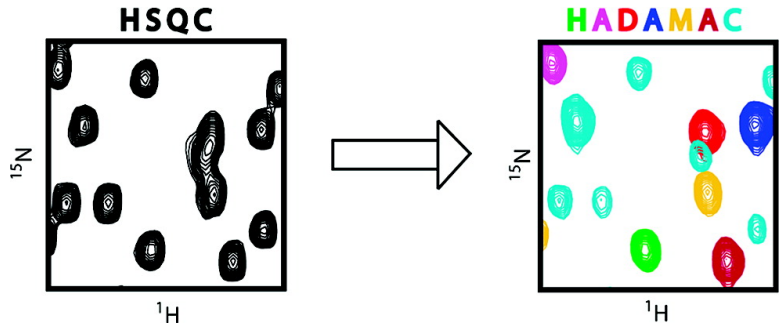
Communication

Hadamard Amino-Acid-Type Edited NMR Experiment for Fast Protein Resonance Assignment

Ewen Lescop, Rodolfo Rasia, and Bernhard Brutscher

J. Am. Chem. Soc., **2008**, 130 (15), 5014-5015 • DOI: 10.1021/ja800914h • Publication Date (Web): 18 March 2008



More About This Article

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

- Supporting Information
- 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 03/18/2008

Hadamard Amino-Acid-Type Edited NMR Experiment for Fast Protein Resonance Assignment

Ewen Lescop, Rodolfo Rasia, and Bernhard Brutscher*
IBS, Institut de Biologie Structurale Jean-Pierre Ebel, 41 rue Jules Horowitz, F-38027 Grenoble, France

Received November 23, 2007; E-mail: bernhard.brutscher@ibs.fr

NMR characterization of protein structure, dynamics, and interaction surfaces requires resonance assignment of all relevant nuclei. This initial step is commonly achieved by the extraction of sequential connectivity and ^{13}C chemical-shift-derived amino-acid-type information from a set of multidimensional NMR spectra. More discriminative, alternative strategies have been proposed in the past for amino-acid-type identification, based either on sign encoding in 3D correlation spectra[1] or on the selection of amino-acid-type specific coherence transfer pathways.[2] Here we present an original Hadamard-based sign-encoding scheme to differentiate among seven amino acid groups from a single experiment that can be performed in a very short acquisition time of only ~30 min as demonstrated for several small proteins. Each $^1H-^{15}N$ correlation peak is unambiguously assigned to one of these seven groups, thus reducing the assignment ambiguity from 20 possible amino acid types to an average of ~5. We demonstrate that the experiment ideally complements the COBRA method[3] introduced recently for correlating ^1H and ^{15}N frequencies within peptide fragments in a hyperdimensional NMR spectrum.[4] Time-optimized NMR data acquisition and analysis will enhance the potential of NMR in the context of structural proteomics, and for the study of lifetime-limited molecular systems.

Hadamard encoding in protein NMR spectroscopy has been proposed as an alternative to standard time-domain sampling methods for speeding up multidimensional data acquisition[5a] and for amino-acid-type editing of methyl groups.[5b] In 2D Hadamard spectroscopy, the frequencies of nuclei S are encoded in the NMR spectrum of the correlated nuclei I via a sign modulation of the detected NMR signal. The experiment is repeated N times using sign encoding of the different S-spin frequency bands according to a Hadamard matrix of order N.[5c] Here, we propose a sign-encoding scheme that combines band-selective frequency manipulation with NMR pulse techniques exploiting spin-coupling topologies to differentiate between amino acid types. The 20 protein amino acid side chains form 7 topology classes (Figure 1a) with respect to the number of hydrogen atoms attached to the C^α and C^β carbons and to the number and type (aliphatic, aromatic, or carbonyl) of carbons at the γ position. Appropriate NMR techniques induce a specific sign inversion of the detected signals originating from specific amino acid bands, based on their different spin-coupling networks. In practice, only odd and even H and C multiplicities can be differentiated via the experimental schemes used here. Therefore, topology classes 2 and 3 cannot be distinguished and form the single amino acid band Val-Ile-Ala. Because of their well-separated $^{13}C^\beta$ frequency range (Figure 1b), a band-selective inversion pulse can be used to specifically change the sign of NMR signals from Ser and Thr residues, resulting in a total of seven amino acid bands for Hadamard encoding (see Figure 2, and Tables S1 and S2).

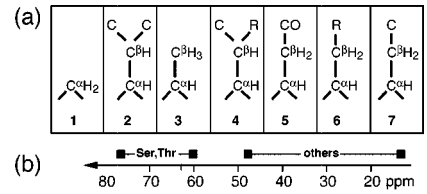


Figure 1. (a) Amino acid topology classes defined by different C−C and C−H bond multiplicities. The carbonyl (CO), hydroxyl/sulfhydril/aromatic (R), and aliphatic (C) groups at the γ position are distinguished because of their distinct NMR chemical shift ranges. The different classes correspond to Gly (class 1), Val, Ile (2), Ala (3), Thr (4), Asn, Asp (5), Phe, Tyr, Trp, His, Cys, Ser (6), and Arg, Glu, Lys, Pro, Gln, Met and Leu (7) side chains. (b) Amino-acid-type dependent $^{13}C^\beta$ chemical shift ranges.

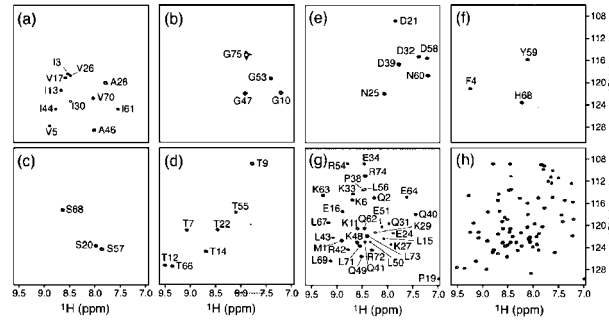


Figure 2. (a−g) HADAMAC ubiquitin spectra (central region) recorded in 20 min acquisition time. The cross peaks, labeled according to the previous residue in the ubiquitin sequence, correspond to the amino acid bands of (a) Val-Ile-Ala, (b) Gly, (c) Ser, (d) Thr, (e) Asn-Asp, (f) *Arom* (Phe, His, Trp, Tyr)-Cys, and (g) *Rest* (Arg, Glu, Lys, Pro, Gln, Met, Leu). Only positive contours are plotted for spectra (a−e). Graphs (f) and (g) correspond to the positive and negative intensity contours of the same HADAMAC subspectrum. (h) Standard $^1H-^{15}N$ correlation spectrum.

The HADAMAC, for HADamard-encoded AMino-ACid-type-editing, experiment is shown in Figure S1. The pulse sequence has been adapted from the DEPT-$H^{\alpha\beta}C^{\alpha\beta}$(CO)NH experiment.[6] A set of eight 2D $^1H-^{15}N$ correlation spectra are recorded where the signs of the cross peaks change from one spectrum to another according to a Hadamard matrix of order 8, reflecting the amino acid type of the preceding residue in the protein sequence. Different spin manipulation techniques are combined to achieve the Hadamard encoding: (i) A DEPT sequence inverts the sign of NMR signals for $C^\beta H/C^\beta H_3$ relative to $C^\beta H_2$ multiplicities; (ii) evolution under C−H scalar coupling allows sign inversion for $C^\alpha H_2$ (Gly residues) relative to $C^\alpha H$ moieties; (iii) the signals of Asn and Asp residues are inverted by evolution under $^{13}C^\beta-^{13}C'$ scalar coupling; (iv) a selective C^β inversion pulse changes the sign for Ser and Thr residues. These four spin manipulations are combined to realize the eight-step Hadamard encoding (Table S2). The spectrum corresponding to a single amino acid band is calculated as a linear combination of the eight data sets with the relative sign taken from the appropriate column in the Hadamard encoding matrix (*Had-*

10.1021/ja800914h CCC: $40.75

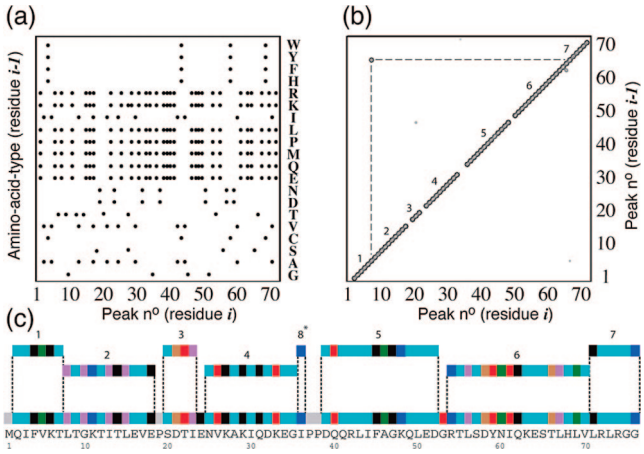


Figure 3. (a) HADAMAC-derived correlation plot for ubiquitin indicating the possible amino acid types of the preceding residue for each amide $^{1}H-^{15}N$ correlation peak (residue). (b) COBRA connectivity map obtained from a pair of intraresidue and sequential BEST[7] H−N−CA correlation spectra recorded for ubiquitin in an overall time of 50 min using targeted time-domain sampling in the ^{13}C dimension.[3] When considering only the most intense COBRA peaks, seven unambiguous peptide fragments are obtained (labeled 1−7). In panel (c), these fragments are represented as color-coded arrays of square blocks, where each block represents one residue, and the color code indicates the possible amino acid types according to the HADAMAC experiment. On the bottom of panel (c), the ubiquitin sequence is represented using the same amino-acid-type color code.

amard transformation). A total of six amino-acid-type edited $^{1}H-^{15}N$ spectra are obtained. The *Arom-Cys* and *Rest* bands are not separated by this Hadamard encoding scheme. Since they only differ by the presence (*Rest*) or absence (*Arom-Cys*) of one aliphatic carbon at the C^{γ} position (Table S1), they are detected in the same subspectrum but with opposite signs. The HADAMAC experiment was tested on four small globular proteins of 70−80 residues differing in sample concentration (0.7–1.9 mM) and amino-acid-type composition. Acquisition times of ∼30 min proved to be sufficient on a 600 MHz spectrometer equipped with a cryogenic triple-resonance probe to yield high-quality HADAMAC data sets for each of these proteins. Figure 2a−g shows the amino-acid-type edited $^{1}H-^{15}N$ correlation spectra obtained for ubiquitin (76 aa), as well as a reference spectrum recorded under identical conditions (h). The HADAMAC spectra recorded for the grx (79 aa), VsvC (73 aa), and Hyl1–2 (77 aa) proteins are provided in Figure S2. A clean, although not perfect, separation of cross peaks from different amino-acid-type bands is observed. Residual peaks detected in the "wrong" HADAMAC subspectrum are due to the sensitivity of the sign inversion techniques used for Hadamard encoding to variations in scalar coupling constants and to pulse imperfections. The intensities of these spectral-editing artifacts, however, do not exceed 13% (worst case) of the intensity of the correlation peak detected in the correct amino-acid-type spectrum (Table S3). The HADAMAC subspectrum with the highest intensity at a given $^{1}H,^{15}N$ frequency pair thus yields unambiguous amino-acid-type discrimination among the seven amino acid bands (types) for all residues with resolved (or only partially overlapping) $^{1}H-^{15}N$ correlations. The HADAMAC spectral information can be translated into a 2D map correlating $^{1}H-^{15}N$ frequency pairs (residues) and amino acid types as shown in Figure 3a for ubiquitin. A significant reduction in the average number of possible amino-acid-type assignments per $^{1}H,^{15}N$ frequency pair is achieved. The correlation peak density is reduced from 100 to 22.1 ± 0.5% for the proteins ubiquitin, grx, VsvC, and Hyl1–2. This amino-acid-type discrimination capability makes HADAMAC a powerful additional tool for fast *de novo* sequential protein backbone resonance assignment. HADAMAC also presents an attractive alternative to 3D triple-resonance experiments to quickly identify and assign shifted correlation peaks for chemical shift mapping purposes after a site-directed protein mutation or ligand binding.

The HADAMAC experiment is ideally suited to complement the sequential connectivity information obtained from pairs of 3D triple-resonance experiments (e.g., HN(CO)CA and intra-HNCA). We have recently introduced the COBRA method[3] for the automated analysis of such NMR data. The output of COBRA is a connectivity map correlating the $^{1}H-^{15}N$ chemical shifts of residues $i-1$ and i. The complementarity of the two methods, COBRA and HADAMAC, for fast protein resonance assignment is demonstrated in Figure 3 for ubiquitin. A total of seven unambiguous peptide fragments are identified from COBRA analysis of NMR data recorded in only 50 min (Figure 3a,c) using fast-pulsing methods[7] and time-optimized sampling schemes. By matching the fragments color-coded according to the HADAMAC-derived amino acid type with the ubiquitin sequence (Figure 3c), every fragment can be unambiguously and correctly assigned to a single position in the peptide sequence. We have further performed a blind test of this approach on the RNA-binding domain of Hyl1 (Hyl1–2), a protein involved in micro-RNA processing in plants, for which no resonance assignment was available prior to this study. Despite a sample concentration of only 0.7 mM and a heterogeneous distribution of $^{1}H-^{15}N$ correlation peak intensities (Figure S3a), complete resonance assignment was obtained (Figure S4) from a combined COBRA and HADAMAC analysis of NMR data recorded in a total acquisition time of only 1 h and 40 min. Our results demonstrate that NMR assignments of small globular proteins can be achieved from sparse NMR data acquired in a few hours or less using appropriate fast acquisition and analysis tools as currently developed by a number of NMR groups.[8] Here we have proposed an original assignment strategy based on the combined analysis of COBRA connectivity information with HADAMAC amino-acid-type discrimination capabilities. This method is well suited for automation, as a single $^{1}H,^{15}N$ frequency list, but no further higher dimensional spectral peak picking is required. Automated on-the-fly analysis of stepwise collected NMR data will allow time-optimized protein resonance assignment, and it will be of special importance for the study of molecular systems with a limited lifetime in the NMR sample tube.

Acknowledgment. Financial support from the CNRS, CEA, UJF, ANR (JCJC05-0077), and the European Commission (I3, Contract No. 026145) is acknowledged. We thank I. Ayala, P. Gans (IBS Grenoble), N. Rouhier (Nancy), M. Jamin, and E. Ribeiro (IVMS Grenoble) for making protein samples available.

Supporting Information Available: Pulse sequence, Hadamard tables, and acquisition parameters for the HADAMAC experiment, as well as additional spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Rios, C. B.; Feng, W. Q.; Tashiro, M.; Shang, Z. G.; Montelione, G. T. *J. Biomol. NMR* **1996**, *8*, 345–350, and included references.
(2) Schubert, M.; Oschkinat, H.; Schmieder, P. *J. Magn. Reson* **2001**, *153*, 186–192, and included references.
(3) Lescop, E.; Brutscher, B. *J. Am. Chem. Soc.* **2007**, *129*, 11916–11917.
(4) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2006**, *128*, 6020–6021.
(5) (a) Kupce, E.; Nishida, T.; Freeman, R. *Prog. NMR Spectrosc.* **2003**, *42*, 95–122. (b) Brutscher, B. *J. Biomol. NMR* **2004**, *29*, 57–64. (c) Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2006**, *178*, 334–339. (d) Van Melckebeke, H.; Simorre, J. P.; Brutscher, B. *J. Am. Chem. Soc.* **2004**, *126*, 9584–959. (e) Hadamard, J. *Bull. Sci. Math.* **1893**, *17*, 240–248.
(6) Brutscher, B. *J. Magn. Reson.* **2004**, *167*, 178–184.
(7) Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2007**, *187*, 163–9.
(8) (a) Fiorito, F.; Hiller, S.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2006**, *35*, 27–3. (b) Malmodin, D.; Billeter, M. *Prog. NMR Spectrosc.* **2005**, *46*, 109–129.

JA800914H