J|A|C|S
COMMUNICATIONS

Published on Web 11/16/2006

Enhanced Covariance Spectroscopy from Minimal Datasets

Yanbin Chen,[†] Fengli Zhang,[†] Wolfgang Bermel,[‡] and Rafael Brüschweiler*,[†]

Department of Chemistry and Biochemistry, National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306, and Bruker BioSpin GmbH, Silberstreifen, D-76287, Rheinstetten, Germany

Received July 31, 2006; E-mail: bruschweiler@magnet.fsu.edu

Speeding up of multidimensional NMR spectroscopy has received significant attention recently, spurred by the vast demands of the rapidly expanding fields of structural genomics and metabolomics. Numerous approaches have been proposed,[1−10] some of which take advantage of the sparse peak distribution of NMR spectra in multiple dimensions. The widely used TOCSY, COSY, and NOESY spectra can be considerably crowded, and the minimal measurement time is often limited by resolution rather than sensitivity.

Covariance NMR spectroscopy[11] has been shown to provide maximal resolution along multiple frequency axes. But when the number of acquired data points is too small, the covariances exhibit poor statistics that manifest themselves as spurious cross-peaks. A novel approach is presented here which addresses this situation. It uses a sparse sampling scheme along the indirect dimension followed by a comprehensive analysis of finite sampling effects. The scheme yields correlation spectra with high spectral resolution from remarkably small datasets.

The enhancement scheme is demonstrated here for a TOCSY spectrum of the cyclic decapeptide antamanide [-Val-Pro-Pro-Ala-Phe-Phe-Pro-Pro-Phe-Phe-], a peptide that acts as antidote of the poisonous mushroom amanita phalloides.[12]

A homonuclear TOCSY spectrum[13] of 1 mM antamanide dissolved in deuterated chloroform exhibits several crowded regions and partial peak overlaps (Figure 1).[14] The aliphatic section of the 2D Fourier transform (FT) reference spectrum with 2048 complex t_1 increments is depicted in panel c. When using only the first 48 complex t_1 increments, that is, about 2% of original time-domain data, the 2D FT spectrum of panel b shows severe line broadening along the indirect dimension ω_1, which impedes simple analysis.

The covariance spectrum is determined using the mixed time−frequency domain data **S**:[11b,c]

$$\mathbf{C} = (\mathbf{S}^{\mathrm{T}}\mathbf{S})^{1/2} \qquad (1)$$

where **S** is the $N_1 \times N_2$ mixed time-frequency domain matrix after Fourier transform along the detection dimension t_2 and N_1, N_2 are the number of data points along the indirect t_1 and detection dimensions, respectively. The matrix square-root is efficiently determined by singular value decomposition.[11c]

An advantage of eq 1 over traditional 2D FT is that the indirect dimension of **S** is not required to be sampled with a time increment Δt_1 that fulfills the Nyquist theorem, $\Delta t_{1,\mathrm{Nyq}} = 1/(\text{spectral width})$.[11] Importantly, if N_1 is to be minimized to achieve maximal speed up, undersampling in t_1 can be advantageous by probing a wider range of t_1 evolution times, which allows better discrimination between true and spurious spin correlations. Undersampling manifests itself in covariance spectra in the appearance of parallel diagonals, which is different from the well-known folding effects in 2D FT spectra. These effects are described in the Supporting Information with application to a covariance 2QF-COSY spectrum[15] of antamanide.

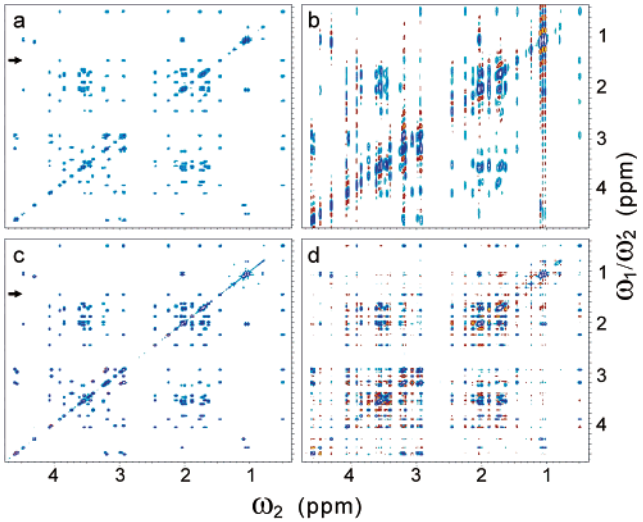

Figure 1. Aliphatic proton region of 2D TOCSY NMR spectra of the 1 mM antamanide peptide in $CDCl_3$ using different methods; covariance spectrum using 48 t_1 increments (a) after and (d) before masking; 2D FT spectrum using (b) 48 t_1 increments and (c) 2048 t_1 increments. All data were collected at 800 MHz proton frequency, 100 ms mixing time, and 300 K using TPPI-states along the indirect dimension. Cross sections at the arrow positions are given in the Supporting Information.

The TOCSY covariance spectrum **C** obtained using 48 t_1 points that are sampled at the Nyquist increment $\Delta t_{1,\mathrm{Nyq}}$ with the first t_1 point at $25\Delta t_{1,\mathrm{Nyq}}$ is depicted in Figure 1d. The delay in t_1 improves the spectral quality while it does not cause first order phasing problems along the indirect dimension, which is a convenient feature of covariance processing.[11] Another characteristic of covariance NMR is the same high resolution along both frequency dimensions. Comparison with the 2D FT spectrum using 2048 increments (Figure 1c) reveals, however, the presence of extra signals reflecting the onset of poor sampling effects because of the small size of the dataset.

Despite the "statistical" character of the covariance analysis, both the sampling scheme and the resonance positions are in fact fully deterministic in nature. As a consequence, the expected amount of spurious correlations can be determined for each point in the 2D spectrum and regions that are prone to artifacts can be identified or "masked" and subsequently removed. In this way, the occurrence of false-positive cross-peaks in the covariance spectrum is minimized without giving up the resolution advantage.

Application of the resulting mask, shown in Figure 2, to the covariance spectrum (Figure 1d) leads to the covariance spectrum of Figure 1a, which is essentially void of any false peaks while only five true off-diagonal peaks are missing.

[†] Florida State University.
[‡] Bruker BioSpin GmbH.

10.1021/ja065522e CCC: $33.50

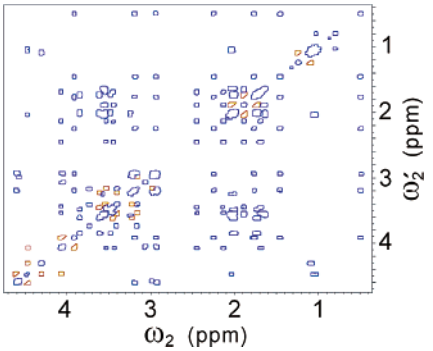

Figure 2. Spectral mask of TOCSY spectrum using the procedure described in the text.

Spectral masking is based on two criteria. First, spurious correlations between resonances are determined on the basis of their intensity, frequencies, and t_1 sampling in the following way. From the diagonal of the original covariance spectrum, a 1D spectrum $F(\omega_2)$ is constructed after applying a baseline correction and the square-root operation. Next, a control matrix is constructed

$$S_{\mathrm{ref}}(t_1,\omega_2) = F(\omega_2)\exp(i\omega_2 t_1) \qquad (2)$$

where t_1 and ω_2 take the same values as in **S**. A covariance spectrum $\mathbf{C}_{\mathrm{ref}}$ is computed from $\mathbf{S}_{\mathrm{ref}}$ using eq 1 and then normalized according to $R_{ij} = \{C_{ij,\mathrm{ref}}\}/\{(C_{ii,\mathrm{ref}}C_{jj,\mathrm{ref}})^{1/2}\}$. The cross-correlation coefficients R_{ij} quantify the spurious correlations between spectral points i and j associated with the applied t_1 sampling scheme. These correlations are independent of the J-coupling mediated TOCSY correlations of interest and are masked when they exceed a threshold value of 5%.

The second criterion is based on cross-peak appearance. Most spurious peaks differ in their shape from a typical cross-peak in that they tend to exhibit sharp isolated features that lack neighboring peaks of similar magnitude. These peaks are easily identified by the absence of a peak cluster surrounding them. If no cluster is present, the peak and its surrounding are masked. The mask constructed from Figure 1d based on the two criteria is shown in Figure 2. The red areas are masked using the first criterion (peaks inside these areas are removed) and the blue areas using the second criterion (peaks outside these areas are removed).

In conclusion, the enhanced covariance method presented here provides high-quality high-resolution 2D spectra from minimal t_1 datasets. The cross-validation and masking scheme efficiently suppress spurious correlations, while only few weak peaks, that are present in the full 2D FT spectrum, are absent. The scheme thereby offers a substantial saving of measurement time for TOCSY-type spectra and related variants of multidimensional experiments. The approach should be particularly well suited for high-throughput screening applications.

Acknowledgment. We thank Dr. David Snyder for discussion. This work was supported by NIH (Grant GM066041).

Supporting Information Available: Cross sections of TOCSY spectra of Figure 1, analysis of undersampling effects in covariance NMR spectroscopy, and application to 2QF-COSY experiment of antamanide.

References

(1) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125*, 1385−1393.
(2) Kupce, E.; Freeman, R. *J. Magn. Reson.* **2003**, *163*, 56−63.
(3) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2004**, *126*, 6429−6440.
(4) Coggins, B. E.; Venters, R. A.; Zhou, P. *J. Am. Chem. Soc.* **2005**, *127*, 11562−11563.
(5) Eghbalnia, H. R.; Bahrami, A.; Tonelli, M.; Hallenga, K.; Markley, J. L. *J. Am. Chem. Soc.* **2005**, *127*, 12528−12536.
(6) Malmodin, D.; Billeter, M. *J. Am. Chem. Soc.* **2005**, *127*, 13486−13487.
(7) Rovnyak, D.; Frueh, D. P.; Sastry, M.; Sun, Z. Y. J.; Stern, A. S.; Hoch, J. C.; Wagner, G. *J. Magn. Reson.* **2004**, *170*, 15−21.
(8) Chen, J. H.; Mandelshtam, V. A.; Shaka, A. J. *J. Magn. Reson.* **2000**, *146*, 363−368.
(9) Frydman, L.; Scherf, T.; Lupulescu, A. *Proc. Nat. Acad. Sci. U.S.A.* **2002**, *29*, 15858−15862.
(10) Schanda, P.; Brutscher, B. *J. Am. Chem. Soc.* **2005**, *127*, 8014−8015.
(11) (a) Brüschweiler, R.; Zhang, F. *J. Chem. Phys.* **2004**, *120*, 5253−5260. (b) Brüschweiler, R. *J. Chem. Phys.* **2004**, *121*, 409−414. (c) Trbovic, N.; Smirnov, S.; Zhang, F.; Brüschweiler, R. *J. Magn. Reson.* **2004**, *171*, 277−283.
(12) Wieland, T.; Faulstich, H. *Crit. Rev. Biochem.* **1978**, *5*, 185−260.
(13) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521−528.
(14) Kessler, H.; Müller, A.; Pook, K.-H. *Liebigs Ann. Chem.* **1989**, 903−912.
(15) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1984**, *117*, 479−485.

JA065522E