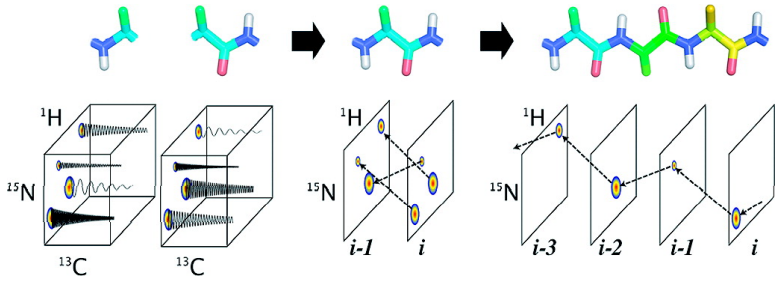
Communication

# Hyperdimensional Protein NMR Spectroscopy in Peptide-Sequence Space

Ewen Lescop, and Bernhard Brutscher

*J. Am. Chem. Soc.*, **2007**, 129 (39), 11916-11917• DOI: 10.1021/ja0751577 • Publication Date (Web): 11 September 2007






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 5 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 09/11/2007

# Hyperdimensional Protein NMR Spectroscopy in Peptide-Sequence Space

Ewen Lescop and Bernhard Brutscher*

*Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF, 41 rue Jules Horowitz, F-38027 Grenoble Cedex 1 France*

Received July 11, 2007; E-mail: bernhard.brutscher@ibs.fr

To extract useful information on protein structure and dynamics from NMR spectra, the observed NMR frequencies need first to be assigned to individual nuclear spins in the protein. Correlating all nuclear spins sequentially along the peptide sequence in a single experiment would be the ideal way of performing NMR resonance assignment. In practice however, even in the most favorable cases spin relaxation-induced signal loss imposes a limit of less than ~10 nuclear spins that can be correlated by NMR pulse sequences. In addition, standard multidimensional (nD) NMR methods require measurement times that increase by about 2 orders of magnitude per additional spectral dimension. Therefore, nD NMR beyond $n = 3$ or 4 is generally impractical using standard methods. Each of the K cross-peaks detected in a 3D (or 4D) spectrum provides correlated frequency information of 3 (or 4) nuclear spins. Such a spectrum can thus be considered as the sum of K projections of the unique correlation peak detected in a virtual hyperdimensional protein NMR spectrum. Protein resonance assignment refers to the problem of reconstructing this hyperdimensional spectrum from a limited number of summed projection spectra. Here, we show that large parts of the spectral space formed by the amide $^1$H and $^{15}$N backbone nuclei of the protein can be reconstructed from a small set of 3D correlation experiments using a correlation-based reconstruction algorithm (COBRA). The COBRA method, closely related to the recently introduced concepts of covariance[1] and hyperdimensional[2] NMR spectroscopy, provides a powerful new tool for automated fast protein resonance assignment.

The COBRA method is based on pairs of unidirectional 3D H−N−C experiments correlating a $^{13}$C nucleus ($C^\alpha$, $C^\beta$, or C′) with either the amide $^1$H and $^{15}$N of the same (*intraresidue* correlation) or the following residue (*sequential* correlation) (Figure 1). After Fourier transformation (FT) along the $^1$H and $^{15}$N dimensions, the two data sets are combined to a 4D spectrum by computing correlation coefficients between $^{13}$C traces extracted from the two data sets. The spectral intensities are given by the real part of the correlation coefficient. Signal intensity close to 1 results from similar frequencies in the two $^{13}$C traces, indicative of two potentially neighboring amide groups forming a bipeptide (Figure 1). To enhance the frequency discrimination capabilities of COBRA, we apply a weighting function to the complex correlation coefficients using an adjustable phase parameter $\phi_0$. Degenerate $^{13}$C resonances will result in ambiguities in the identification of neighboring amide groups (false peaks). This is a common feature of reconstruction algorithms that build the higher dimensional spectral space from a few projection spectra.[3] To resolve these ambiguities and remove the false cross-peaks from the 4D spectrum additional projection spectra are required. In our case, a second pair of unidirectional H−N−C experiments is performed with a different $^{13}$C nuclear spin (among $C^\alpha$, $C^\beta$, C′), and the intensities in the reconstructed 4D spectrum are given by the product of the phase-weighted correlation coefficients calculated from different pairs of 3D spectra. More details about COBRA are provided in the Supporting Information.

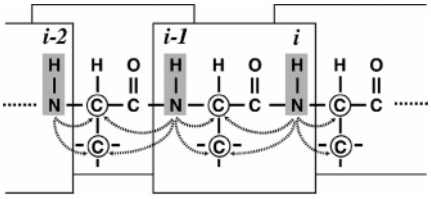



***Figure 1.*** Hyperdimensional protein NMR spectroscopy in peptide sequence space. COBRA first correlates neighboring $^1$H and $^{15}$N amide resonance frequencies within the same bipeptide based on pairs of unidirectional H−N−C experiments. In a second reconstruction step, the correlations within bipeptides are extended to longer protein fragments.

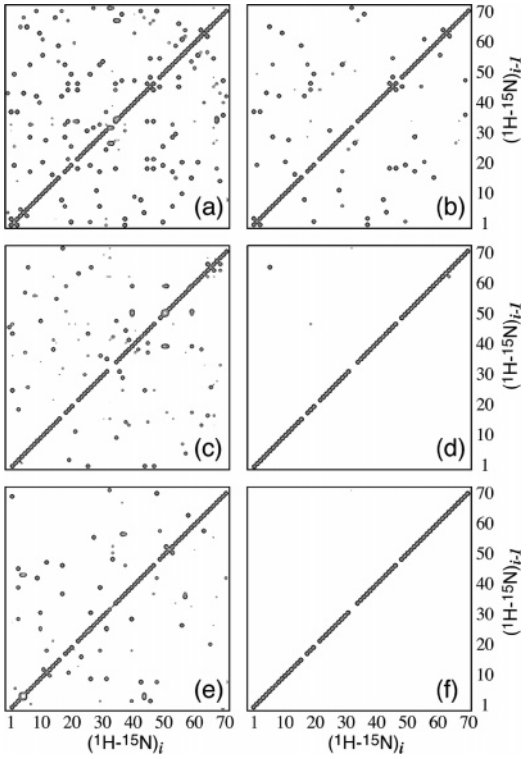



***Figure 2.*** Two-dimensional $i/(i-1)$ correlation maps of reconstructed 4D ubiquitin spectra using different experimental data sets: intraresidue and sequential H−N−CA with (a) $t_1^0 = 0$ ms, $t_1^{max} = 3$ ms, (b) $t_1^0 = 0$ ms, $t_1^{max} = 11.7$ ms, and (c) $t_1^0 = 25$ ms, $t_1^\Delta = 3$ ms, $t_1^{max} = 28$ ms; (d) combination of data sets from panels a and c; (e) intraresidue and sequential H−N−CB with $t_1^0 = 0$ ms, $t_1^{max} = 3$ ms; (f) combination of data sets from panels c and e. The COBRA phase parameter was set to $\phi_0 = 15°$. Because of the ($^1$H-$^{15}$N) ordering, cross-peaks are only expected along the shifted diagonal, correlating the frequency pairs $j$ on the horizontal axis and $j-1$ on the vertical axis.

We have tested the performance of COBRA for the protein ubiquitin (76 residues) using a set of 3D BEST-type[4] sequential


**10.1021/ja0751577 CCC: $37.00** 

and intraresidue H−N−CA and H−N−CB spectra that have been recorded in an experimental time of only a few hours (Supporting Information, Table 1). Pairs of 3D data sets were correlated via the $^{13}C$ time-domain signals as explained above. Figure 2 shows the resulting 4D spectra as simplified 2D representations correlating ($^1H$-$^{15}N$) frequency pairs, characteristic of single ubiquitin residues. In Figure 2 the 70 ($^1H$-$^{15}N$) pairs have been numbered according to the order of the corresponding residues in the ubiquitin sequence. Therefore only cross-peaks detected along the shifted diagonal in these i/i-1 correlation maps correspond to true correlations, whereas all other peaks are reconstruction artifacts. $C^\alpha$ only based COBRA reconstructions are shown in Figure 2a−d. In analogy to FT data processing, the number of false peaks decreases with increased sampling time $t_1^{max}$ corresponding to a higher spectral resolution in the $^{13}C$ dimension (Figure 2a,b). Shifting the sampled data points by $t_1^0$ along the time domain also changes the spectral appearance. The result of targeted sampling,[5] setting $t_1^0 \approx 1/J_{CC}$ for optimized sensitivity in the presence of scalar coupling evolution ($\cos(\pi J t_1^0) \approx -1$), is shown in Figure 2c. Most false cross-peaks present in the spectrum of Figure 2a are absent in the spectrum of Figure 2c, while new reconstruction artifacts appear. This observation is explained by the complementary frequency discrimination capabilities of COBRA for standard and time-shifted data sampling (Supporting Information Figure S4). Therefore, the combined sampling of a few data points at the beginning of the time-domain signal, and a few more shifted in time, suppresses most of the false cross-peaks in the spectrum (Figure 2d), while retaining short overall acquisition times. The few resulting false peaks correlate residues with $C^\alpha$ resonance frequencies that differ by less than 2 Hz. To solve remaining ambiguities arising from degenerate $C^\alpha$ chemical shifts, a second (and eventually third) pair of 3D H−N−C experiments is required. For ubiquitin, an unambiguous $i/(i-1)$ correlation map (without false peak) is obtained when combining $C^\alpha$ and $C^\beta$-based correlation information (Figure 2f). Let us emphasize here that the spectrum shown in Figure 2f is the result of automated processing of four 3D data sets recorded in less than 2 h.

Once an unambiguous $i/(i-1)$ correlation spectrum ($M_2$) is obtained, the information contained in these data allows reconstruction of the hyperdimensional spectral space formed by all amide $^1H$ and $^{15}N$ nuclei of the protein backbone, or more precisely within a protein fragment. Indeed, a proline residue or a missing correlation peak in the NMR spectra splits the peptide sequence into fragments of different size. For ubiquitin five such fragments are obtained comprising 17 (Q2−E18), 4 (S20−I23), 11 (N25−G35), 14 (D39−D52), and 23 (R54−G76) residues, respectively. In a second COBRA step, 2$k$-dimensional spectra $M_k$ can be reconstructed in an iterative manner by computing the matrix elements $M_k(\omega_k, \omega_{k-1}, ..., \omega_1) = M_2(\omega_k, \omega_{k-1}) \times M_2(\omega_{k-1}, \omega_{k-2}) \times ... \times M_2(\omega_2, \omega_1)$ with $\omega$ representing the resonance frequencies of a $^1H$-$^{15}N$ pair. The coordinates of the individual cross-peaks in the 2$k$-D hyperdimensional spectra directly yield the correlated frequency information of each combination of $k$ sequentially connected residues ($i$, $i-1$, ..., $i-k+1$) in the peptide sequence. Figure 3b shows 4D $i/(i-k+1)$ projections of the reconstructed 2$k$-D hyperdimensional spectra $M_k$ for different $k$ values. These spectra are again represented as 2D maps correlating the ($^1H$-$^{15}N$) frequencies of residues $i$ and $(i-k+1)$. No further reconstruction is possible when the $M_k$ spectrum contains only one unique cross-peak that correlates all residues within the (longest) protein fragment. For ubiquitin this situation is reached for $k = 23$ corresponding to a 46 dimensional NMR spectrum (Figure 3b).

So far we have not discussed the effect of degenerate ($^1H$-$^{15}N$) frequency pairs on the outcome of the COBRA reconstruction.

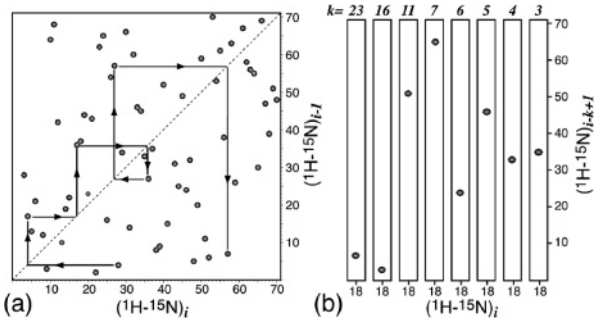



***Figure 3.*** Reconstruction of hyperdimensional spectral space. (a) Same $i/(i-1)$ correlation map as Figure 2f, but with different ordering of the ($^1H$-$^{15}N$) frequency pairs. Ubiquitin bipeptides are correlated by cross-peaks forming a characteristic directional pattern (illustrated by lines and arrows). (b) Strip plots of $i/(i-k+1)$ correlation maps correlating the $^1H$-$^{15}N$ frequency pair 18 (corresponding to residue G76) with the $(k-1)$th preceding residue in the peptide sequence. The cross-peak detected in the $k = 23$ projection of a 46-dimensional NMR spectrum correlates the first and last residues of the 23-residues fragment R54 to G76 of ubiquitin.

Unresolved $^1H$-$^{15}N$ correlations for different protein residues will result in false peaks in the $M_2$ correlation map, and as a consequence give rise to additional (false) peaks in the reconstructed hyperdimensional $M_k$ spectra. Therefore degenerate ($^1H$-$^{15}N$) frequency pairs should be excluded from the second COBRA analysis step, resulting in the reconstruction of shorter protein fragments. Alternatively, the degeneracy may be solved experimentally, for example, by recording 4D H−N−CO−C correlation spectra instead of the 3D H−N−C experiments performed here for ubiquitin. We have recently shown that such 4D spectra can be recorded in a few hours using ASCOM and BEST optimization.[6]

In summary, we have introduced COBRA, a statistical method based on the computation of phase-weighted correlation coefficients providing correlated chemical shift information from pairs of unidirectional NMR correlation experiments. A hyperdimensional NMR spectrum is computed that correlates all amide $^1H$ and $^{15}N$ nuclear spins within a same protein fragment directly from the raw NMR data without any user intervention. Nonuniform targeted sampling schemes may be used advantageously to increase spectral resolution while keeping short overall acquisition times. COBRA reconstruction thus presents a convenient new way for fast and automated sequential resonance assignment of proteins. Depending on the size of the protein and the quality of the NMR spectra, different amounts of data are required for unambiguous NMR assignment. The short computation times of only a few seconds for COBRA processing, make this method also particularly well suited for real-time estimation of data quality and completeness during NMR data acquisition.

**Supporting Information Available:** Table containing acquisition parameters used for the BEST experiments and additional information on the COBRA algorithm. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Brüschweiler, R.; Zhang, F. L. *J. Chem. Phys.* **2004**, *120*, 5253−5260. Zhang, F. L.; Brüschweiler, R. *J. Am. Chem. Soc.* **2004**, *126*, 13180−13181. Hu, B. W.; Zhou, P.; Noda, I.; Zhao, G. Z. *Anal. Chem.* **2005**, *77*, 7534−7538.
(2) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2006**, *128*, 6020−6021.
(3) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2004**, *126*, 6429−6440.
(4) Schanda, P.; Van Melckebeke, H.; Brutscher, B. *J. Am. Chem. Soc.* **2006**, *128*, 9042−9043. Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2007**, *129*, 2756−2757.
(5) Jaravine, V. A.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2006**, *128*, 13421−13426.
(6) Lescop, E.; Schanda, P.; Rasia, R.; Brutscher, B. *J. Am. Chem. Soc.* **2007**, *129*, 2756−2757.

JA0751577