K. Kazimierczuk, A. Zawadzka, W. Kozminski, I. Zhukov, Random sampling of evolution time space and Fourier transform processing, *J. Biomol. NMR* 36 (2006) 157–168.
- [12] K. Kazimierczuk, A. Zawadzka, W. Kozminski, I. Zhukov, Lineshapes and artifacts in multidimensional Fourier transform of arbitrary sampled NMR data sets, *J. Magn. Reson.* 188 (2007) 344–356.
- [13] N. Pannetier, K. Houben, L. Blanchard, D. Marion, Optimized 3D NMR sampling for resonance assignment of partially unfolded proteins, *J. Magn. Reson.* 186 (2007) 142–149.
- [14] H.R. Eghbalnia, A. Bahrami, M. Tonelli, K. Hallenga, J.L. Markley, High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection, *J. Am. Chem. Soc.* 127 (2005) 12528–12536.
- [15] E. Kupce, R. Freeman, Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy, *J. Am. Chem. Soc.* 126 (2004) 6429–6440.
- [16] J.W. Yoon, S. Godsil, E. Kupce, R. Freeman, Deterministic and statistical methods for reconstructing multidimensional NMR spectra, *Magn. Reson. Chem.* 44 (2006) 197–209.
- [17] T. Szyperki, H.S. Atreya, Principles and applications of GFT projection NMR spectroscopy, *Magn. Reson. Chem.* 44 Spec. No. (2006) S51–S60.
- [18] D.H. Eberly, 3D Game Engine Design: A Practical Approach to Real-Time Computer Graphics, Morgan Kaufmann, San Francisco, 2000.
- [19] J.M. Gledhill, A.J. Wand, Phasing arbitrarily sampled multidimensional NMR data, *J. Magn. Reson.* 187 (2007) 363–370.
- [20] A.J. Wand, J.L. Urbauer, R.P. McEvoy, R.J. Bieber, Internal dynamics of human ubiquitin revealed by  $^{13}\text{C}$ -relaxation studies of randomly fractionally labeled protein, *Biochemistry* 35 (1996) 6116–6125.
- [21] D.R. Muhandiram, L.E. Kay, Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity, *J. Magn. Reson. Ser. B* 103 (1994) 203–216.
- [22] T.D. Goddard, D.G. Kneller, SPARKY 3, University of California, San Francisco.