Generalized Indirect Covariance NMR Formalism for Establishment of Multidimensional Spin Correlations

David A. Snyder[†,‡] **and Rafael Brüschweiler***[,‡]

Department of Chemistry, William Paterson University, 300 Pompton Road, Wayne, New Jersey 07470, and Chemical Sciences Laboratory, Department of Chemistry and Biochemistry, and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306

Received: July 23, 2009; Revised Manuscript Received: September 11, 2009

Multidimensional nuclear magnetic resonance (NMR) experiments measure spin−spin correlations, which provide important information about bond connectivities and molecular structure. However, direct observation of certain kinds of correlations can be very time-consuming due to limitations in sensitivity and resolution. Covariance NMR derives correlations between spins via the calculation of a (symmetric) covariance matrix, from which a matrix-square root produces a spectrum with enhanced resolution. Recently, the covariance concept has been adopted to the reconstruction of nonsymmetric spectra from pairs of 2D spectra that have a frequency dimension in common. Since the unsymmetric covariance NMR procedure lacks the matrix-square root step, it does not suppress relay effects and thereby may generate false positive signals due to chemical shift degeneracy. A generalized covariance formalism is presented here that embeds unsymmetric covariance processing within the context of the regular covariance transform. It permits the construction of unsymmetric covariance NMR spectra subjected to arbitrary matrix functions, such as the square root, with improved spectral properties. This formalism extends the domain of covariance NMR to include the reconstruction of nonsymmetric NMR spectra at resolutions or sensitivities that are superior to the ones achievable by direct measurements.

Introduction

Multidimensional nuclear magnetic resonance (NMR) is a powerful tool for probing molecular connectivity and structure by displaying magnetization transfer between nuclear spins due to their magnetic interaction as correlation peaks in a multidimensional spectrum.[1] However, multidimensional NMR spectra with high resolution and sensitivity require the acquisition of a large number of scans, which is NMR spectrometer time intensive.[2] Establishment of direct correlations between insensitive nuclei, such as ^{13}C and ^{15}N, requires particularly long measurement times.[3]

Indirect covariance NMR[4] offers a linear algebraic approach to establish correlations between pairs of heteronuclei that are coupled to a common set of protons. Formally, the indirect covariance transform of the $N_1 \times N_2$ NMR spectrum $\mathbf{X}$ produces the (symmetric) spectrum $\mathbf{C} = (\mathbf{X}\mathbf{X}^{\mathrm{T}})^{1/2}$ (where the superscripts T and 1/2 denote the matrix transpose and matrix-square root, respectively). Unsymmetric covariance NMR[5−8] generates asymmetric spectra via matrix multiplication of two distinct spectra that share (at least) one common dimension. An example is the multiplication of an ^{13}C−^{1}H HSQC[9] with a ^{1}H−^{1}H TOCSY[10] to correlate all ^{1}H and ^{13}C nuclei in the same spin system. This reconstructs a ^{13}C−^{1}H HSQC−TOCSY spectrum from two standard 2D experiments without requiring additional measurement time and thereby yields additional ^{13}C,^{1}H correlations, which can facilitate chemical shift assignment by linking unassigned ^{13}C chemical shifts to already assigned ^{1}H and ^{13}C chemical shifts.[6] Hyperdimensional NMR reconstructs high-dimensional spectra, which are often asymmetric, from lower dimensional spectra for the purpose of protein resonance assignment.[11,12] COBRA[13,14] and Burrow-Owl[15] apply linear algebraic spectral manipulations for the same purpose.

An important property of unsymmetric covariance NMR is that the sensitivity of the covariance spectrum is limited only by the sensitivity of the experiments it combines.[16] For example, unsymmetric covariance of an ^{13}C−^{1}H HMBC[17] with a ^{13}C−^{1}H HSQC spectrum establishes carbon−carbon correlations with the enhanced sensitivity characteristic of an inverse detected ^{13}C−^{1}H heteronuclear spectra rather than that of a direct detected ^{13}C−^{13}C correlation spectrum.[4]

A key difference between symmetric and unsymmetric covariance NMR is the applicability of the matrix-square root transform. The matrix-square root, which minimizes artifacts due to relay effects and chemical shift (near) degeneracy ("pseudo-relay effects")[4,18−20] is properly defined only for symmetric and positive semidefinite covariance spectra, e.g., when the product matrix is a regular covariance matrix.

In this paper, a general approach is presented for constructing a covariance matrix from multiple NMR spectra. Since the standard covariance transform is recovered as a special case when identical spectra are used as input, the generalized covariance matrix formalism reconciles symmetric and unsymmetric covariance processing. The generalized covariance matrix is symmetric, which makes it amenable to the extraction of arbitrary matrix functions, including the matrix-square root and other matrix powers λ. Depending on the types of spectra that are correlated, application of the square root suppresses false positives. It is found that the analysis of the variation of covariance peak intensity as a function of λ is an effective indicator for the identification of false positives in unsymmetric covariance spectra. Covariation of a ^{13}C−^{1}H HMBC with a

* To whom correspondence should be addressed. Tel.: 850-644-1768. Fax: 850-644-8281. E-mail: bruschweiler@magnet.fsu.edu.
† William Paterson University.
‡ Florida State University.

10.1021/jp9070168 CCC: $40.75
Published on Web 10/07/2009

^{1}H−^{1}H TOCSY spectrum to obtain reliable ^{13}C,^{1}H correlations not detectable in the HMBC experiment demonstrates the utility of this method. The generalized covariance formalism therefore expands the power of covariance NMR to the reconstruction of nonsymmetric spectra.

Theory

Unsymmetric indirect covariance NMR[5−8] takes an $N_{1,1} \times N_2$ 2D spectrum $\mathbf{X}_1$ (matrix) and an $N_{1,2} \times N_2$ 2D spectrum $\mathbf{X}_2$ and "concatenates" them into a single $N_{1,1} \times N_{1,2}$ spectrum $\mathbf{C}$ via matrix multiplication:

$$\mathbf{C} = \mathbf{X}_1 \cdot \mathbf{X}_2^{\mathrm{T}} \tag{1}$$

Matrix element C_{ij} of $\mathbf{C}$ is a measure of the correlation between the pair (i,j) of spins belonging to the ith row vector of $\mathbf{X}_1$ and the jth row vector of $\mathbf{X}_2$. Such a correlation either indicates a direct interaction between the two spins, a mutual correlation to a common third spin, e.g., via spin-diffusion in NOESY spectra,[18,21] or a pseudo-relay effect due to correlations to different spins with identical chemical shift. In the symmetric case, i.e. $\mathbf{X}_1 = \mathbf{X}_2$, extraction of the matrix-square root effectively reduces both relay and pseudo-relay effects.[18,19,22]

Generalized (indirect) covariance (GIC) NMR provides a framework in which unsymmetric covariance spectra are embedded in symmetric covariance spectra amenable to general matrix functions. GIC starts out with the construction of a stacked spectrum from n 2D spectra of dimensions $N_{1i} \times N_2$ ($i = 1, ..., n$):

$$\mathbf{S} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \tag{2}$$

A generalized covariance matrix is then defined as

$$\mathbf{C} = \mathbf{S} \cdot \mathbf{S}^{\mathrm{T}} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \cdot [\mathbf{X}_1^{\mathrm{T}} \quad \cdots \quad \mathbf{X}_n^{\mathrm{T}}] = \begin{bmatrix} \mathbf{X}_1\mathbf{X}_1^{\mathrm{T}} & \cdots & \mathbf{X}_1\mathbf{X}_n^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \mathbf{X}_n\mathbf{X}_1^{\mathrm{T}} & \cdots & \mathbf{X}_n\mathbf{X}_n^{\mathrm{T}} \end{bmatrix} \tag{3}$$

Because of Parseval's theorem, eq 3 yields (up to a constant prefactor) the same result irrespective of whether the direct dimensions of $\mathbf{X}_1$, ..., $\mathbf{X}_n$ are in the time domain or in the frequency domain.[18] Matrix $\mathbf{C}$ is symmetric and semipositive definite, which permits the straightforward calculation of arbitrary matrix functions, including matrix roots. For $n = 1$, eq 3 reduces to the indirect covariance NMR spectrum.[4] For $n \geq 2$, $\mathbf{C}$ contains the unsymmetric covariance matrix given in eq 1 as an off-diagonal submatrix. For simplicity, the GIC spectrum from $\mathbf{X}_1$ and $\mathbf{X}_2$ ($n = 2$) is denoted by $\mathbf{X}_1$*$\mathbf{X}_2$ and, when raised to the matrix power λ, by $[\mathbf{X}_1$*$\mathbf{X}_2]^\lambda$.

After application of singular value decomposition (SVD) to matrix $\mathbf{S}$ of eq 2, $\mathbf{S} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^{\mathrm{T}}$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\mathbf{D}$ is diagonal, eq 3 becomes

$$\mathbf{C} = \mathbf{S} \cdot \mathbf{S}^{\mathrm{T}} = (\mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^{\mathrm{T}})(\mathbf{V} \cdot \mathbf{D} \cdot \mathbf{U}^{\mathrm{T}}) = (\mathbf{U} \cdot \mathbf{D}^2 \cdot \mathbf{U}^{\mathrm{T}}) \tag{4}$$

For the matrix-square root, $\lambda = 1/2$, it follows $\mathbf{C}^{0.5} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{U}^{\mathrm{T}}$ and for general powers

$$\mathbf{C}^\lambda = \mathbf{U} \cdot \mathbf{D}^{2\lambda} \cdot \mathbf{U}^{\mathrm{T}} \tag{5}$$

Of practical importance, calculation of a series of spectra with different powers λ of $\mathbf{C}$ only requires a single SVD, which makes such calculations efficient.

The unsymmetric covariance matrix given by eq 1 constitutes an off-diagonal submatrix of the generalized covariance matrix $\mathbf{C}$ of eq 3. The same submatrix of $\mathbf{C}^\lambda$ defines the λth power of the unsymmetric covariance matrix including the matrix-square root of an unsymmetric covariance matrix.

GIC is applicable to a stack of spectra, $\mathbf{X}_1$, ..., $\mathbf{X}_n$, as long as each combination of covariance spectra $\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}$, $\mathbf{X}_1\mathbf{X}_2^{\mathrm{T}}$, ..., gives rise to non-diagonal blocks and thereby expands the block-diagonal parts stemming from the "auto-covariances" $\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}$. GIC can reconstruct any spectrum that factors into individually measurable NMR experiments. For example, a [^{13}C−^{1}H-HMBC*^{1}H−^{1}H-TOCSY]$^\lambda$ covariance spectrum reconstructs a 2D ^{13}C−^{1}H HMBC−TOCSY spectrum while [^{13}C−^{1}H-HMBC*^{15}N−^{1}H-HSQC]$^\lambda$ yields a 2D through-bond ^{13}C−^{15}N correlation spectrum.[23] Experiments probing spin-diffusion, relay, or multispin correlation effects (NOESY, TOCSY, HMBC) are particularly suitable for GIC analysis due to the analogy between the matrix (square) root operation of covariance NMR and the shortening of the experimental mixing time.[18]

In symmetric covariance, the matrix-square root minimizes artifacts due to pseudo-relay effects.[18,19,22] Likewise, the square root of the generalized covariance matrix suppresses artifacts in submatrices belonging to the unsymmetric covariance spectra. Hence, the intensities of pseudo-relay correlation peaks are systematically weakened by the root operation as compared to the intensities of bona fide signals. Generally, the more rapidly the covariance cross-peak intensity $C_{ij}(\lambda)$ increases with λ, the less likely is that peak to be a valid signal. Hence, the slope of log $[C_{ij}(\lambda)]$ as a function of λ serves as a useful metric by complementing signal intensity alone for assessing the veracity of the signal for matrix element (i,j).

Equation 5 may be rewritten in terms of matrix elements (where D_k denotes the kth singular value and U_{ik} the ith component of the kth singular vector)

$$C_{ij}(\lambda) = \sum_k U_{ik} D_k^{2\lambda} U_{jk} \tag{6}$$

Thus the slope of the natural log $[C_{ij}(\lambda)]$ is

$$\frac{\partial\{\log[C_{ij}(\lambda)]\}}{\partial\lambda} = \frac{2}{C_{ij}(\lambda)} \sum_k U_{ik} D_k^{2\lambda} \log D_k U_{jk} \tag{7}$$

Note that the plot of log $[C_{ij}(\lambda)]$ vs λ is typically a straight line (Figure 2) and thus the slope given by eq 7 is constant over a broad range of λ values.

Materials and Methods

2D ^{1}H−^{1}H-TOCSY[10] (90 ms mixing time using MLEV-17[24]) and ^{13}C−^{1}H-HMBC spectra[17] were recorded at 18.8 T and 298

K for a mixture of seven common metabolites at natural ^{13}C abundance (D-carnitine, D-glucose, L-glutamine, L-histidine, L-lysine, *myo*-inositol, and shikimic acid) each at a concentration of 10 mM in D_2O. The direct ^1H dimension of each spectrum was acquired with 2048 complex points and a spectral width of 8013 Hz. The indirect ^1H dimension of the TOCSY was acquired with 1024 complex points and the same spectral width as the direct dimension. The indirect ^{13}C dimensions of the HMBC spectrum was acquired with 1024 complex points and a spectral width of 32 206 Hz, respectively.

Additionally, 2D $^1H-^1H$-TOCSY (50 ms mixing time using DIPSI-2[25]) and $^{13}C-^1H$ HMBC spectra were also recorded at 298 K using a sample of the MDM2-binding p53 peptide construct with sequence ETFSDLWKLLPEN, described previously.[26] The spectra were acquired with the same spectral widths as above but with half the number of complex points along each dimension, except for the indirect dimension of the TOCSY having only 256 complex points, and with a spectral width of 44643 Hz in the indirect (^{13}C) dimension of the HMBC spectrum.

All spectra were recorded on a Bruker AVANCE 800 spectrometer equipped with a cryogenic probe and processed in NMRPipe.[27] For the HMBC spectra, a magnitude spectrum was calculated after 2D FT.[17] All other calculations were performed in MATLAB.[28]

Results

To demonstrate the approach, a generalized indirect covariance (GIC) HMCB*TOCSY spectrum for a two-component mixture was calculated from a simulated $^{13}C-^1H$ HMBC spectrum (Figure 1A) and $^1H-^1H$ TOCSY spectrum (Figure 1B) with sharp lines. The mixture consists of two molecules represented by two different spin systems: the first has three linked $^{13}C,^1H$ pairs, X−Y−Z, and the second has two pairs, U−V. To simulate the effects of overlap, the protons of pairs Y and U are assigned degenerate chemical shifts. Related models with different degenerate chemical shifts were explored, but all gave results similar to those reported here. $\lambda = 1$ gives rise to a false peak in the generalized indirect covariance spectrum between C_X-H_V, as indicated in (Figure 1C).

Figure 2A shows the suppression of the false positive C_X-H_V peak (red) achieved by varying the exponent λ in eq 5. This log−linear plot demonstrates the higher slope (eq 7) associated with the false positive signal (red) relative to the true signals (black).

Figure 2B shows the analogous plot for a GIC HMBC*TOCSY spectrum derived from experimental $^{13}C-^1H$-HMBC and $^1H-^1H$ TOCSY spectra of a metabolite mixture sample. The false positive signal, which incorrectly correlates a ^{13}C resonance of *myo*-inositol to a ^1H resonance of carnitine, exhibits a systematically stronger λ scaling compared to the true positive signals. Its intensity in the $\lambda = 1$ covariance matrix lies between the intensities of two true positive signals, a glucose cross-peak and a *myo*-inositol cross-peak, but when $\lambda = 0.5$, its intensity is only as high as the weaker of the two true signals and the slope of its intensity buildup as a function of λ is higher than the slope of the true signals. The higher slope and weaker intensity at $\lambda = 0.5$ provide a signature that this peak is a false positive.

Figure 3 demonstrates the preferential suppression of artifact signals via the matrix-square root in two GIC HMBC*TOCSY covariance spectra calculated from two experimental pairs of $^{13}C-^1H$-HMBC and $^1H-^1H$ TOCSY spectra recorded of the metabolite mixture (Figure 3A,B) and the p53 peptide (Figure 3C,D). The peak intensity better separates false peaks (red dots)

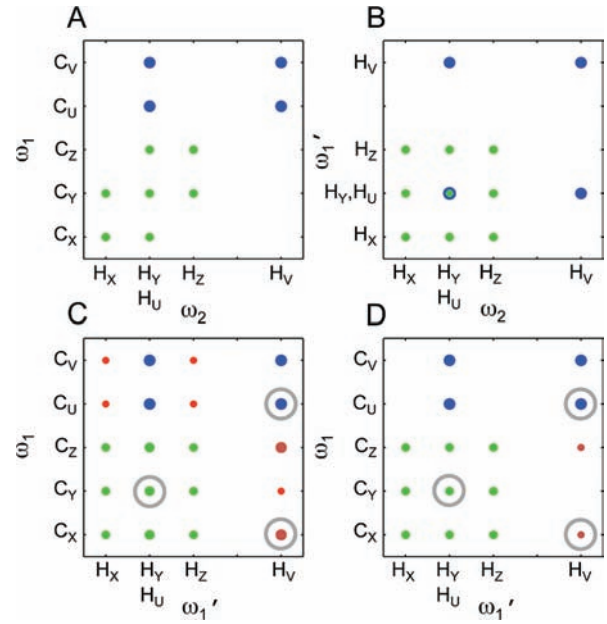

Figure 1. Schematic HMBC (A), TOCSY (B), [HMBC*TOCSY]^1 (C), and [HMBC*TOCSY]$^{0.5}$ (D) spectra for a model mixture containing one spin system with three connected $^{13}C-^1H$ pairs, X−Y−Z, and one spin system consisting of two $^{13}C-^1H$ pairs, U−V, where the carbon/carbon connectivities are among X−Y, Y−Z, and U−V. Note the degeneracy in chemical shift for the protons of $^{13}C-^1H$ pairs Y and U, which leads to false positive (red) signals in the [HMBC*TOCSY]$^{\lambda=1}$ spectrum. Application of the matrix-square root in the GIC formalism eliminates most false positives. The two most intense false positives (dark red) are not completely suppressed with decreasing λ. They can be identified as false positives because of their large slope as a function of λ. The peaks circled in gray are those whose traces are displayed in Figure 2.

from true peaks (black dots) in the $\lambda = 0.5$ spectrum than in the $\lambda = 1$ spectrum (Figure 3A,C and Table 1). However, while intensity in the $\lambda = 1$ spectrum alone is a relatively poor indicator of peak veracity, deviations from the trend visible among the true peaks in Figure 3A,C are indicative of peak authenticity: peaks lying on the upper left-hand side of the distribution marked by the ellipse, i.e., peaks for which the matrix-square root reduces peak intensity by a large amount, are most likely to be false.

Plotting the slope (eq 7) versus the intensity at $\lambda = 0.5$ also separates true from false peaks (Figure 3B,D). Peaks characterized by especially high slopes relative to their intensity (above and to the left of the ellipse surrounding most peaks) are most likely to be false. In fact, plotting the slope versus the intensity at $\lambda = 0.5$ identifies false peaks more effectively than does plotting intensity at $\lambda = 1$ versus that at $\lambda = 0.5$.

The selection procedure can be formalized by applying principal component analysis (PCA) in two dimensions,[29] which in good approximation reproduces the ellipses drawn in Figure 3. The major axis of the ellipse is given by the first principal component and the minor axis by the second principal component. PCA transforms intensity and slope into a new variable pair of independent statistics that is a linear combination of the original pair. The first principal component adjusts peak intensity using slope information, while the second component combines intensity and slope information into a measure of peak quality. Under the assumption that the principal components are Gaussian distributed, the value for the second principal component calculated for a given peak can be transformed into a *p*-value that quantifies the probability that this peak is real rather than an artifact arising from spurious chemical shift degeneracy.

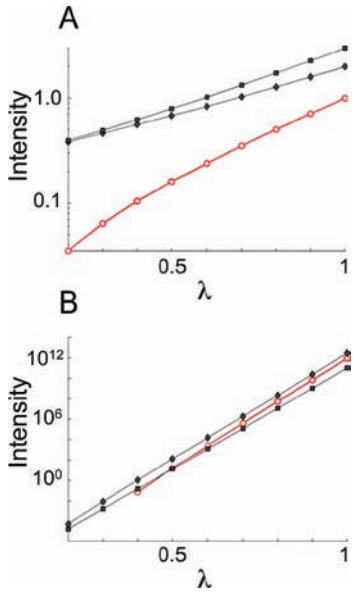

Figure 2. Increase of covariance peak intensity with respect to the exponent (λ) used in transforming the generalized covariance matrix. (A) Log−linear plot tracking the intensity buildup with increasing λ for three example traces from the simulated spectra of Figure 1. (B) Analogous plot for an experimental generalized indirect covariance (GIC) HMBC*TOCSY spectrum of a metabolite mixture. In panel B, the black curves belong to *myo*-inositol (stronger peak) and glucose (weaker peak). In all panels, black traces with filled circles correspond to expected signals while red traces with open circles correspond to false positive signals. Note the characteristically higher slopes of the false positive traces.

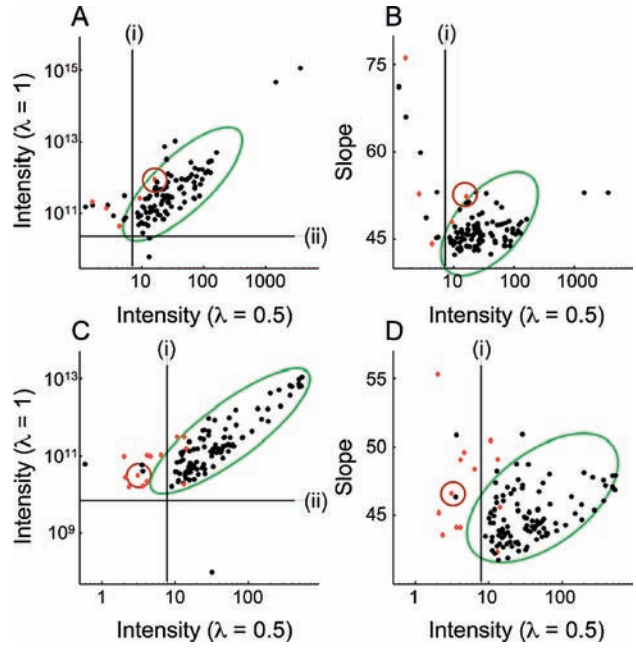

Figure 3. Suppression and identification of false positive signals via matrix-square root and λ scaling. (A, C) Comparison of intensity with $\lambda = 1$ and $\lambda = 0.5$ of (unsplit) peaks in covariance [HMBC*TOCSY]$^{\lambda}$ spectra of (A) a metabolite mixture and (C) the p53 peptide. (B, D) Comparison of slope vs intensity at $\lambda = 0.5$. Panels B and D show data corresponding to the pairs shown in panels A and C, respectively. Black dots represent data derived from true peaks, while red dots belong to false peaks. Line (i) demarcates the minimum intensity for which peaks are picked in the $\lambda = 0.5$ spectrum, while line (ii) demarcates the minimum intensity for which peaks are picked in the $\lambda = 1$ spectrum. The green ellipses surround the bulk of data to guide the eye. The red circles enclose (A, B) the false positive peak shown in Figures 2 and 4 and (C, D) the false positive peak shown in Figure 5.

TABLE 1: Reduction in False Positive Rate via Square-Root Extraction

	metabolite mixture		p53 (MDM2 binding peptide)[a]	
	$\lambda = 1$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 0.5$
true peaks	107	103	103	101
false peaks	6	2	15	4
false positive rate (%)	5	2	13	4

[a] Aliphatic/aliphatic $^{13}C-^{1}H$ peaks.

The following procedure allows one to edit peaks picked from a GIC-derived spectrum: (i) perform PCA as described above on (only) the peaks picked in the $\lambda = 0.5$ spectrum and (ii) reject peaks for which the *p*-value calculated (as in a one-tailed test) from the second principal component is less than 5%. Application of this procedure cuts the false-positive rate (reported for the $\lambda = 0.5$ spectra in Table 1) in half while only rejecting one (p53 peptide) and two (metabolite mixture) true peaks. The peaks plotted in Figure 3 include only those peaks reported in Table 1 whose line shapes do not qualitatively change as a function of λ as illustrated in Figure 4. This figure shows a region of the metabolite mixture GIC [HMBC*TOCSY]$^{\lambda}$ spectrum for different λ values. The unsymmetric covariance spectrum ($\lambda = 1$) displays a noise ridge (cross-hatched box)[16] due to the covariance of a signal arising from the carnitine methyl groups with noise. This ridge is suppressed after application of the matrix roots using the GIC formalism.

The decrease in intensity with decreasing λ for the false positive is again much more pronounced than for the other peaks: relative to the other peaks in panel A, peak 3 is quite strong, whereas it is weak relative to the other peaks in panel C and negative in panel D. The slope given by eq 7 at $\lambda = 0.5$ for this peak is 52, while a slope of 45 is typical for this data set. This peak appears in the upper left of Figure 3B (encircled in red) outside of the ellipse surrounding true peaks. Due to its high slope and low intensity at $\lambda = 0.5$, this peak can be easily identified and eliminated, improving the analysis of the GIC HMBC*TOCSY spectrum.

Application of $\lambda \leq 0.5$ also recovers the splitting present in the direct dimensions of the HMBC and TOCSY spectra of this mixture, which is lost by covariation of the direct dimension in the unsymmetric covariance process. However, the onset of distortions in line-shape (e.g., peak 2 in Figure 3D) and signal reduction generally preclude the use of very low λ values ($\lambda \leq 0.25$).

Figure 5 shows a region of the GIC HMBC*TOCSY spectrum of the p53 peptide. Again, the matrix-square root suppresses a false positive peak and a ridge, demonstrating the applicability of generalized covariance to larger systems, such as peptides. Unlike an experimentally recorded HSQC−TOCSY, the GIC HMBC*TOCSY exhibits correlations connecting quaternary and other nonprotonated carbons, such as carbonyl and carboxyl carbons, as illustrated in Figure 6. Thus, GIC provides a powerful representation of spectral information for the resonance assignment of small and large molecules, including peptides.

Discussion and Conclusions

Many informative spin correlations are not directly accessible by experiment by multidimensional NMR due to measurement

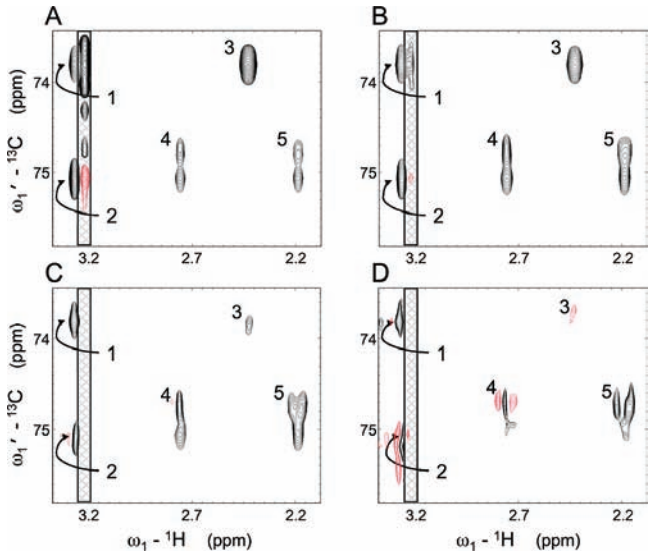

Figure 4. Spectral region of $[HMBC*TOCSY]^{\lambda}$ spectrum of a metabolite mixture with (A) $\lambda = 1$, (B) $\lambda = 0.75$, (C) $\lambda = 0.5$, (D) $\lambda = 0.25$. Black contours indicate positive signals, and red contours, negative signals. The cross-hatched region indicates a noise ridge, which is suppressed by the matrix power of $\lambda \leq 0.5$. Decreasing the value of λ also effectively suppresses peak 3, an artifact due to chemical shift near-degeneracy (pseudo-relay) between *myo*-inositol and carnitine ^{1}H resonances. Peaks 1 and 2 arise from *myo*-inositol, while peaks 4 and 5 arise from the geminal protons attached to C6 in the cyclohexene ring of shikimic acid. Their distorted line-shapes, particularly pronounced with $\lambda = 0.25$ (C, D) reflect *J*-splittings in the underlying HMBC and TOCSY spectra, corresponding to those observed in the 1D ^{1}H spectrum of shikimic acid available via the Biological Magnetic Resonance Bank.[38]

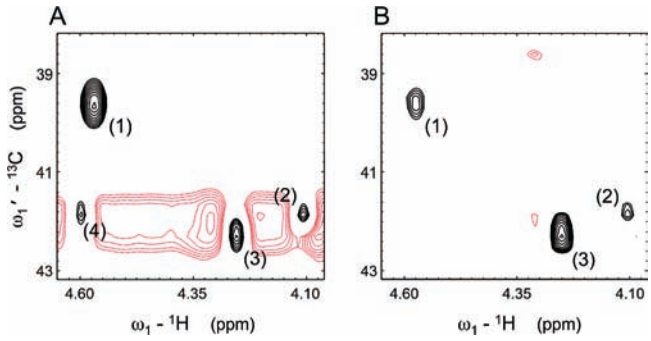

Figure 5. Selected region of the generalized indirect covariance $[HMBC*TOCSY]^{\lambda}$ spectrum of the p53 peptide calculated using (A) $\lambda = 1$ and (B) $\lambda = 0.5$. Black contours indicate positive signals, and red contours indicate negative signals. Peaks 1, 2, and 3 are Phe3 (CB-HA), Lys8 (CE-HA), and Leu10 (CB-HA), respectively. Peak 4 is a pseudo-relay artifact caused by accidental near-degeneracy that is suppressed by the matrix-square root, which also eliminates the horizontal ridge in panel A.

and sensitivity considerations. For instance, correlations between insensitive nuclei can often be observed only indirectly, i.e., via correlations between those nuclei via protons. Other spectra, such as heteronuclear NOESY and TOCSY, which contain useful information for resonance assignment and structure determination of complex molecules, are often not collected due to limited sensitivity and spectrometer time constraints. However, unsymmetric covariance NMR can reconstruct heteronuclear TOCSY and NOESY spectra from homonuclear NOESY and TOCSY spectra and common heteronuclear ^{13}C−^{1}H HSQC or HMBC spectra.[7]

Similarly, the high-dimensional correlation information required to make chemical shift assignments in polypeptides can

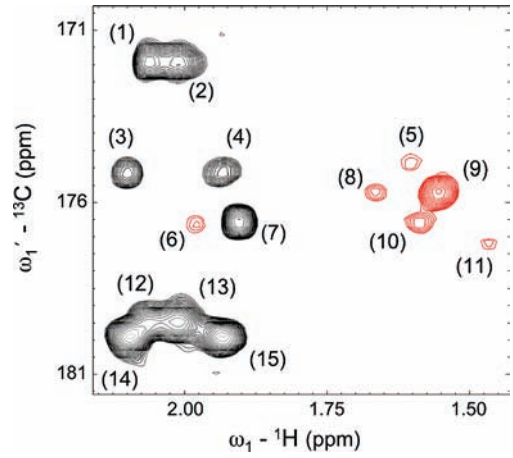

Figure 6. Novel long-range carbonyl-proton and carboxyl-proton correlations of p53 peptide derived from GIC $[HMBC*TOCSY]^{1/2}$. Cross-peaks initially present in the ^{13}C−^{1}H-HMBC spectrum are depicted by black contours. Peaks arising from covariance of the HMBC with the ^{1}H−^{1}H TOCSY spectrum are colored in red. The assignments of the peaks are as follows: (1) Glu12 (C′-HB2), (2) Glu12 (C′-HB3), (3) Glu1 (C′-HB2), (4) Glu1 (C′-HB3), (5) Leu9 (C′-HB2), (6) Pro11 (C′-HG3), (7) Pro11 (C′-HB3), (8) Lys8 (C′-HB2), (9) Lys8 (C′-HG2), (10) Leu10 (C′-HG), (11) Leu6 (C′-HG), (12) Glu12 (CD-HB2), (13) Glu12 (CD-HB3), (14) Glu1 (CD-HB2), and (15) Glu1 (CD-HB3). While ^{13}C−^{1}H HSQC and HSQC-TOCSY spectra lack carbonyl/carboxyl-proton cross-peaks, the ^{13}C−^{1}H HMBC spectrum correlates carbonyl and side-chain carboxyl carbons (such as the δ-carbons of glutamic acid) with protons via two and three bond correlations. The inclusion of TOCSY information via GIC processing results in longer-range correlations, such as those shown here in red.

often only be practically measured by a series of lower dimensional spectra. A typical manual analysis of NMR spectra establishes higher order correlations via a comparison of strip plots. Visual assessment of a nonvanishing correlation of peaks between slices (strip plots) in two NMR spectra links the spin-systems associated with the strip plots being compared. Automated analysis methods, particularly those for protein backbone assignment,[30−37] often work with peak lists rather than with the underlying spectra. However, such methods generally require high-quality peak lists that are manually curated. Recently developed methods such as hyperdimensional NMR,[11,12] COBRA,[13] and Burrow-Owl[15] use unsymmetric covariance[5,7] to automate the traditional manual approach of establishing spin correlations via comparison of strip plots, prior to peak picking. However, the application of such methods can confound downstream analysis due to the presence of spurious correlations between strip plots caused by (near-)degenerate chemical shifts and therefore may benefit from the generalized indirect covariance approach presented here. GIC establishes correlations between spectra rather than peak lists and thereby "delays" the otherwise iterative and sometimes difficult process of peak picking until true peaks become self-evident.

The GIC formalism generalizes the use of the matrix-square root for the suppression of relay effects and pseudo-relay effects, originally demonstrated for symmetric covariance NMR spectra,[18,19] to unsymmetric covariance spectra.[6] Previous work in covariance reconstruction of unsymmetric spectra compared unsymmetric and indirect covariance results in order to identify artifacts in each.[20] The generalized covariance matrix (eq 3) presented here computes both unsymmetric and symmetric covariance spectra in the same step. Furthermore, the GIC formalism allows for the extraction of multiple roots in a single covariance calculation. For the examples given here, extraction of the square root via the generalized covariance matrix reduces

the false positive count of a HMBC*TOCSY spectrum by about a factor of 3. Removal of peaks characterized by weak intensity following extraction of the square root concomitant with a rapid intensity buildup with λ further reduces the false positive rate.

The generalized covariance formalism addresses the issue of false positives in unsymmetric covariance spectra caused by resonance overlap and extends the applicability of unsymmetric covariance NMR to systems with an increased number of signals of greater resonance degeneracy, including complex mixtures, for example, of metabolites and biological macromolecules, such as peptides and proteins. By providing a mechanism to identify false positive correlations, generalized indirect covariance lays a linear-algebraic foundation for the accurate and sensitive identification of spin correlations that are distributed over multiple 2D NMR spectra. The establishment of spin correlations that are not easily experimentally observable via an automated method analogous to the comparison of strip plots marks a path toward the development of computer-based assignment procedures that are as robust as are the most expert manual analyses of NMR data.

Acknowledgment. We thank Fengli Zhang and Scott Showalter for kindly providing us with the metabolite mixture and p53 peptide NMR spectra, respectively, and Wolfgang Bermel for useful discussion. This work was supported by the National Institutes of Health (Grant GM 066041). The NMR experiments were conducted at the National High Magnetic Field Laboratory (NHMFL) supported by cooperative agreement DMR 0654118 between the NSF and the State of Florida.

References and Notes

(1) Ernst, R. Bodenhausen, G. Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press: Oxford, 1987.

(2) Jaravine, V.; Ibraghimov, I.; Orekhov, V. Y. *Nat. Methods* **2006**, *3*, 605–607.

(3) Friebolin, H. *Basic One-and Two-Dimensional NMR Spectroscopy*; Wiley-VCH: Weinheim, 2005.

(4) Zhang, F.; Brüschweiler, R. *J. Am. Chem. Soc.* **2004**, *126*, 13180–13181.

(5) Blinov, K. A.; Larin, N. I.; Kvasha, M. P.; Moser, A.; Williams, A. J.; Martin, G. E. *Magn. Reson. Chem.* **2005**, *43*, 999–1007.

(6) Blinov, K. A.; Larin, N. I.; Williams, A. J.; Mills, K. A.; Martin, G. E. *J. Heterocycl. Chem.* **2006**, *43*, 163–166.

(7) Blinov, K. A.; Larin, N. I.; Williams, A. J.; Zell, M.; Martin, G. E. *Magn. Reson. Chem.* **2006**, *44*, 107–109.

(8) Blinov, K. A.; Williams, A. J.; Hilton, B. D.; Irish, P. A.; Martin, G. E. *Magn. Reson. Chem.* **2007**, *45*, 544–546.

(9) Bodenhausen, G.; Ruben, D. G. *Chem. Phys. Lett.* **1980**, *69*, 185–189.

(10) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521–528.

(11) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2006**, *128*, 6020–6021.

(12) Kupce, E.; Freeman, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *52*, 22–30.

(13) Lescop, E.; Brutscher, B. *J. Am. Chem. Soc.* **2007**, *129*, 11916–11917.

(14) Lescop, E.; Rasia, R.; Brutscher, B. *J. Am. Chem. Soc.* **2008**, *130*, 5014–5015.

(15) Benison, G.; Berkholz, D. S.; Barbar, E. *J. Magn. Reson.* **2007**, *189*, 173–181.

(16) Snyder, D. A.; Ghosh, A.; Zhang, F.; Szyperski, T.; Brüschweiler, R. *J. Chem. Phys.* **2008**, *129*, 104511.

(17) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.

(18) Brüschweiler, R. *J. Chem. Phys.* **2004**, *121*, 409–414.

(19) Trbovic, N.; Smirnov, S.; Zhang, F. L.; Brüschweiler, R. *J. Magn. Reson.* **2004**, *171*, 277–283.

(20) Martin, G. E.; Hilton, B. D.; Blinov, K. A.; Williams, A. J. *Magn. Reson. Chem.* **2008**, *46*, 138–43.

(21) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95–117.

(22) Snyder, D. A.; Zhang, F.; Brüschweiler, R. *J. Biomol. NMR* **2007**, *39*, 165–175.

(23) Kupce, E.; Freeman, R. *Magn. Reson. Chem.* **2007**, *45*, 103–105.

(24) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355–360.

(25) Shaka, A.; Lee, C.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274–293.

(26) Showalter, S. A.; Bruschweiler-Li, L.; Johnson, E.; Zhang, F.; Brüschweiler, R. *J. Am. Chem. Soc.* **2008**, *130*, 6472–6478.

(27) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–93.

(28) *MATLAB*, 7.1.0.183 ed.; The MathWorks: Natick, MA, 2005.

(29) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer: New York, 2002.

(30) Altieri, A. S.; Byrd, R. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 547–553.

(31) Moseley, H. N. B.; Monleon, D.; Montelione, G. T. *Nucl. Magn. Reson. Biol. Macromol., Part B* **2001**, *339*, 91–108.

(32) Li, K. B.; Sanctuary, B. C. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 359–366.

(33) Xu, Y. Z.; Wang, X. X.; Yang, J.; Vaynberg, J.; Qin, J. *J. Biomol. NMR* **2006**, *34*, 41–56.

(34) Wang, J. Y.; Wang, T. Z.; Zuiderweg, E. R. P.; Crippen, G. M. *J. Biomol. NMR* **2005**, *33*, 261–279.

(35) Jung, Y. S.; Zweckstetter, M. *J. Biomol. NMR* **2004**, *30*, 11–23.

(36) Coggins, B. E.; Zhou, P. *J. Biomol. NMR* **2003**, *26*, 93–111.

(37) Atreya, H. S.; Sahu, S. C.; Chary, K. V. R.; Govil, G. *J. Biomol. NMR* **2000**, *17*, 125–136.

(38) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H. Y.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36*, D402−D408.

JP9070168