J Biomol NMR (2007) 39:165–175
DOI 10.1007/s10858-007-9187-1

ARTICLE

Covariance NMR in higher dimensions: application to 4D NOESY spectroscopy of proteins

David A. Snyder · Fengli Zhang · Rafael Brüschweiler

Received: 1 June 2007 / Accepted: 6 August 2007 / Published online: 18 September 2007

Abstract Elucidation of high-resolution protein structures by NMR spectroscopy requires a large number of distance constraints that are derived from nuclear Overhauser effects between protons (NOEs). Due to the high level of spectral overlap encountered in 2D NMR spectra of proteins, the measurement of high quality distance constraints requires higher dimensional NMR experiments. Although four-dimensional Fourier transform (FT) NMR experiments can provide the necessary kind of spectral information, the associated measurement times are often prohibitively long. Covariance NMR spectroscopy yields 2D spectra that exhibit along the indirect frequency dimension the same high resolution as along the direct dimension using minimal measurement time. The generalization of covariance NMR to 4D NMR spectroscopy presented here exploits the inherent symmetry of certain 4D NMR experiments and utilizes the trace metric between donor planes for the construction of a high-resolution spectral covariance matrix. The approach is demonstrated for a 4D ^{13}C-edited NOESY experiment of ubiquitin. The 4D covariance spectrum narrows the line-widths of peaks strongly broadened in the FT spectrum due to the necessarily short number of increments collected, and it resolves otherwise overlapped cross peaks allowing for an increase in the number of NOE assignments to be made from a given dataset. At the same time there is no significant decrease in the positive predictive value of observing a peak as compared to the corresponding 4D Fourier transform spectrum. These properties make the 4D covariance method a potentially valuable tool for the structure determination of larger proteins and for high-throughput applications in structural biology.

Keywords Covariance NMR · 4D NMR spectroscopy · NOESY · Cross peak assignment · Distance constraints

D. A. Snyder · F. Zhang · R. Brüschweiler (✉)
Department of Chemistry and Biochemistry, National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32306, USA
e-mail: bruschweiler@magnet.fsu.edu

Introduction

Nuclear magnetic resonance (NMR) spectroscopy provides a powerful approach for the elucidation of the three-dimensional structures of proteins in solution (Wüthrich 1986). It is largely based on the retrieval of inter-proton distances by the nuclear Overhauser effect using NOESY-type experiments. Protons within a protein produce NMR signals at specific frequencies (chemical shifts). A cross peak in a two-dimensional (2D) NOESY experiments (Kumar et al. 1980; Macura and Ernst 1980; Solomon 1955) at position (ω_1, ω_2) indicates that a proton with chemical shift ω_1 is nearby in space to a proton with chemical shift ω_2.

In proteins, however, it is unlikely that only one proton belongs to a given chemical shift preventing the unique attribution of 2D NOESY cross peaks. One way of resolving such ambiguities of attribution is to move to higher dimensional NMR spectra (Clore and Gronenborn 1991): a 2D NOESY cross-peak attributable to multiple pairs of protons can be resolved, for example, in a 3D NOESY experiment which separates the proton–proton peaks by the chemical shift of the attached carbon or nitrogen spin. Similarly, 4D NOESY separates proton–proton peaks by the chemical shifts of hetero-atoms attached to each proton.

However, recording higher dimensional NMR experiments naturally involves trade-offs: maintaining the same resolution in each dimension requires the recording of exponentially more transients when the dimensionality is increased. In practice, keeping the total measurement time reasonable, which is especially a concern in the context of high-throughput structure determination (Yee et al. 2006), requires a reduction in the measured number of increments resulting in a loss of resolution in the corresponding FT spectra along the indirect dimensions.

A way to circumvent the trade-off between higher dimensionality and reduced resolution is the application of resolution enhancement techniques during processing. Some resolution enhancement techniques such as linear prediction (Koehl 1999; Led and Gesmar 1991) extrapolate (predict) from the truncated experimental time series additional signal that, following Fourier transform, helps to recover some of the lost resolution. Maximum entropy reconstruction (Hoch and Stern 2001; Rovnyak et al. 2004) operates on non-linearly sampled datasets allowing resolution optimization for a given measurement time. On the downside, it is not straightforward to subject the experimental data to standard Fourier transform for comparison. More recently, modern Fourier transform variants have been introduced that can be applied to non-linearly or non-Cartesian sampled data (Coggins and Zhou 2007; Marion 2006; Misiak and Kozminski 2007). Methods such as the filter diagonalization method (FDM) avoid the use Fourier transform altogether, also allowing for the use of optimized non-linear sampling schemes (Hu et al. 2000).

A potential weakness of these methods is that the underlying model or criterion might be of limited appropriateness for the resolution enhancement problem at hand. For example, linear prediction is often only valid for a certain amount of extrapolation, typically no more than by a factor of two, and maximum entropy does not necessarily produce a spectrum that is identical to the idealized (i.e. high-resolution) FT spectrum.

Reduced dimensionality (RD) (Brutscher et al. 1995; Szyperski et al. 1993), G-Matrix Fourier transform (GFT) (Kim and Szyperski 2003; Szyperski and Atreya 2006), projection reconstruction (PR) (Freeman and Kupce 2006; Kupce and Freeman 2006), APSY (Hiller et al. 2005) and high-resolution iterative frequency identification (HIFI) (Eghbalnia et al. 2005) techniques avoid many of these assumptions by focusing on sampling schemes that provide high-resolution spectra in suitable lower dimensions (projections) that, after FT, provide peak positions of interest (RD and GFT) or allow the reconstruction of higher dimensional spectra (PR, APSY, HIFI).

Another approach to resolution enhancement is to apply the techniques of matrix algebra to the analysis of NMR spectra (Havel et al. 1994). Multi-dimensional decomposition techniques (Jaravine et al. 2006; Malmodin and Billeter 2005; Orekhov et al. 2003; Tugarinov et al. 2005) achieve resolution enhancement by treating NMR spectra as matrices/tensors, using algebraic operations previously defined for such mathematical objects, to build implicit or explicit models from which to calculate high resolution frequency domain spectra from sparsely and potentially non-linearly sampled mixed time/frequency domain data.

A method for resolution enhancement based on the properties of a matrix representation of NMR data, but without requiring that peak shapes follow any particular model, is covariance NMR (Brüschweiler 2004; Brüschweiler and Zhang 2004; Trbovic et al. 2004). Because a covariance is a general measure of correlation between a pair of objects, such as a pair of resonances, covariances naturally lend themselves to the representation of spin correlations contained in 2D NMR data. Of practical importance, covariance NMR endows the indirect dimension with the same high resolution as the direct dimension (Brüschweiler and Zhang 2004), which allows significant savings in the number of increments collected along the indirect dimension (Chen et al. 2006; Trbovic et al. 2004). Covariance NMR represents a 2D spectrum with N_1 points along the indirect dimension and N_2 points along the direct dimension as an $N_1 \times N_2$ matrix, $\mathbf{S}$. As a consequence of Parseval's theorem, the $N_2 \times N_2$ matrix resulting from the covariance transformation of $\mathbf{S}$, $\mathbf{C} = (\mathbf{S}^{\dagger}\mathbf{S})^{1/2}$ where the † symbol denotes the conjugate transpose, is independent of whether the indirect dimension is in the frequency domain (Brüschweiler 2004) or the time domain (after an affine adjustment to each column to ensure that all column averages of $\mathbf{S}$ are zero).

Frequency domain covariance need only consider the real absorptive parts of $\mathbf{S}$ in which case the covariance transform is $(\mathbf{S}^{\mathbf{T}}\mathbf{S})^{1/2}$, where the T symbol denotes the matrix transpose. If $\mathbf{S}$ is symmetric and positive-semidefinite (all eigenvalues ≥ 0), then $(\mathbf{S}^{\mathbf{T}}\mathbf{S})^{1/2} = \mathbf{S}$. NOESY spectra usually possess large diagonal signals that ensure $\mathbf{S}$ is positive semidefinite (Brüschweiler 2004), although regularization techniques obviate this requirement (Chen et al. 2007). In the case where $\mathbf{S}$ is asymmetric due (only) to fewer points being sampled in the indirect rather than the direct dimension, the result of the covariance transform not only has equal resolution in the indirect and direct dimensions but also provides an estimator of the spectrum sampled so as to be symmetric: with the same number points in the direct dimension as in the indirect dimension.

The matrix square root operation is of particular importance in the covariance transform of NOESY spectra. Let a signal associated with element S_{ij} of $\mathbf{S}$ indicate that a proton having a chemical shift that is the jth point along the direct dimension experiences an NOE due to a proton having a chemical shift corresponding to the ith point along

the indirect dimension. The standard definition of a covariance matrix (Johnson and Wichern 2002), $\mathbf{C}^2 = \mathbf{S}^\mathbf{T}\mathbf{S}$, results in a non-zero signal for element ${}^2S_{ij}$ of $\mathbf{C}^2$ so long as column i and column j of $\mathbf{S}$ have non-zero values at the same positions, which would occur if the protons, H_i and H_j, whose chemical shifts fall at positions i and j in the direct dimension, experience NOEs from one or more protons in common even if the NOE between H_i and H_j is negligible. This relates to the well-known NOESY relay effect: in the case where resonance overlap is minimal, $\mathbf{C}^2$ is in fact the spectrum obtained by doubling the mixing time used to obtain $\mathbf{S}$ (Brüschweiler 2004).

However, resonance overlap between distinct protons allows column i and column j of $\mathbf{S}$ to have non-zero values at the same position due to H_i and H_j each experiencing an NOE from a distinct proton having a common chemical shift. Such non-physical "pseudo-relays" mean that $\mathbf{C}^2$ cannot in general be interpreted as a NOESY spectrum (i.e. the one that would have been obtained instead of $\mathbf{S}$ had the mixing time been doubled), however, the properties of the matrix square root ensure, since as described above in the symmetric case, $\mathbf{C} = \mathbf{S}$, that even non-physical pseudo-relays are suppressed by the square root operation.

This work presents a generalization of the covariance method to specific four-dimensional (4D) spectra that, under ideal sampling conditions, exhibit an inherent symmetry between donor and acceptor resonances as is the case for certain 4D NOESY and 4D TOCSY experiments. The covariance transform is applied to a ^{13}C-edited 4D NOESY spectrum measured on a ubiquitin sample. Peak picking heuristics for 4D covariance transformed data, presented in this paper, provide peak lists, independent of knowledge of ubiquitin's structure that, when evaluated based on ubiquitin's known three-dimensional (3D) structure, demonstrate that covariance NMR can resolve peaks yielding additional restraints not readily extractable from a 4D FT NOESY spectrum obtained from the same raw data. Thus, as for 2D spectra, covariance NMR is found to be a useful estimator of a spectrum obtained with resolution in the "donor" dimension(s) equal to that of the acceptor dimension(s), which includes the high resolution direct proton dimension.

Theory of covariance NMR of 4D NOESY spectra

Key to 2D covariance NMR is the ability to treat an NMR spectrum as a collection of vectors. A 2D experiment associates with each point in the direct dimension a "donor vector" of points along the indirect dimension. The matrix product in covariance NMR, as described in the introduction, is a notationally compact method of tabulating the inner-products of all pairs of such vectors into a matrix, the matrix square root of which yields an estimate for a symmetric spectrum.

Just as a 2D spectrum is a line of lines, a 4D spectrum can be viewed as a plane of planes. A 4D NMR spectrum associates for each sampled quadruple of chemical shift values (ω_1, ω_2, ω_3, ω_4), a (real valued) signal $S(\omega_1, \omega_2, \omega_3, \omega_4)$. Taking a ^{13}C-edited 4D NOESY as an exemplar of a 4D NMR spectrum, fixing the first two chemical shift values to a specific CH "donor pair" (ω_1, ω_2), where in the present case ω_1 corresponds to ^{13}C and ω_2 to ^{1}H, and letting the last two values vary yields a CH "acceptor plane"

$$\boldsymbol{A}_{\omega_1,\omega_2} = \boldsymbol{S}(\omega_1, \omega_2, *, *). \quad (1)$$

Similarly, fixing the last two values to a specific "acceptor pair" (ω_3, ω_4), where in the present case ω_3 corresponds to ^{13}C and ω_4 to ^{1}H, and letting the first two values vary yields a "donor plane"

$$\boldsymbol{D}_{\omega_3,\omega_4} = \boldsymbol{S}(*, *, \omega_3, \omega_4). \quad (2)$$

In Fig. 1, panels a and b schematically illustrate the representation of a 4D spectrum, respectively, as a plane of acceptor planes and a plane of donor planes. In the case of a ^{13}C-edited 4D NOESY, each donor pair and each acceptor pair correspond to points in a ^{13}C-^{1}H-HSQC spectrum. If (ω_1, ω_2) (respectively, (ω_3, ω_4)) corresponds to the chemical shifts of a covalently bonded carbon–hydrogen pair, H_d–C_d (respectively, H_a–C_a), $\boldsymbol{A}_{\omega_1,\omega_2}$ (respectively, $\boldsymbol{D}_{\omega_3,\omega_4}$) contains the subset of the signals in an ^{13}C-^{1}H-HSQC spectrum for which the proton associated with a signal in that subset experiences an NOE from the proton H_d (respectively, H_a). The spectrum is ideally symmetric: $\boldsymbol{A}_{\omega_1,\omega_2} = \boldsymbol{D}_{\omega_1,\omega_2}$, ignoring the higher resolution possible for the acceptor planes (which contain the direct dimension) in practice.

In order to save computational time and to reduce computer memory requirements, it is useful to restrict analysis of a 4D spectrum to signal containing regions. Replacing each acceptor plane in the plane of planes shown in Fig. 1a with its maximal element yields a plane of points, the donor (skyline) projection. Similarly replacing donor planes with their maximal elements yields the acceptor (skyline) projection shown in Fig. 1c. An "acceptor footprint" consisting of those acceptor pairs associated with elements in the acceptor projection whose value is greater than some threshold value (and which lie outside of the water line) provides a reduced set of acceptor pairs to consider for subsequent analysis. Similarly, a "donor footprint", obtained by mapping the acceptor footprint onto the lower resolution donor projection and filling in any remaining gaps (for example, the water line excluded from the acceptor footprint which is not as noisy

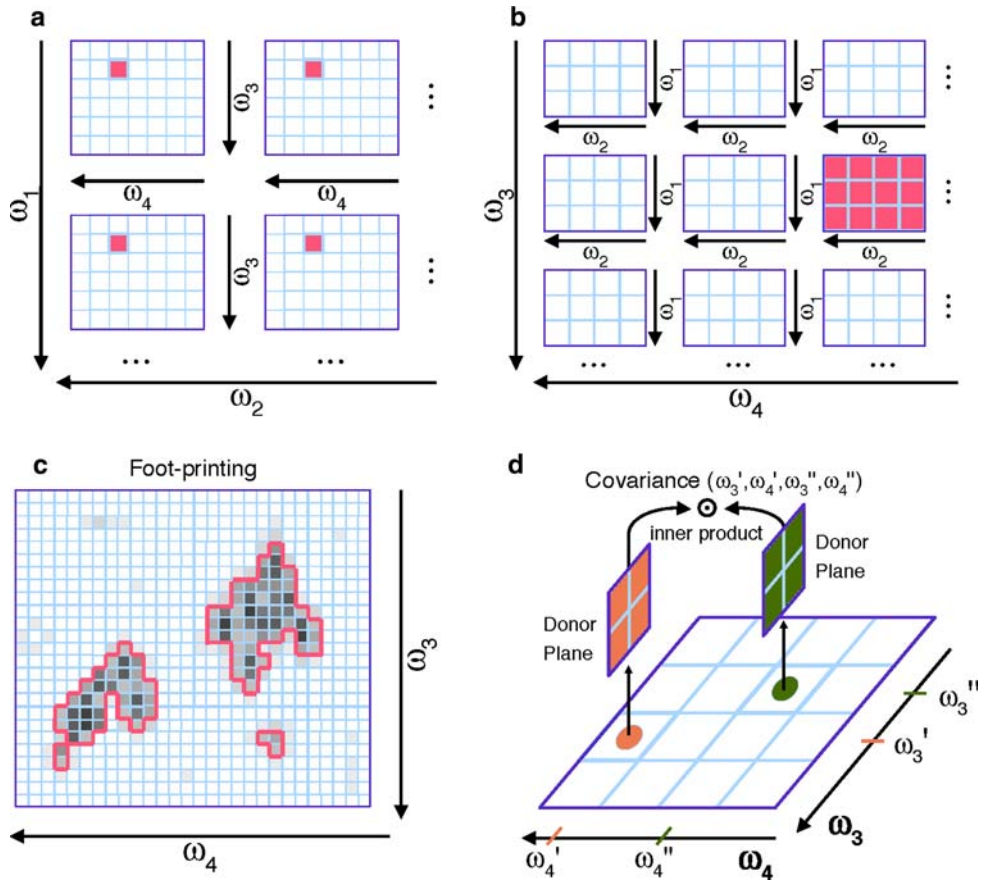

Fig. 1 Representations of a 4D NOESY NMR spectrum. (**a**) Associated with each donor carbon/proton (ω_1, ω_2) pair is an acceptor plane: a signal at a particular location (ω_3, ω_4) within the acceptor plane indicates an NOE between a donor proton with chemical shift ω_2, attached to a donor carbon with chemical shift ω_1, and an acceptor proton with chemical shift ω_4 attached to an acceptor carbon with chemical shift ω_3. (**b**) Transposing the view shown in panel **a** leads to a representation of the 4D NOESY spectrum as a plane of donor planes, each associated with an acceptor carbon/proton pair. As a guide through the transposition between panel **a** and **b**, a fixed (ω_3, ω_4) location in each acceptor plane shown in panel **a** is painted red; these points comprise a particular donor plane which is painted red in panel **b**. (**c**) Replacing each of the donor planes shown in panel **b** with its maximal element yields an acceptor projection as the one sketched in this panel. The set of points in this plane associated with an intensity greater than a critical value constitutes the "acceptor footprint". Covariance NMR, when using acceptor foot-printing, ignores donor planes mapping to points in the acceptor projection outside of the acceptor footprint. (**d**) Diagram of the 4D covariance NMR procedure: the covariance signal associated with the point $(\omega_3', \omega_4', \omega_3'', \omega_4'')$ is the inner product of the two donor planes associated with the acceptor pairs (ω_3', ω_4') and (ω_3'', ω_4'')

in the donor dimensions), provides a useful reduced set of donor pairs to consider.

The 4D spectrum can be represented as a collection of donor planes (see Fig. 1b), which is analogous to the representation of a 2D spectrum as a collection of donor vectors. The 2D covariance transform calculates the squared covariance spectrum, $\mathbf{C}^2$, as a matrix tabulating the inner product between all pairs of donor vectors; the natural 4D extension of this procedure calculates the inner products between donor planes (Fig. 1d), each of which is finitely sampled and hence a matrix. The 'inner product' between the donor plane $\boldsymbol{D}_{\omega_3',\omega_4'} = \boldsymbol{S}(*,*,\omega_3',\omega_4')$ and $\boldsymbol{D}_{\omega_3'',\omega_4''} = \boldsymbol{S}(*,*,\omega_3'',\omega_4'')$ by the trace metric

$$^2\mathrm{C}_{\mathrm{ij}} = \mathrm{Tr}(\boldsymbol{D}_{\omega_3',\omega_4'}{}^T \boldsymbol{D}_{\omega_3'',\omega_4''}) \quad (3)$$

(since the matrix trace (Tr) of Eq. 3 yields the sum of the element by element product of two donor planes).

Equation 3 associates a unique index value to each possible acceptor pair, e.g. index value i with (ω_3', ω_4') and index value j with (ω_3'', ω_4''). The possible acceptor pairs might include all (ω_3, ω_4) pairs sampled in the frequency domain or be restricted to those in the acceptor footprint (the process of acceptor foot-printing). Tabulating the $^2\mathrm{C}_{\mathrm{kl}}$ obtained by allowing k and l to vary over all possible index values yields a matrix $\mathbf{C}^2$, whose matrix square root, $\mathbf{C}$, is the 2D representation of the 4D covariance spectrum.

Back-conversion of this matrix representation to a 4D array is straightforward:

$$C(\omega_3', \omega_4', \omega_3'', \omega_4'') = C_{ij} \quad (4)$$

where i is the index value associated with the acceptor pair (ω_3', ω_4') and j is the index value associated with the acceptor pair (ω_3'', ω_4''). Thus, as in the 2D case, the covariance transform endows both the proton and carbon donor dimensions with the same resolution as the acceptor dimensions.

Materials and methods

Ubiquitin sample and NMR spectroscopy

NMR experiments were performed on a uniformly ^{13}C,^{15}N doubly labeled human ubiquitin sample. Experiments were measured in aqueous (H_2O) solution at pH 5.5 in 95% H_2O and 5% D_2O. All NMR experiments were performed at 298 K using a Bruker 800 MHz spectrometer equipped with a cryoprobe. The 4D NOESY (Clore et al. 1991; Kay et al. 1990a; Morshauser and Zuiderweg 1999; Vuister et al. 1993) was performed with 8, 32 and 16 complex points in the donor carbon, donor hydrogen and acceptor carbon dimensions, respectively. Acquisition of these indirect dimensions used the States-TPPI acquisition mode and gradient-assisted sensitivity enhancement (Kay et al. 1992; Rance 1994) was applied to the acceptor carbon dimension. The direct, i.e. acceptor proton, dimension was recorded, using quadrature detection of 955 complex points. Each point represents the average of eight appropriately phase cycled transients, and the spectrum took 130 h to record. Data processing, including the linear prediction of 32 additional points in the donor proton dimension together with exponential to Gaussian apodization in that dimension, shifted sine-bell apodization in the other dimensions, zero filling, phase adjustment and baseline correction, were performed using NMRPipe (Delaglio et al. 1995).

The ^{13}C-edited 2D ^{1}H-^{1}H NOESY was obtained using the same pulse sequence as for the 4D NOESY, but setting the number of points obtained along the carbon dimensions to unity. The number of complex points along the indirect proton dimension is 256. Additionally, ^{13}C-^{1}H HSQC and ^{15}N-^{1}H HSQC spectra, a three-dimensional HNCA spectrum (Farmer et al. 1992; Grzesiek and Bax 1992; Kay et al. 1990b) and three C-C TOCSY spectra, an H(C)CH (Bax et al. 1990), an HcccoNH and an (H)CccoNH spectrum (Celda and Montelione 1993) provided the information required to select from among the many chemical shift assignments available via the BMRB (Seavey et al. 1991) for ubiquitin and to refine those assignments so that they matched the ubiquitin sample under study.

Covariance NMR

4D covariance NMR procedure described in the Theory section was implemented by means of a Matlab program, which is available from the authors upon request. The threshold value for acceptance into the acceptor footprint was set to 20 times the median absolute value of the acceptor projection. The acceptor footprint also excluded the water line identified as those acceptor pairs with proton chemical shifts between 4.65 and 4.85 ppm. This reduced the number of all possible acceptor pairs from 15,280 (in the absence of zero-filling in the acceptor dimensions) to 3,138 pairs (with zero filling). Donor planes corresponding to acceptor pairs outside of the acceptor footprint were excluded from covariance processing and elements within the donor planes outside of the donor footprint were set to zero. Following covariance analysis, the program "unpacks" the 2D matrix representation of the 4D spectrum, filling in zeros in locations corresponding to those excluded from the acceptor footprint, into an NMRPipe format file, which can be viewed in NMRDraw or converted to other formats for further analysis.

This program was applied to a frequency domain 4D NOESY spectrum that was processed as follows: The apodization in the donor proton dimension used for visualizing the 4D FT NOESY used sufficient line broadening to prevent sinc-wiggles due to the relatively small number of points collected and further linear predicted in t_2. However, in order to ensure a more symmetric apodization scheme and to prevent the suppression of higher resolution information contained in the later t_2 increments, the Gaussian line broadening applied to the donor proton dimension in the spectrum further processed by the covariance transform was set to just above the typical natural line-width for proton resonances in this data set (as judged from the line-width of protons in the direct dimension, processed without apodization), as is the customary procedure for apodization with linear prediction.

Display of spectra and NOESY peak picking

Spectra were displayed in Sparky (Goddard & Kneller), which was used to automatically peak pick 2D spectra. The 4D FT NOESY was manually peak picked using the ^{13}C-^{1}H HSQC spectrum and the theoretical symmetry of the 4D NOESY experiment as guides to peak picking.

This manually generated peak list of peaks outside of the water line was filtered (using a tolerance of 0.2 ppm, i.e. equal to the resolution of the donor proton dimension following linear prediction, for the donor proton dimension and a tolerance of 0.05 ppm for the direct dimension) against the peaks found in the (covariance) 2D NOESY spectrum, culled to remove diagonal and near-diagonal peaks and then sorted according to peak intensity. The covariance 2D spectrum is used here because it is simpler and more robust to filter against a higher resolution symmetric spectrum than a lower resolution asymmetric spectrum.

To peak pick the 4D covariance spectrum, the 4D FT NOESY was overlaid onto the covariance spectrum. Associated with each peak P, centered at (ω_1, ω_2, ω_3, ω_4) and manually picked from the FT spectrum, is a collection of local maxima in the covariance spectrum such that each local maximum L in this collection approximately has the same acceptor chemical shifts as P and lies, in viewing the donor plane $\boldsymbol{D}_{\omega_3,\omega_4}$, within the contour lines associated with P. The heuristic for inclusion of the local maximum L, centered at $(\omega'_1, \omega'_2, \omega'_3, \omega'_4)$, in the list of covariance peaks is that, within the acceptor planes $\boldsymbol{A}_{\omega'_3,\omega'_4}$ and $\boldsymbol{A}_{\omega'_1,\omega'_2}$, the shape and position of L—and its symmetry peak L', centered at $(\omega'_3, \omega'_4, \omega'_1, \omega'_2)$—approximately coincide with the contours in the FT spectrum in $\boldsymbol{A}_{\omega'_3,\omega'_4}$ in the neighborhood of (ω'_1, ω'_2) and in $\boldsymbol{A}_{\omega'_1,\omega'_2}$ around (ω'_3, ω'_4).

The motivation for this heuristic is that covariance peaks should reflect the line-shape of the peak in the FT spectrum from which the covariance transform resolves them because the covariance transform only enhances the resolution in the donor dimensions and hence should leave the cross section of a peak in the acceptor plane largely unchanged both with respect to resolution and line-shape. The peaks picked in the covariance spectrum were further filtered (using a tolerance of 0.05 ppm in each proton dimension) against the peaks found in the (covariance) 2D NOESY spectrum.

Comparison of NOESY results with the known ubiquitin structure

The AutoQF module (Huang et al. 2005) of the automatic NOESY assignment software package, AutoStructure (Huang et al. 2006), evaluated the peak lists extracted from NOESY spectra, using automatic peak picking for 2D spectra but with manual intervention, as described above, required for 4D spectra. AutoStructure's AutoQF module, in addition to computing other statistics, internally filters the peak lists against the given assignments to count the number of "assignable" peaks and calculates the "recall" or "positive predictive value" of picking a peak, which is the percentage of picked peaks that are consistent with a given three-dimensional structure.[1]

Additionally, AutoStructure, via its module for using a known structure to bootstrap an NMR-based structure calculation, allows for the assignment of peaks to specific NOE observable pairs of protons based on a given structure. Re-running AutoStructure with assignments adjusted using the chemical shifts of the assigned NOESY cross-peaks provides additional assignments and renders some peaks consistent that were considered inconsistent in the first round of AutoStructure analysis.

The structure used for comparison purposes was the ubiquitin structure, PDB (http://www.pdb.org/) (Berman et al. 2000) entry 1G6J (Babu et al. 2001), and matching to assignments was performed using tolerances equal to one-half the digital resolution (assuming zero-filling to twice the data size) in each indirect dimension of the 4D NOESY and 0.035 ppm in the direct dimension of all NOESY spectra (as well as for the indirect dimension for the 2D NOESY analyses). Additional statistical tests were performed using the Simple Interactive Statistical Analysis (SISA) website (Uitenbroek 1997).

Results

Covariance 2D NOESY analysis

In order to interpret 4D NMR data, it is useful to start from lower dimensional, higher resolution data. Previous work has established that the covariance NMR result obtained from a 2D ^{1}H-^{1}H NOESY spectrum of ubiquitin with 256 complex points in the indirect dimension accurately reproduces the data quality and resolution obtainable in a 2D FT NOESY spectrum with 1,024 (real, TPPI) t_1 increments (Brüschweiler and Zhang 2004).

AutoStructure (Huang et al. 2005, 2006) analysis of automatically generated peak lists from both an FT spectrum with 512 (complex) t_1 increments and a covariance spectrum calculated using 256 (complex) t_1 increments demonstrated the applicability of covariance NMR to ^{13}C-edited 2D NOESY spectra. AutoStructure assigned 1,389 peaks in the FT spectrum and found 71% of those peaks picked which matched the available proton chemical shift assignments to be consistent with the structure. The

[1] AutoStructure, a program designed to determine and validate protein structures, views the protein structure as querying a peak list, the peaks being "relevant documents" retrieved by matching the structure. Thus, AutoStructure labels the percentage of peaks that are true as the "recall". In evaluating spectra, e.g. comparing Covariance and Fourier Transform spectra, the peak list is the query, therefore AutoStructure's "recall" statistic measures the positive predictive value in the context of peak list or spectrum validation.

covariance NMR spectrum, though obtained by truncating the available data to 256 complex points, with its higher nominal resolution (i.e. having the same resolution in the indirect as in the direct dimension), resulted in having 1,496 peaks assigned by AutoStructure based on the Ubiquitin structure as well as a AutoQF "recall" score of 74%.

4D covariance NOESY analysis

Since it is impossible in practice to obtain a 4D FT spectrum at the same resolution provided by covariance NMR, validation of covariance NMR proceeds by comparing peaks found in the covariance NMR spectrum to those found in the FT spectrum. A successful application of covariance NMR to 4D data resolves additional peaks not distinguishable in the FT spectrum without resulting in a significant increase in the rate of artifacts found in the covariance spectrum relative to the rate found in the FT spectrum. AutoStructure provides validation tools comparing NOESY peak lists to 3D protein structures (Huang et al. 2005, 2006) and is thus used here to analyze and compare covariance and FT spectra.

Manual peak picking of the FT spectrum resulted in 415 peaks. Associated, using the heuristics described in the "Materials and Methods" section, with 302 of those peaks were 416 peaks in the covariance spectrum. The remaining 113 peaks in the FT spectrum had no associated peaks in the covariance spectrum, acceptable according to the peak selection heuristics used.

Table 1 lists the AutoStructure provided evaluation statistics for the 386 assignable peaks from the 4D covariance spectrum and the 279 assignable peaks from the 302 4D FT spectrum peaks corresponding to peaks in the covariance spectrum. The tabulation provides results based on the initial chemical shift assignments as verified for the ubiquitin sample used in this study as well as results incorporating the additional chemical shift information provided via assigning NOESY cross-peaks. These results demonstrate that, by increasing the nominal resolution of the spectrum, the covariance transform yields peak lists providing additional long range constraints without a significant increase in the false positive rate (Likelihood ratio χ^2 p-value > 0.10).

Figure 2 illustrates the resolution enhancement provided by covariance NMR. Figure 2a,c shows orthogonal views of the 4D FT NOESY spectrum of ubiquitin, intersecting at the V5CG1/HG1 donor and I13CA/HA acceptor pairs. The FT donor plane (Fig. 2c) cannot resolve whether the signal present between V5CG1/HG1 and I13CA/HA arises from an NOE between the donor and acceptor protons or whether the intensity at this region evidences an NOE from another donor pair with chemical shifts effectively degenerate with the V5CG1/HG1 chemical shifts at the donor plane resolution. Covariance NMR resolves the broad linewidth peak of panel c (full-width at half-height is 2511.4 Hz in ω_1 and 168.2 Hz in ω_2) into multiple narrower peaks (full-width at half-height of each peak is approximately cut in half, corresponding to the doubling of resolution, in ω_1 and by over three-fourths in ω_2) including peaks from not only the (1) V5HG1/I13HA NOE but also from (2) I13HG2/I13HA. Also distinguishable are the (3) I23HG2/K27HA as well as the (4) V26HG1/K27HA and L8HD2/HA NOEs, which are distinct local maxima associated with acceptor pairs with similar chemical shifts to the I13CA/HA chemical shift pairs.

Table 1 Performance comparison of 4D covariance and 4D FT NOESY spectroscopy

		Initial assignments[a]		Revised assignments[b]	
		FT	Covariance	FT	Covariance
Peaks	Assignable[c]	279	386	279	386
	Inconsistent[d]	45	86	45	80
	Recall (%)	83.9	77.7	83.9	79.3
Assignments	Total[e]	206	284	211	286
	Intra-residue	141	200	143	200
	$0 < \lvert i - j \rvert < 5$	28	34	30	34
	$\lvert i - j \rvert \geq 5$	24	40	24	42
	$\langle \lvert i - j \rvert \rangle$	3.32	3.19	3.27	3.33

[a] Evaluation based on assignments as validated/adjusted using HSQC, HNCA and TOCSY spectra

[b] Including in the count of assignments those made using chemical shifts as adjusted according to the "Initial Assignments" of NOESY cross-peaks. These columns also exclude from the counts of inconsistent peaks those rendered consistent following this adjustment

[c] Corresponding to both chemical shift assignments and a peak in the 2D NOESY

[d] Not consistent with any pair of protons (effectively) < 5 Å apart in structure

[e] Total number of peaks assigned (including ambiguous assignments) by AutoStructure

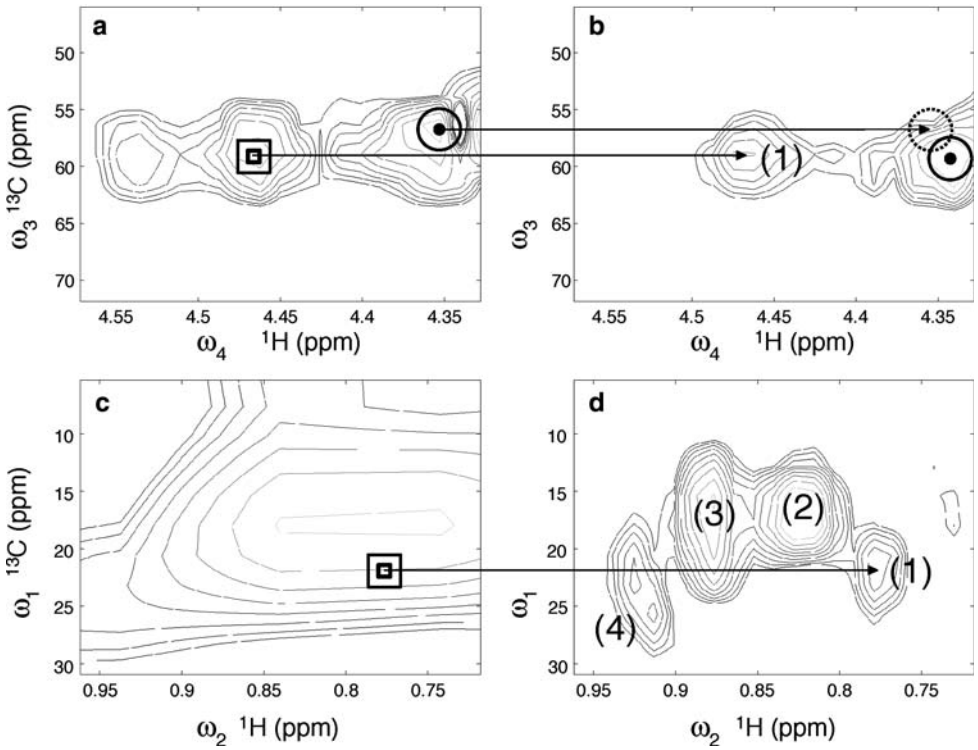

Fig. 2 Spectral resolution enhancement via covariance NMR. (**a**, **c**) Portions of mutually orthogonal FT acceptor (**a**) and donor (**c**) planes of a 4D NOESY spectrum. The square marks the V5CG1/HG1 donor (ω_1 = 21.191 ppm, ω_2 = 0.780 ppm), I13CA/HA acceptor (ω_3 = 58.908 ppm, ω_4 = 4.473 ppm) quadruple at which these planes intersect. Acceptor (**b**) and donor (**d**) planes after the covariance transformation: covariance NMR maintains the acceptor plane resolution with which it also endows the donor plane, extracting multiple distinct peaks from a single peak in the lower resolution FT spectrum. Cross peak (1) reflects the V5HG1/I13HA NOE and peak (2) reflects the I13HG2/HA NOE. Peaks (3) and (4) both are centered in nearby donor planes: (3) corresponds to the I23HG2/K27HA NOE and (4) comprises two distinct peaks, V26HG1/K27HA and L8HD2/HA. Additionally, covariance NMR suppresses the leftmost signal of panel **a** and displaces the rightmost FT signal of panel **a** (indicated by dashed and solid circles), neither of which corresponds to NOEs consistent with ubiquitin's structure

Figure 3 further illustrates how 4D covariance NMR identifies constraints not distinguishable in 4D FT spectra. Figure 3a,d shows orthogonal views of the 4D covariance NOESY spectrum, intersecting at a peak (1) attributable to the T55HG2/S20HA NOE. However, this NOE is not distinguishable in the FT spectrum (Fig. 3b, c, e, f): while the acceptor planes corresponding to donor pairs T55CG2/HG2 (Fig. 3b) and S20CA/HA (Fig. 3f) show isolated signals, respectively, corresponding to the chemical shifts of S20CA/HA (1) and T55CG2/HG2 (1′), the low resolution of the orthogonal donor planes for the S20CA/HA (Fig. 3e) and T55CG2/HG2 (Fig. 3c) acceptor pairs preclude the identification of the T55HG2/S20HA NOE.

Since the primary indication of covariance NMR is to provide resolution enhancement and thus disambiguate highly overlapped features in an FT spectra into multiple resolved peaks, the above analysis only considers for comparison those peaks picked in the FT spectrum for which associated covariance peaks can be picked. Interestingly, in comparison to the 83.9% positive predictive rate of those FT peaks corresponding to the heuristically picked covariance peaks, out of the 106 assignable peaks not associated with heuristically picked covariance peaks, only 56.6% are consistent with ubiquitin's structure. This difference in the positive predictive rate indicates a highly statistically significant ($p < 0.001$ by the Likelihood ratio χ^2 test) association between an FT NOESY peak arising from an actual NOE interaction and such a peak yielding at least one reasonable covariance peak. This association likely arises from the symmetry and line-shape requirements implicitly imposed on FT peaks for the corresponding covariance peaks to be picked according to the method described above.

Discussion

An ideal evaluation of whether covariance NMR, applied to a data set with low donor resolution, provides an appropriate estimator of an FT spectrum obtained with donor resolution equal to the acceptor resolution, compares the covariance result obtained from a lower resolution data

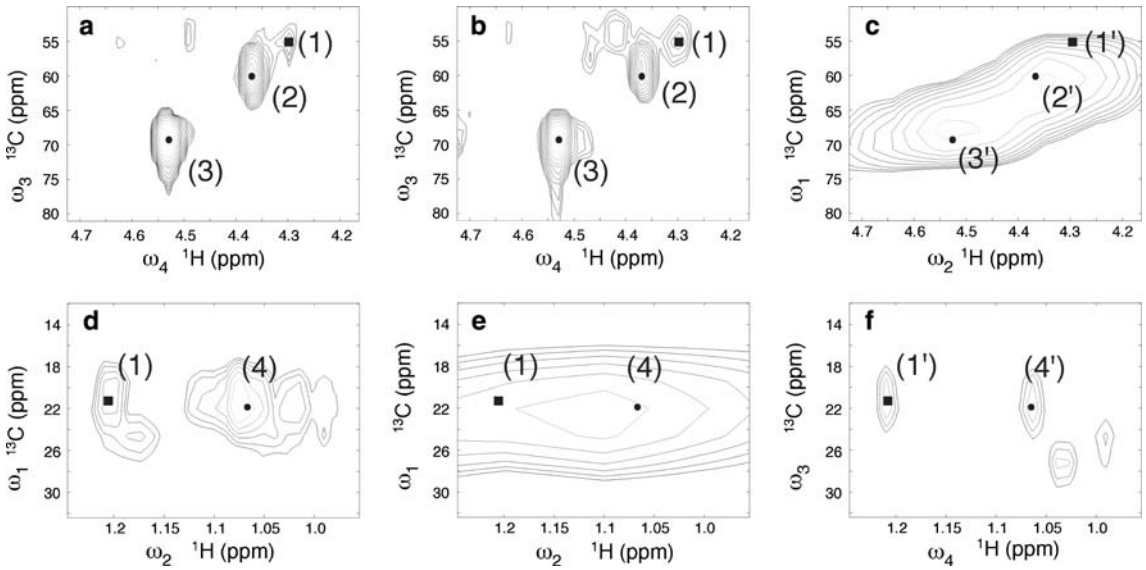

Fig. 3 Peak overlap resolution by covariance NMR. The symbol ■ denotes the point of intersection of orthogonal planes (**a**, **d**), (**b**, **e**) and (**c**, **f**). (**a**) Portion of the covariance acceptor plane associated with the three threonine donor pairs, T9, T22 and T55 CG2/HG2 (ω_1 = 22.044 ppm, ω_2 = 1.207 ppm), showing peaks arising from the (1) T55HG2/S20HA, (2) T9HG2/HA and (3) T9HG2/HB NOEs. (**d**) Portion of the covariance donor plane associated with the S20CA/HA acceptor (ω_3 = 56.565 ppm, ω_4 = 4.295 ppm) indicating that the 4D covariance spectrum indeed resolves peak (1) as a local maximum. Also visible in this region of the donor plane is peak (4) belonging to the T22HG2/S20HA NOE. (**b**, **e**) acceptor (**b**) and donor (**e**) FT planes, which after covariance transformation yield the planes of panels **a** and **d**, respectively. (**c**, **f**) donor (**c**) and acceptor (**f**) FT planes symmetric to the planes shown in panels **b** and **e**, respectively—i.e. respectively associated with the T9CG2/HG2,T22CG2/HG2,T55CG2/HG2 *acceptor* pairs and the S20CA/HA *donor* pair. Prior to covariance transformation, neither feature (1) nor its symmetry feature (1′) mark peak maxima in the 4D spectrum. The higher resolution with which covariance NMR endows donor planes disambiguates (1) as a distinct signal arising from the T55HG2/S20HA NOE rather than merely being associated with the shoulders of broad unresolved peaks associated with other NOEs

set with the high resolution "target" FT spectrum. For 4D datasets such comparisons are simply not feasible. This study evaluates 4D covariance results using a more pragmatic approach: covariance NMR provides an acceptable estimator of a higher resolution spectrum if the information obtained from the covariance spectrum is the information expected from a high resolution spectrum. In particular, does covariance NMR disentangle peaks that are highly overlapped due to the necessarily low donor resolution of 4D FT spectra without significantly decreasing the positive predictive value for accurate NOE identification from any given peak?

The results presented in Table 1 show that this is the case: the peak list obtained from the covariance spectrum has more peaks, yielding more assigned NOEs and more assigned long-range constraints, than do the corresponding peaks in the FT spectrum. Yet, the false positive rates for the two peak lists do not differ significantly. Furthermore, the analysis presented in this table uses tolerances commensurate with the resolution of the spectrum under consideration: covariance NMR allows for the assignment of more NOESY cross peaks at *tighter* (donor) tolerances than does the FT spectrum. The ability to use tighter tolerances and hence avoid ambiguities is particularly useful in iterative structure calculation where one must distinguish between multiple possible assignments in order to determine which constraint to extract from a given peak. Having a shorter list of possible assignments makes structure calculation more robust.

Figures 2 and 3 demonstrate how the increased donor plane resolution provided by covariance NMR leads to the identification of additional long-range constraints from 4D NOESY data. Figure 3 in particular shows that even taking advantage of the symmetry of the NOESY spectrum—mapping high resolution acceptor planes onto low resolution donor planes, which potentially serves as a means of disambiguating the latter—fails to disambiguate the T55HG2/S20HA NOE because the donor plane in a mutually orthogonal pair of intersecting planes in a 4D FT NOESY always will have lower resolution. By endowing the donor planes with the same resolution as the acceptor planes, 4D covariance NMR allows for the identification of NOEs indistinguishable using a 4D FT NOESY obtained at an experimentally feasible resolution.

Comparison between the FT and covariance spectra also significantly improves the positive predictive value of peaks obtained from the FT spectrum, as described in the Results section. By symmetrizing the NOESY spectrum, the covariance transform can suppress local maxima that are less reliable as indicated by a lack or gross shift in the

position of the symmetric local maximum. Features that are robust with respect to their characteristics before and after the covariance transform—this robustness being the criterion used for peak picking the covariance NMR spectrum—are significantly less likely to be artifacts than are less robust features. Figure 2 shows an example of this: the feature in the center of Fig. 2a, while not resolvable as a peak in the FT spectrum, is relatively unchanged, except for the increased donor resolution disambiguating it as a distinct peak, upon covariance. On the other hand, the feature to the left in Fig. 2a is suppressed by covariance NMR while the feature to the right has a gross shift in the position of its most intense portions. These non-robust characteristics do not correspond to NOEs given ubiquitin's known chemical shifts and structure.

The problem of identifying which peaks in an NMR spectra arise from expected through-bond or through-space interactions and which peaks are artifacts, prior to full knowledge of the chemical shifts and/or conformation of the molecule under study, is a difficult problem a solution to which will greatly enhance the automation of NMR-based structure determination (Altieri and Byrd 2004). By providing a perturbation of NMR spectra after which the true peaks are more likely to remain relatively un-perturbed, covariance NMR provides another view of a spectrum that aids in the identification of peaks that have a high likelihood to be valid.

Conclusions

4D covariance NMR provides substantial resolution enhancement of 4D FT spectra. In the case of NOESY, the method enables the resolution of broad, highly overlapped NOESY cross peaks into multiple distinct peaks with the resolution enhancement yielded by covariance NMR providing apparent line-widths closer to expected inherent line-widths. Even when using tighter tolerances, commensurate with the enhanced resolution of covariance spectra for matching donor chemical shifts, covariance NMR allows for the identification of additional NOEs, including long-range NOEs in terms of the protein sequence, not identifiable or distinguishable in FT NOESY spectra. The ability to extract more constraints, using tighter tolerances, increases the utility of 4D NOESY data in iterative structure calculations. Moreover, the decreased false positive rate of FT cross peaks associated with heuristically acceptable covariance cross peaks suggests a constructive role for covariance NMR in cross peak identification. 4D covariance spectra can be analyzed manually, as shown here, and they have unique potential for incorporation into fully automated procedures. An advantage of the method is that it can be applied to the same raw data that is used for FT processing enabling direct comparison and cross validation of the methods. Together, these properties make 4D covariance NMR a potentially valuable tool for resonance assignment and structure determination of larger proteins and for high-throughput structural biology applications.

Acknowledgements The authors thank Wolfgang Bermel for helpful discussions and acknowledge the support of NIH grant R01-GM066041.

References

Altieri AS, Byrd RA (2004) Automation of NMR structure determination of proteins. Curr Opin Struc Biol 14:547–553

Babu CR, Flynn PF, Wand AJ (2001) Validation of protein structure from preparations of encapsulated proteins dissolved in low viscosity fluids. J Am Chem Soc 123:2691–2692

Bax A, Clore GM, Gronenborn AM (1990) H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13-enriched proteins. J Magn Reson 88:425–431

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucl Acids Res 28:235–242

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

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

Brutscher B, Cordier F, Simorre JP, Caffrey M, Marion D (1995) High-resolution 3D HNCOCA experiment applied to a 28-kDa paramagnetic protein. J Biomol NMR 5:202–206

Celda B, Montelione GT (1993). Total correlation spectroscopy (TOCSY) of proteins using coaddition of spectra recorded with several mixing times. J Magn Reson Ser B 101: 189–193

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

Chen Y, Zhang F, Snyder DA, Gan Z, Brüschweiler-Li L, Brüschweiler R (2007) Quantitative covariance NMR by regularization. J Biomol NMR 38:73–77

Clore GM, Gronenborn AM (1991) Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. Science 252:1390–1399

Clore GM, Kay LE, Bax A, Gronenborn AM (1991) 4-Dimensional C-13/C-13-edited nuclear Overhauser enhancement spectroscopy of a protein in solution—application to interleukin 1-beta. Biochemistry 30:12–18

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

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

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

Farmer BT, Venters RA, Spicer LD, Wittekind MG, Muller L (1992) A refocused and optimized HNCA—increased sensitivity and resolution in large macromolecules. J Biomol NMR 2:195–202

Freeman R, Kupce E (2006) New ways to record multidimensional NMR spectra. Curr Anal Chem 2:101–105

Goddard TD, Kneller DG (1998) SPARKY v. 3. University of California, San Francisco, CA

Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein. J Magn Reson 96:432–440

Havel TF, Najfeld I, Yang J-X (1994) Matrix decompositions of two-dimensional nuclear magnetic resonance spectra. Proc Natl Acad Sci USA 91:7962–7966

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

Hoch JC, Stern AS (2001) Maximum entropy reconstruction, spectrum analysis and deconvolution in multidimensional nuclear magnetic resonance. Method Enzymol 338:159–178

Hu HT, De Angelis AA, Mandelshtam VA, Shaka AJ (2000) The multidimensional filter diagonalization method—II. Application to 2D projections of 2D, 3D, and 4D NMR experiments. J Magn Reson 144:357–366

Huang YJ, Powers R, Montelione GT (2005) Protein NMR recall, precision and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 127:1665–1674

Huang YJ, Tejero R, Powers R, Montelione GT (2006) A topology constrained distance network algorithm for protein structure determination from NOESY data. Proteins: Struct Funct Bioinfo 15:587–603

Jaravine V, Ibraghimov I, Orekhov VY (2006) Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. Nat Methods 3:605–607

Johnson RA, Wichern DW (2002) Applied multivariate statistical analysis. Prentice Hall

Kay LE, Clore GM, Bax A, Gronenborn AM (1990). 4-Dimensional heteronuclear triple-resonance NMR-spectroscopy of interleukin-1-beta in solution. Science 249:411–414

Kay LE, Ikura M, Tschudin R, Bax A (1990) 3-Dimensional triple-resonance NMR-spectroscopy of isotopically enriched proteins. J Magn Reson 89:496–514

Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665

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

Koehl P (1999) Linear prediction spectral analysis of NMR data. Prog Nucl Mag Res Sp 34:257–299

Kumar A, Ernst RR, Wüthrich K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. Biochem Bioph Res Co 95:1–6

Kupce E, Freeman R (2006). Hyperdimensional NMR spectroscopy. J Am Chem Soc 128:6020–6021

Led JJ, Gesmar H (1991) Application of the linear prediction method to NMR spectroscopy. Chem Rev 91:1413–1426

Macura S, Ernst RR (1980) Elucidation of cross relaxation in liquids by two-dimensional NMR-spectroscopy. Mol Phys 41:95–117

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

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

Misiak M, Kozminski W (2007) Three-dimensional NMR spectroscopy of organic molecules by random sampling of evolution time space and multidimensional Fourier transform. Magn Reson Chem 45:171–174

Morshauser RC, Zuiderweg ERP (1999) High-resolution four-dimensional HMQC-NOESY-HSQC spectroscopy. J Magn Reson 139:232–239

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

Rance M (1994) Sensitivity improvement in multidimensional NMR spectroscopy. Bull Magn Reson 16:54

Rovnyak D, Frueh DP, Sastry M, Sun ZYJ, Stern AS, Hoch JC, Wagner G (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. J Magn Reson 170:15–21

Seavey BR, Farr EA, Westler WM, Markley JL (1991) A relational database for sequence-specific protein NMR data. J Biolmol NMR 1:217–236

Solomon I (1955) Relaxation processes in a system of 2 spins. Phys Rev 99:559–565

Szyperski T, Atreya HS (2006) Principles and applications of GFT projection NMR spectroscopy. Magn Reson Chem 44:S51–S60

Szyperski T, Wider G, Bushweller JH, Wüthrich K (1993) Reduced dimensionality in triple-resonance NMR experiments. J Am Chem Soc 115:9307–9308

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

Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY (2005). High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. J Am Chem Soc 127:2767–2775

Uitenbroek DG (1997). SISA: simple interactive statistical analysis. (http://www.home.clara.net/sisa/)

Vuister GW, Clore GM, Gronenborn AM, Powers R, Garrett DS, Tschudin R, Bax A (1993) Increased resolution and improved spectral quality in four-dimensional 13C/13C-separated HMQC-NOESY-HMQC spectra using pulsed field gradients. J Magn Reson B 101:210–213

Wüthrich K (1986) NMR of proteins and nucleic acids. John Wiley and Sons, New York

Yee A, Gutmanas A, Arrowsmith CH (2006) Solution NMR in structural genomics. Curr Opin Struct Biol 16:611–617