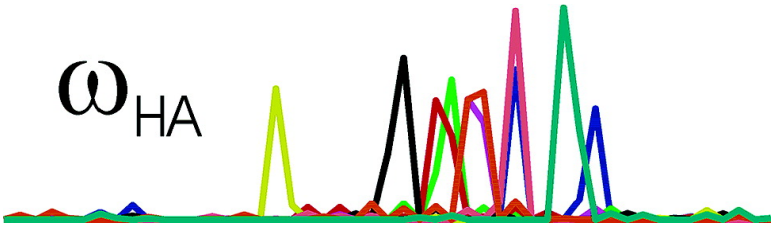
Communication

Multiway Decomposition of NMR Spectra with Coupled Evolution Periods

Daniel Malmodin, and Martin Billeter

J. Am. Chem. Soc., **2005**, 127 (39), 13486-13487• DOI: 10.1021/ja0545822 • Publication Date (Web): 13 September 2005



J|A|C|S
COMMUNICATIONS

Published on Web 09/13/2005

Multiway Decomposition of NMR Spectra with Coupled Evolution Periods

Daniel Malmodin and Martin Billeter*

Biophysics Group, Department of Chemistry, Göteborg University, Box 462, 405 30 Göteborg, Sweden

Received July 11, 2005; E-mail: martin.billeter@chem.gu.se

NMR applied to proteins relies on multinuclear experiments to uniquely identify all signals and assign all resonances. Observation of an additional nucleus is typically achieved by adding a spectral dimension, with significant consequences on both experiment time and spectrum size. By coupling different evolution periods during the experiment, a few lower-dimensional spectra can be recorded instead (for recent reviews see refs 1 and 2). Signal frequencies in these spectra are linear combinations of the frequencies of individual nuclei, and thus correlations exist among signals from different spectra. Ideally, spectra with coupled evolution times yield in much shorter time identical information as the corresponding full-dimensional spectrum.

The NMR response of an *N*-dimensional experiment with evolution/acquisition times t_i can be described as[3]

$$s(t_1 \ldots t_N) = \sum_k f_1^k(t_1) \otimes f_2^k(t_2) \ldots \otimes f_N^k(t_N) \qquad (1)$$

The sum runs over different response components; the outer product $\otimes$ indicates the relation between the one-dimensional functions $f_j(t_j)$ and the *N*-dimensional response $s(t_1 \ldots t_N)$. Fourier transform of eq 1 yields the following description for an *N*-dimensional spectrum *S* with frequencies ω_i

$$S(\omega_1 \ldots \omega_N) = \sum_k F_1^k(\omega_1) \otimes F_2^k(\omega_2) \ldots \otimes F_N^k(\omega_N) \qquad (2)$$

Equation 2 represents the basis for a general tool[4] introduced to NMR as "three-way decomposition" (TWD, implementation MUNIN[5]), where the functions $F_j^k(\omega_j)$ are referred to as shapes and the terms in the sum as components. In experiments with coupled evolution periods, time variables become correlated, and the outer products are replaced by normal products indicating the dimensionality reduction. Thus, the NMR response of a 5D experiment with coupling of all four evolution periods becomes (t_1,t_2,t_3,$t_4 \sim t$):

$$p_m(t,t_5) = \sum_k f_1^k(t,c_{1m}) \cdot f_2^k(t,c_{2m}) \cdot f_3^k(t,c_{3m}) \cdot f_4^k(t,c_{4m}) \otimes f_5^k(t_5) \qquad (3)$$

The index *m* enumerates different coupling schemes used (i.e., different projections recorded), the c_{jm} reflects the effect of a given type of coupling on dimension *i*.[6] Fourier transform of eq 3 yields convolutions of the first four shapes $F_j^k(\omega_j)$:[7]

$$P_m(\omega,\omega_5) = \sum_k [F_1^k(c_{1m}) * F_2^k(c_{2m}) * F_3^k(c_{3m}) * F_4^k(c_{4m})](\omega) \otimes F_5^k(\omega_5) \qquad (4)$$

The symbol "*" represents the convolution operation.

Equation 4 represents a model description[4,5] of a set of 2D spectra $P_m(\omega,\omega_5)$, derived from a 5D spectrum, with arbitrary coupling schemes among the evolution times. Due to noise and other artifacts, the model to the right of this equation will not exactly describe the experimental input on the left, but a residual will remain. The novel tool PRODECOMP determines the optimal 1D shapes $F_j^k(\omega_j)$, and therewith the shifts for all nuclei, from a set of spectra P_m recorded with coupled evolution times by minimizing the size of this residual. The three examples below illustrate the major advantages of this tool: (a) all spectra enumerated by *m* are used *simultaneously*, avoiding sensitivity loss associated with individual examination of the spectra,[8] (b) folding (aliasing) caused by the linear combinations of individual shifts is resolved, allowing an increase in resolution, (c) various combinations and types of coupling schemes are possible, (d) high-dimensional spectra can be reconstructed from the shapes $F_j^k(\omega_j)$ using eq 2, and (e) the approach is robust with respect to overlap in both the projected and the directly detected dimensions. Finally, no initial assumptions or start values are needed for the shapes F_j^k.

In this study, PRODECOMP is applied to various regions of spectra resulting from a (5,2)D HACACONHN-type GFT[9] experiment, which was recorded for the 14 kD protein azurin.[10,11] To better illustrate capabilities (a) and (c) of this tool, we exclude here spectra with coupling of only one or two nuclei (^{15}N HSQC, HNCO) that typically exhibit best resolution and S/N (this is not meant as a recommendation!). Therefore, only spectra with the following 12 coupling schemes were used: $\omega_N \pm \omega_{CO} \pm \omega_{C\alpha}$ and $\omega_N \pm \omega_{CO} \pm \omega_{C\alpha} \pm \omega_{H\alpha}$. A consequence of this data exclusion follows from the general observation that projections at angles $\pm\alpha$ can normally not provide unique line shapes along the basic axes (e.g., for signals with either (a,b) or (b,a) as line widths along (ω_1,ω_2), projections at $+45°$ and $-45°$ yield the same result). With the presently used constant time experiment, line shape information would be of little interest anyway. Thus, the four shapes $\omega_N + \omega_{CO} + \omega_{C\alpha}$, $2 \cdot \omega_{CO}$, $2 \cdot \omega_{C\alpha}$, and $\omega_{H\alpha}$ were used in addition to ω_{NH}.[6] Note that this does not prevent the determination of all chemical shifts, including ω_N! CPU times on a Linux-PC increase approximately with the square of the number of components starting with about 1 min for a one-component problem.

All 12 input spectra are shown in Figure 1A for an interval near $\omega_{HN} = 8.0$ ppm containing nine spin systems with overlapping signals along ω_{HN}. A PRODECOMP run with user specification of the interval and 9 as the number of components completely separates all nine spin systems. From the output shapes (Figure 1B) unique and correct[10] chemical shifts can be derived for all nuclei, including two Hα of a glycine. Consistency and correctness of the shapes in Figure 1B is also indicated by the reconstructions shown in Figure 1C of selected input spectra based on these shapes. The major difference between Figure 1A and 1C is the removal of some noise in the reconstructions.

For the present experiment, all 12 input spectra contain a very similar number of signals, which is related to the number of residues. Due to the presence of linear combinations of chemical shifts, accommodating all peaks without folding would require a wide spectral width, which has a negative impact on resolution. The input spectra in Figure 2A show that for Gly-Thr124 some of the combined frequencies exceed the spectral width, resulting in

 10.1021/ja0545822 CCC: $30.25

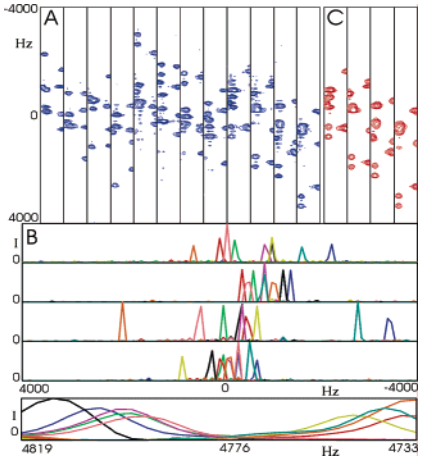


Figure 1. Analysis of the spectral region with $8.03 > \omega_{NH} > 7.89$ ppm. (A) All 12 input spectra with $\omega_N \pm \omega_{CO} \pm \omega_{C\alpha}$ and $\omega_N \pm \omega_{CO} \pm \omega_{C\alpha} \pm \omega_{H\alpha}$. (B) PRODECOMP output shapes $\omega_N + \omega_{CO} + \omega_{C\alpha}$, $2{\cdot}\omega_{CO}$, $2{\cdot}\omega_{C\alpha}$, $\omega_{H\alpha}$, and ω_{HN} (top to bottom); color code: S94 black, L68 blue (preceded by glycine), D71 magenta, L86 green, L102 pink, F111 yellow, V59 cyan, V99 red, M44 orange (arbitrary units used for intensities I). (C) Reconstructions of the four rightmost input spectra of (A), using the same contour levels.

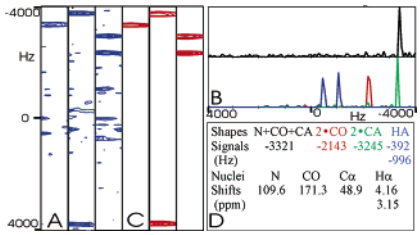


Figure 2. Folding of signals in spectra with coupled evolution periods. (A) Selected input spectra ($\omega_N + \omega_{CO} + \omega_{C\alpha}$ and $\omega_N + \omega_{CO} + \omega_{C\alpha} \pm \omega_{H\alpha}$) for Gly-Thr124 with $\omega_{NH} = 7.89$ ppm. (B) PRODECOMP output shapes $\omega_N + \omega_{CO} + \omega_{C\alpha}$ (black), $2{\cdot}\omega_{CO}$ (red), $2{\cdot}\omega_{C\alpha}$ (green), and $\omega_{H\alpha}$ (blue) (arbitrary intensity units). (C) Reconstructions using the shapes from (B) of the input spectra shown in (A). (D) List with signal positions in the output shapes of (B) and the resulting chemical shifts for N, CO, Cα, and Hα.

folded peaks. This becomes obvious when noting that, in the spectrum with $\omega_N + \omega_{CO} + \omega_{C\alpha} + \omega_{H\alpha}$, the two peaks caused by the presence of two α-protons are at opposite sides of the spectrum, while they are next to each other in the $\omega_N + \omega_{CO} + \omega_{C\alpha} - \omega_{H\alpha}$ spectrum. The matrix-based calculations used in PRODECOMP implicitly perform a "wrap-around" during multiplication steps,[6] eliminating any consequences of folding. As Figure 2B shows, well-defined shapes without folding effects are obtained, providing correct results (Figure 2D). Consequently, in future experiments of this type, folding caused by coupling of several evolution periods can be ignored, and the spectral widths in the projected dimensions may be chosen as if only one nucleus was present.

Sensitivity is a particular critical issue in all types of fast spectroscopy.[1,2] In experiments with coupled evolution periods, signal intensity is typically distributed over several spectra. For best sensitivity, it is therefore crucial to analyze all spectra simultaneously (e.g., to avoid independent peak picking in the individual spectra).[11] Asn38, preceded by a glycine, exhibits a very weak HN signal at 11.36 ppm. Figure 3A shows cross sections with fixed ω_{HN} for three input spectra with the following coupling of frequencies: $\omega_N + \omega_{CO} + \omega_{C\alpha}$ and $\omega_N + \omega_{CO} + \omega_{C\alpha} \pm \omega_{H\alpha}$. Signal intensities in all individual input spectra hardly exceed the largest noise intensities, and in some cases the signal peaks are actually lower than the largest noise peaks (e.g., green line in Figure 3A): this would make any individual analysis (e.g., peak picking)

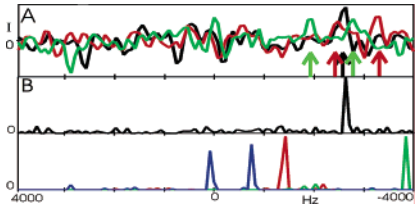


Figure 3. Analysis of Gly-Asn38 ($\omega_{HN} = 11.36$ ppm). (A) S/N in the input spectra illustrated by cross sections along the indirectly detected dimension for the spectra with $\omega_N + \omega_{CO} + \omega_{C\alpha}$ (black), $\omega_N + \omega_{CO} + \omega_{C\alpha} + \omega_{H\alpha}$ (red), and $\omega_N + \omega_{CO} + \omega_{C\alpha} - \omega_{H\alpha}$ (green). Signal peaks are identified by arrows. (B) Output shapes with $\omega_N + \omega_{CO} + \omega_{C\alpha}$ (black), $2{\cdot}\omega_{CO}$ (red), $2{\cdot}\omega_{C\alpha}$ (green), and $\omega_{H\alpha}$ (blue). Translation from Hz to ppm yields the correct chemical shifts[10] $\omega_{CO} = 173.8$, $\omega_{C\alpha} = 47.4$, $\omega_{H\alpha} = 3.68/5.03$, and $\omega_N = 119.1$ ppm (arbitrary intensity units are used on the vertical axes of all panels).

of the single spectra unreliable. However, the intensity of the real signals is in all 12 spectra clearly above zero, a coincidence that is statistically highly unlikely. Because PRODECOMP is working directly on the input spectra and not on a peak list, it detects the signals among the surrounding noise, providing a clear-cut answer in the form of the shapes shown in Figure 3B. Confidence in the reliability of this result stems from the large intensity differences in the shapes of Figure 3B between the peaks and the surrounding baseline (which can exceed the S/N in the input spectra), while in runs on spectral regions containing only noise (not shown) no detectable intensity difference among the strongest peaks is present, in particular in the $\omega_N + \omega_{CO} + \omega_{C\alpha}$ shape.

The above applications illustrate the successful combination of two general approaches in protein NMR: coupling of evolution periods and multiway decomposition. While further details (e.g., optimal fitting of shapes in eq 4 or reconstructions) will be discussed elsewhere, the examples show that PRODECOMP can be applied to various schemes of coupling evolution periods, reliably resolving situations with overlap, folding, or poor S/N. Peak lists from 1D peak picking on the shapes can be directly used for automated assignments. This method represents another NMR application of multiway decomposition,[5] in addition to the analysis of high-dimensional NOESYs, relaxation data, or nonuniformly sampled (sparse) data sets, as well as in drug discovery.

Acknowledgment. This project was supported by the Swedish Research Council (VR Grant 621-2003-4048) and the Swedish NMR Centre.

Supporting Information Available: Application of eq 4 to the spectra, matrix notation for convolutions, complete spectral input for examples 2 and 3, and the computational procedure. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Freeman, R.; Kupce, E. *Concepts Magn. Reson.* **2004**, *23A*, 63−75.
(2) Malmodin, D.; Billeter, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *45*, 109−129.
(3) Levitt, M. H. *Spin Dynamics*; Wiley: Chichester, 2001.
(4) Smilde, A.; Bro, R.; Geladi, P. *Multi-way Analysis*; Wiley: Chichester, 2004.
(5) Orekhov, V. Y.; Ibraghimov, I. V.; Billeter, M. *J. Biomol. NMR* **2001**, *20*, 49−60.
(6) See Supporting Information.
(7) Folland, G. B. *Fourier analysis and its applications*; Brooks/Cole: Pacific Grove, CA, 1992.
(8) Kupce, E.; Freeman, R. *J. Magn. Reson.* **2005**, *173*, 317−321.
(9) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125*, 1385−1393.
(10) Leckner, J. Ph.D. Dissertation, Chalmers University of Technology, Göteborg, Sweden, 2001.
(11) Malmodin, D.; Billeter, M. *Magn. Reson.* **2005**, *176*, 47−53.

JA0545822