

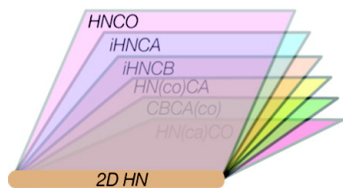
Article

Hyperdimensional NMR Spectroscopy with Nonlinear Sampling

Victor A. Jaravine, Anastasia V. Zhuravleva, Perttu Permi, Ilgis Ibraghimov, and Vladislav Yu. Orekhov

J. Am. Chem. Soc., **2008**, 130 (12), 3927-3936 • DOI: 10.1021/ja077282o

Downloaded from <http://pubs.acs.org> on February 8, 2009

$$8D H = \sum F^N \otimes F^H \otimes F^{Co1} \otimes F^{Ca} \otimes F^{Cb} \otimes F^{Ca1} \otimes F^{Ca1Cb1} \otimes F^{CoCo1}$$


3D HNCO	=	$\sum F^N \otimes F^H$	$\otimes F^{Co1}$
3D iHNCA	=	$\sum F^N \otimes F^H$	$\otimes F^{Ca}$
3D iHN(ca)CB	=	$\sum F^N \otimes F^H$	$\otimes F^{Cb}$
3D HN(co)CA	=	$\sum F^N \otimes F^H$	$\otimes F^{Ca1}$
3D CBCA(co)	=	$\sum F^N \otimes F^H$	$\otimes F^{Ca1Cb1}$
3D HN(ca)CO	=	$\sum F^N \otimes F^H$	$\otimes F^{CoCo1}$

More About This Article

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

- Supporting Information
- Links to the 3 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](#)



ACS Publications
High quality. High impact.

Hyperdimensional NMR Spectroscopy with Nonlinear Sampling

Victor A. Jaravine,[†] Anastasia V. Zhuravleva,^{†,‡} Perttu Permi,[‡] Ilgis Ibraghimov,[§] and Vladislav Yu. Orekhov^{*,†,||}

Swedish NMR Centre, Göteborg University, Box 465, 40530 Göteborg, Sweden, NMR Laboratory, Structural Biology and Biophysics Program, Institute of Biotechnology, University of Helsinki, P.O. Box 65, Helsinki FIN-00014, Finland, Saarbrücken University, Mathematical Department, Saarbrücken D-66041, Germany, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115

Received September 20, 2007; E-mail: orov@nmr.gu.se

Abstract: An approach is described for joint interleaved recording, real-time processing, and analysis of NMR data sets. The method employs multidimensional decomposition to find common information in a set of conventional triple-resonance spectra recorded in the nonlinear sampling mode, and builds a model of hyperdimensional (HD) spectrum. While preserving sensitivity per unit of measurement time and allowing for maximal spectral resolution, the approach reduces data collection time on average by 2 orders of magnitude compared to the conventional method. The 7–10 dimensional HD spectrum, which is represented as a set of deconvoluted 1D vectors, is easy to handle and amenable for automated analysis. The method is exemplified by automated assignment for two protein systems of low and high spectral complexity: ubiquitin (globular, 8 kDa) and ζ_{cyt} (naturally disordered, 13 kDa). The collection and backbone assignment of the data sets are achieved in real time after approximately 1 and 10 h, respectively. The approach removes the most critical time bottlenecks in data acquisition and analysis. Thus, it can significantly increase the value of NMR spectroscopy in structural biology, for example, in high-throughput structural genomics applications.

Introduction

Over the past two decades, NMR spectroscopy has evolved as one of the prime techniques for protein structure determination at the atomic level and for characterizing proteins, protein–ligand complexes, or nucleic acids. X-ray crystallography and NMR are two biophysical methods for determining protein structures¹ that have proven to be the most useful in structural genomics, which aims to ascribe a three-dimensional protein structure to each gene product of the human and other genomes. Because these methods rely on distinctly different physical principles and experimental procedures, crystallography and NMR are highly complementary for high-throughput (HTP) structure determination; both are important to ultimate project success.^{2–4} When used to determine structure in the pipeline with highly automated and parallelized target selection and

protein expression, contemporary NMR often represents a major time bottleneck.⁴ Weeks of data collection using an expensive NMR spectrometer are required for every protein target. Measurements are followed by data analysis, which is at least as lengthy and is usually performed manually. The NMR community^{5,6} has devoted significant attention to the need to save spectrometer time and to automate the analysis steps.

With modern sensitive NMR spectrometer hardware, the duration of a multidimensional experiment is determined by the time needed for one measurement and the number of measurements. Both factors are targeted in ongoing efforts to speed up the experiments. Recording individual data points can be accelerated by reducing the delay between consecutive measurements^{7–12} or by parallel acquisition as in single scan NMR.¹³

[†] Göteborg University.

[‡] University of Helsinki.

[§] Saarbrücken University.

^{||} Harvard Medical School.

^{*} Current address: Department of Biochemistry & Molecular Biology 1232P, University of Massachusetts, Amherst, Massachusetts 01003.

(1) Brünger, A. T. *Nat. Struct. Biol.* **1997**, *4* (Suppl), 862–865.
(2) Yee, A. A.; Savchenko, A.; Ignachenko, A.; Lukin, J.; Xu, X. H.; Skarina, T.; Evdokimova, E.; Liu, C. S.; Semesi, A.; Guido, V.; Edwards, A. M.; Arrowsmith, C. H. *J. Am. Chem. Soc.* **2005**, *127* (47), 16512–16517.
(3) Snyder, D. A., et al. *J. Am. Chem. Soc.* **2005**, *127* (47), 16505–16511.
(4) Yee, A.; Gutmanas, A.; Arrowsmith, C. H. *Curr. Opin. Struct. Biol.* **2006**, *16* (5), 611–617.

(5) Wuthrich, K. Nobel Lecture, Stockholm, 2002.

(6) Moseley, H. N. B.; Montelione, G. T. *Curr. Opin. Struct. Biol.* **1999**, *9* (5), 635–642.

(7) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of NMR in one and two dimensions*; Clarendon: Oxford, 1987.

(8) Ross, A.; Salzman, M.; Senn, H. *J. Biomol. NMR* **1997**, *10* (4), 389–396.

(9) Pervushin, K.; Vogeli, B.; Eletsky, A. *J. Am. Chem. Soc.* **2002**, *124* (43), 12898–12902.

(10) Atreya, H. S.; Szyperski, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (26), 9642–9647.

(11) Schanda, P.; Brutscher, B. *J. Am. Chem. Soc.* **2005**, *127* (22), 8014–8015.

(12) Schanda, P.; Van Melckebeke, H.; Brutscher, B. *J. Am. Chem. Soc.* **2006**, *128* (28), 9042–9043.

(13) Frydman, L.; Scherf, T.; Lupulescu, A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (25), 15858–15862.

Novel sampling schemes reduce the number of data points without losing essential information. In particular, in G-matrix Fourier transform (GFT)¹⁴ and *projection reconstruction*,^{15,16} only specific spectral projections of lower dimensionality are used to obtain information present in complete multidimensional spectra. As an alternative, nonlinear sampling (NLS), which is also referred to as nonuniform or *sparse sampling*, allows reconstruction of a complete spectrum from only a small number of optimally selected experimental data points.¹⁷ By taking into account prior information about signal properties, a NLS schedule can be optimized for maximum spectral sensitivity and resolution.¹⁷ The approach requires nontraditional signal processing schemes such as nonlinear Fourier transform,^{18–20} maximum entropy,^{21–23} or multidimensional decomposition (MDD),^{24–27}

In addition to rapid data collection, novel sampling techniques offer a framework for highly efficient automated spectra analysis by making possible spectroscopy of high dimensionality. For example, to ultimately simplify the problem of signal assignments, one may choose an experiment that contains only one signal, whose position in multidimensional space encodes frequencies in all spin systems throughout the protein amino acid sequence. Although the hundreds of dimensions that would be needed for such an experiment are out of reach for contemporary NMR methodology, clear progress in this area has been recently demonstrated. Five- and six-dimensional experiments for backbone resonance assignment that make use of reduced dimensionality and the GFT approach to sample the huge spectral space during a reasonable measurement time were reported.^{10,28} Although carefully and optimally designed, these experiments represent a limit for physical transfer of the magnetization through the long chain of spin–spin interactions and thus are often seen as an unaffordable compromise when sensitivity is in demand.²⁸ It is generally accepted that extending NMR spectroscopy beyond four dimensions is impractical. An alternative approach to high-dimensional NMR is found in the concept of hyperdimensional (HD) spectroscopy.^{29,30} According to its original definition, the HD technique derives all possible direct and indirect correlation spectra from a limited set of low-

dimensional measurements. In other words, a spectrum of high dimensionality or any of its projections is obtained using a suite of sensitive 2D to 4D experiments. Generally, HD spectroscopy opens an avenue for obtaining spectra of any dimensionality and for the optimization of data acquisition.

Here, we demonstrate that a model of HD spectrum can be built using MDD of a standard set of triple-resonance NMR experiments, which are rapidly recorded using NLS. The spectrum represented by a number of HD MDD components is amenable to fully automated real-time backbone assignments performed concurrently with data collection.

Theory

HD Spectroscopy. HD spectroscopy was introduced by Kupce and Freeman.²⁹ In the following section, we point to an important distinction between conventional multidimensional spectroscopy and HD spectroscopy. The former is implemented in a pulse sequence, thereby directly exploiting physical correlations between the spin coherences in a single NMR experiment. HD spectroscopy is essentially a joint recording and processing of spectral sets in the frame of a single mathematical model.

An HD spectrum shows both direct and indirect correlations within a spin system. The indirect correlations are inferred from direct correlations using a set of low-dimensional experiments, which are essentially the orthogonal projections of the HD spectrum. For example, a pair of conventional experiments with direct correlations $A \rightarrow B$ and $A \rightarrow C$ can be used to construct a hyperspectrum with additional indirect correlation $B \rightarrow C$. The idea of obtaining indirect correlations from direct ones is reminiscent of the “covariance spectroscopy” of Zhang and Brüschweiler.³¹

It should be emphasized that, in general, a hyperspectrum may contain less information than a genuine spectrum of the same dimensionality. The latter, however, is typically not possible to obtain due to the practical limitations and thus cannot be used for comparison. In the above example, the difference appears if signal A overlaps with another signal A' , which, in turn, has direct correlations with B' and C' . The overlap may result in spurious correlations $B \rightarrow C'$ and $B' \rightarrow C$ that appear in the HD spectrum but not in the spectrum exploiting direct coherence transfer $B \rightarrow C$. In other words, while a conventional N -dimensional spectrum resolves signals overlapped in $N-1$ dimensions, resolution in the hyperspectrum corresponds to that in the original set of low-dimensional spectra. Currently, protein HD spectroscopy is based on sets of conventional 3D and 4D experiments, where the problem of overlap is largely alleviated. Indeed, these experiments have been used so far to obtain signal assignments for most of the protein systems studied by NMR, including disordered and large proteins,^{32–36} which are the cases with the most crowded spectra. The HD approach is general and can be used with sensitive 2–4D spectra as well as with 5–6Ds^{10,37} that are optimized for ultimate resolution. This approach does not require developing new, specialized NMR

- (14) Szyperki, T.; Yeh, D. C.; Sukumaran, D. K.; Moseley, H. N. B.; Montelione, G. T. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (12), 8009–8014.
- (15) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2003**, *125* (46), 13958–13959.
- (16) Coggins, B. E.; Venters, R. A.; Zhou, P. *J. Am. Chem. Soc.* **2004**, *126* (4), 1000–1001.
- (17) Barna, J. C. J.; Laue, E. D.; Mayger, M. R.; Skilling, J.; Worrall, S. J. P. *J. Magn. Reson.* **1987**, *73* (1), 69–77.
- (18) Marion, D. *J. Biomol. NMR* **2005**, *32* (2), 141–150.
- (19) Kazimierzczuk, K.; Kozminski, W.; Zhukov, I. *J. Magn. Reson.* **2006**, *179* (2), 323–328.
- (20) Kazimierzczuk, K.; Zawadzka, A.; Kozminski, W.; Zhukov, I. *J. Biomol. NMR* **2006**, *36* (3), 157–168.
- (21) Hoch, J. C.; Stern, A. S. *NMR data processing*; Wiley-Liss: New York, 1996.
- (22) Stern, A. S.; Li, K. B.; Hoch, J. C. *J. Am. Chem. Soc.* **2002**, *124* (9), 1982–1993.
- (23) Rovnyak, D.; Frueh, D. P.; Sastry, M.; Sun, Z. Y. J.; Stern, A. S.; Hoch, J. C.; Wagner, G. *J. Magn. Reson.* **2004**, *170* (1), 15–21.
- (24) Orekhov, V. Y.; Ibraghimov, I.; Billeter, M. *J. Biomol. NMR* **2003**, *27* (2), 165–173.
- (25) Tugarinov, V.; Kay, L. E.; Ibraghimov, I.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2005**, *127* (8), 2767–2775.
- (26) Luan, T.; Jaravine, V.; Yee, A.; Arrowsmith, C. H.; Orekhov, V. Y. *J. Biomol. NMR* **2005**, *33* (1), 1–14.
- (27) Jaravine, V.; Ibraghimov, I.; Orekhov, V. Y. *Nat. Methods* **2006**, *3* (8), 605–607.
- (28) Fiorito, F.; Hiller, S.; Wider, G.; Wuthrich, K. *J. Biomol. NMR* **2006**, *35* (1), 27–37.
- (29) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2006**, *128* (18), 6020–6021.
- (30) Lescop, E.; Brutscher, B. *J. Am. Chem. Soc.* **2007**, *129* (39), 11916–11917.

- (31) Zhang, F. L.; Brüschweiler, R. *J. Am. Chem. Soc.* **2004**, *126* (41), 13180–13181.
- (32) Yao, J.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **1997**, *419* (2–3), 285–289.
- (33) Zhang, O. W.; Forman-Kay, J. D.; Shortle, D.; Kay, L. E. *J. Biomol. NMR* **1997**, *9* (2), 181–200.
- (34) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J. Mol. Biol.* **1999**, *288* (4), 705–723.
- (35) Liu, A. Z.; Riek, R.; Wider, G.; von Schroetter, C.; Zahn, R.; Wuthrich, K. *J. Biomol. NMR* **2000**, *16* (2), 127–138.
- (36) Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124* (34), 10025–10035.

pulse sequences. The experiments in a set, however, should be performed in compatible way to ensure equality in the common parts. In particular, to ensure good matches between the line shapes from different experiments, the spectral width, the carrier positions, and the number of time increments should be the same for the corresponding common dimensions. It is also helpful to run the experiments in the interleaved fashion to avoid differences in the sample condition between the experiments in the set. Another reason for the interleaved acquisition relates to the incremental data collection with real-time analysis, which is used in this work and described below.

Accumulation of half a dozen 3–4D spectra using conventional uniform sampling is time-consuming. Previously, the measurement time was reduced in HD spectroscopy by employing the projection reconstruction technique²⁹ or suite of fast band-selective excitation short-transient (BEST)-type experiments.¹² In this work, full information from the experiments is rapidly harvested using optimal nonlinear sampling.

The MDD Model of the HD Spectrum. The model of multidimensional decomposition (MDD) assumes that all essential features of an M -dimensional matrix can be described as the sum of a small number of tensor products of one-dimensional vectors. MDD has been used in a variety of fields as a tool for data analysis and signal processing since the early 1970s under various names, such as parallel factor analysis, canonical decomposition, and three-way decomposition.³⁸ Previously, the MDD model was successfully used for the reconstruction of individual multidimensional experiments recorded using either nonlinear sampling^{24–26,39} or projections.⁴⁰ In this work, we generalize the approach for an HD spectrum \mathbf{H} , where data are collected as a suite of experiments \mathbf{S} of lower dimensionalities. Unlike a single experiment, in which dimensionality is limited by the efficiencies of all magnetization transfer steps, the HD spectrum has no limits on the number of dimensions.

When applied to NMR spectra, the MDD can be formulated as follows. Given a spectrum \mathbf{H} with sizes N_m of its M dimensions ($m = 1 \dots M$) and elements H_{n_1, n_2, \dots, n_M} , find scalar numbers βa and normalized vectors $\beta \mathbf{F}^m$, with elements $\beta F^{m-}(n_m)$ ($n_m = 1 \dots N_m$), such that the following norm becomes minimal:⁴¹

$$|\mathbf{G} \cdot [\mathbf{H} - \sum_{\beta} (\beta a \beta \mathbf{F}^1 \otimes \beta \mathbf{F}^2 \dots \otimes \beta \mathbf{F}^M)]|^2 + \lambda \sum_{\beta} (\beta a)^2 \quad (1)$$

Here, the symbol \otimes denotes the tensor product operation. In the case of NLS, only a fraction of the elements in \mathbf{H} is measured, and the matrix \mathbf{G} , which contains elements $g_{n_1, n_2, \dots, n_M} \in [0 \text{ or } 1]$, indicates the absence or presence of a particular data point. Accordingly, the symbol \cdot describes the element-wise multiplication of matrices. The last term represents Tikhonov regularization with the parameter λ , which can be used to improve the convergence of the MDD algorithm. The regularization also ensures stability of the minimization algo-

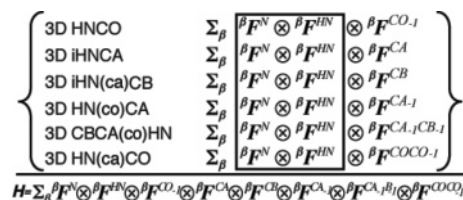


Figure 1. Construction of the HD spectrum from six 3D experiments. The amide nitrogen and proton shapes are the same for all 3Ds and serve to bind HD components over the experiments. Each experiment contributes one indirect shape that constitutes a dimension in the HD spectrum.

rithm in case of a very sparse \mathbf{G} matrix. Equation 1 with sparse matrix \mathbf{G} is most useful for describing a nonlinearly sampled NMR signal in the time domain. It is worth noting that an alternative equation was recently published, which describes the spectrum represented by frequency domain projections.⁴⁰ The summation index β runs over the number of components used for decomposition. The range of this index depends on the type of spectrum. For example, for the 3D HNCO spectrum, it is roughly equal to the number of protein backbone amide groups.

In the time domain, the experiments \mathbf{S}_{α} , which are enumerated by index α , can be considered as orthogonal cross sections of the full HD spectrum \mathbf{H} . Assuming that all the experiments have the same sensitivity, eq 1 for the M -dimensional HD spectrum can be solved using all \mathbf{S}_{α} measurements simultaneously, with matrix \mathbf{G} taking care of the lower dimensionalities and NLS in individual spectra. In the real case, we need to account for different efficiencies of magnetization transfers and signal intensities in the set of experiments, that is, to find scaling factors for the components in the individual experiments. In the frame of the MDD model, this is done most naturally using eq 1, with an additional dimension for spectra weighting. Thus, an M -dimensional hyperspectrum can be constructed from an appropriate set of low-dimensional spectra using $(M + 1)$ -dimensional decomposition. Below, we describe a simplified practical algorithm for building the model of HD spectra and illustrate it using examples of real-time experiments.

The HD Algorithm. Equation 1 gives a general theoretical framework for constructing a HD spectrum from a number of experiments using a joint MDD model. In this work, we use a simplified two-stage algorithm that is derived from eq 1. The algorithm is equivalent to the direct application of eq 1 with the assumption that experiments in the set share several dimensions (amide H^{N} and N) and one of the experiments (HNCO in this work) significantly exceeds all others in sensitivity. In the general case, for example, if some experiments do not have HN dimensions, we could envisage alternative multistage protocols or direct use of eq 1.

The most reliable, albeit computationally demanding, algorithm for solving the minimization problem defined by eq 1 is based on alternating least-squares (ALS) iterations.^{39,42} In this algorithm, at each iteration, amplitudes βa and shapes $\beta \mathbf{F}^m$ along one of the dimensions are optimized simultaneously for all components, while the shapes for other dimensions are fixed. This naturally allows fixation of the shapes for selected dimensions, which is achieved by skipping updates along these dimensions. On the other hand, when updating a particular dimension of a HD spectrum, the $(M + 1)$ -dimensional matrix \mathbf{G} selects only measurements from the experiments with the corresponding dimension. Thus, Figure 1 exemplifies a special case where each experiment from the suite contributes only one

(37) Hiller, S.; Fiorito, F.; Wuthrich, K.; Wider, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (31), 10876–10881.

(38) Kruskal, J. B. Rank, decomposition, and uniqueness for 3-way and N-way arrays. In *Multway data analysis*; Coppi, R., Bolasco, S., Eds.; North-Holland Elsevier Science Pub. Co.: Amsterdam, New York, 1989.

(39) Orekhov, V. Y.; Ibraghimov, I. V.; Billeter, M. *J. Biomol. NMR* **2001**, *20* (1), 49–60.

(40) Malmodin, D.; Billeter, M. *J. Am. Chem. Soc.* **2005**, *127* (39), 13486–13487.

(41) Ibraghimov, I. *Numer. Linear Algebra Appl.* **2002**, *9* (6–7), 551–565.

additional dimension, while the remaining dimensions are shared by all experiments. In this case, fixing shapes of the r -shared dimensions splits the $(M + 1)$ -dimensional decomposition problem into $M-r$ independent MDD problems for individual $(r + 1)$ -dimensional experiments:

$$|\mathbf{G}_\alpha \bullet [\mathbf{S}_\alpha - \sum_\beta (\beta a_\alpha \beta \mathbf{F}^1 \otimes \beta \mathbf{F}^2 \dots \otimes \beta \mathbf{F}^r \otimes \beta \mathbf{F}^\alpha)]|^2 + \lambda \sum_\beta (\beta a_\alpha)^2 \quad (2)$$

where the MDD shapes $\beta \mathbf{F}^i$ ($i = 1 \dots r$) are the same in all spectra and fixed, while $\beta \mathbf{F}^\alpha$ denotes a shape which is specific for a particular 3D experiment α ; \mathbf{G}_α is a subset of \mathbf{G} corresponding to the NLS schedule of the experiment; and βa_α is the amplitude adjusted for the individual experiment and thus corresponds to the shape of the *weighting* dimension. The common shapes $\beta \mathbf{F}^i$ are obtained in a decomposition of the most sensitive *base* experiment.

In this work, without losing generality of the consideration and merely as an example, we focus on a set of eight triple-resonance experiments, typically used in structural studies of proteins. The first six experiments are used for backbone assignment (Figure 1) including HNCO, HN(co)CA, CBCA-(co)HN, HN(ca)CO, and two intrasidues experiments iHNCA and iHN(ca)CB,⁴³ while the last two serve to assign side chain protons and calculate spatial structure: H(cco)NH and ¹⁵N NOESY-HSQC.

For residue i and its preceding residue $i - 1$, HD spectrum \mathbf{H} correlates frequencies of any of the backbone nuclei H^{N}_i , N_i , C^{A}_i , C^{B}_i , $\text{C}^{\text{A}}_{i-1}$, $\text{C}^{\text{B}}_{i-1}$, C^{O}_i , and $\text{C}^{\text{O}}_{i-1}$ plus protons relayed by the last two experiments. \mathbf{H} has up to $M = 10$ dimensions and is constructed from eight 3D experiments sharing $r = 2$ dimensions, namely of amide proton and nitrogen. Accordingly, the shapes $\beta \mathbf{F}^{\text{HN}}$, $\beta \mathbf{F}^{\text{N}}$, $\beta \mathbf{F}^{\text{CA}}$, $\beta \mathbf{F}^{\text{CB}}$, $\beta \mathbf{F}^{\text{CO}-1}$, and $\beta \mathbf{F}^{\text{CA}-1}$ contain a single peak; $\beta \mathbf{F}^{\text{COCO}-1}$ and $\beta \mathbf{F}^{\text{CA}-1\text{CB}-1}$ contain two peaks corresponding to frequencies of C^{O}_i , $\text{C}^{\text{O}}_{i-1}$ and $\text{C}^{\text{A}}_{i-1}$, $\text{C}^{\text{B}}_{i-1}$, respectively; $\beta \mathbf{F}^{\text{TOCSY}-1}$ and $\beta \mathbf{F}^{\text{NOESY}}$ represent groups of peaks from H(cco)NH and ¹⁵N NOESY-HSQC dimensions.

To solve the problem posed by eq 2, we first determine the amide proton and nitrogen shapes from the base HNCO experiment, which is the most sensitive in the set. These shapes are then considered as known and not adjusted in the MDD calculations for the remaining five experiments. Fixation of shapes $\beta \mathbf{F}^{\text{NH}}(t2)$ and $\beta \mathbf{F}^{\text{HN}}(\omega3)$ during the decomposition reduces the number of adjustable parameters in the MDD model and thus saves additional spectrometer time relative to the original R-MDD analysis.²⁷ Total saving of the measurement time relative to the commonly used uniform sampling is given by factor ϕ (derivation of the equation in the Supporting Information)

$$\phi = f N_p \log_2(N_\alpha) / (4 N_\alpha N_{\text{NH}}) \quad (3)$$

where N_α and N_{NH} are the adjusted sizes of carbon and fixed nitrogen indirect spectral dimensions, respectively; N_p is number of amides in a protein; and $f = 0.3$ is the empiric factor, which is the same for all experiments.

Skipping updates for two of the three dimensions improves the convergence of minimization in eq 2 and dramatically speeds

up the calculations. With several shapes defined, the algorithm effectively performs filtering of the input data, that is, largely disregards spectral artifacts that may appear in the areas beyond the defined H^{N} and N^{H} positions. However, the most important aspect is that the two fixed shapes tag each HD component by associating it with a given $\text{H}^{\text{N}}-\text{N}^{\text{H}}$ pair throughout the set of 3D experiments and thus collect all peaks of a spin system into one component.

Figure 2 exemplifies the output of the HD calculations (a MDD model of HD spectrum). It is completely defined by the set of line shapes for all dimensions (Figure 2a–i). From this representation, any region or projection of HD spectrum \mathbf{H} can be reconstructed. Figure 2j, for example, depicts the reconstruction of a 2D projection, which correlates α -carbons of two consecutive residues in the amino acid sequence of ubiquitin. The 2D to 4D projections can be input to existing programs for spectral analysis. On the other hand, the 1D representation of the HD spectrum is natural for the method and suitable for signal identification. In this way, we directly obtain a peak list for the whole HD spectrum without the need to identify signals in the individual 3D spectra or in higher/lower dimensional projections and without needing to match peaks from different data sets. Straightforward and robust peak-picking in 1D vectors and inherent peak grouping by $n\text{D}$ components simplify the analysis and are beneficial for the automation of signal assignments.

Experimental Section

NMR Data Measurement and Processing. Table 1 details the performed NMR experiments. The proteins were uniformly ¹³C and ¹⁵N labeled: ubiquitin⁴⁴ and 13 kDa naturally disordered ζ_{cyt} (cytoplasmic domain of the T cell receptor ζ subunit).^{45–49} ζ_{cyt} was expressed and purified as previously described.⁴⁷ The samples were in 10% D₂O/90% H₂O: ubiquitin, 1.7 mM, pH 4.6; and ζ_{cyt} , 0.5 mM, 20 mM phosphate, pH 6.7. The NMR experiments on ubiquitin (ζ_{cyt}) were performed at 25 °C (15 °C) on Varian spectrometers with Larmor frequencies of 800 MHz (900 MHz) equipped with room temperature (cold) pulsed-field gradient triple resonance probes. The sets of triple-resonance 3D experiments were performed using standard gradient sensitivity enhanced pulse sequences from the BioPack library (Varian Inc.) in the NLS mode (Table S1 in the Supporting Information). HNCO, HN(co)CA, CBCA(co)NH, and “intra” iHNCA and iHN(ca)CB spectra⁴³ were recorded for both proteins. In addition, for ζ_{cyt} , a HN(ca)CO experiment was performed. All experiments were recorded in an interleaved order as a sequence of incremental steps,⁵⁰ except for HNCO and ¹H-¹⁵N NOESY-HSQC and H(cco)NH spectra for ubiquitin, which were not included in the real-time scheme.

At each step, every experiment from the suite has a short run with a newly generated NLS schedule. The schedules at different steps follow the same probability density distribution and differ only in the value of the seed for the random number generator. The distributions are calculated using the matched acquisition principle^{17,23,26,39,51} with parameters presented in Table S1 in the Supporting Information. By

(44) Wang, A. C.; Grzesiek, S.; Tschudin, R.; Lodi, P. J.; Bax, A. *J. Biomol. NMR* **1995**, *5* (4), 376–382.

(45) Weissenhorn, W.; Eck, M. J.; Harrison, S. C.; Wiley, D. C. *Eur. J. Biochem.* **1996**, *238* (2), 440–445.

(46) Aivazian, D.; Stern, L. *J. Nat. Struct. Biol.* **2000**, *7* (11), 1023–1026.

(47) Sigalov, A.; Aivazian, D.; Stern, L. *Biochemistry* **2004**, *43* (7), 2049–2061.

(48) Sigalov, A. B.; Zhuravleva, A. V.; Orekhov, V. Y. *Biochimie* **2007**, *89* (3), 419–421.

(49) Duchardt, E.; Sigalov, A. B.; Aivazian, D.; Stern, L. J.; Schwalbe, H. *ChemBioChem* **2007**, *8* (7), 820–827.

(50) Jaravine, V. A.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2006**, *128* (41), 13421–13426.

(42) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38* (2), 149–171.

(43) Tossavainen, H.; Permi, P. *J. Magn. Reson.* **2004**, *170* (2), 244–251.

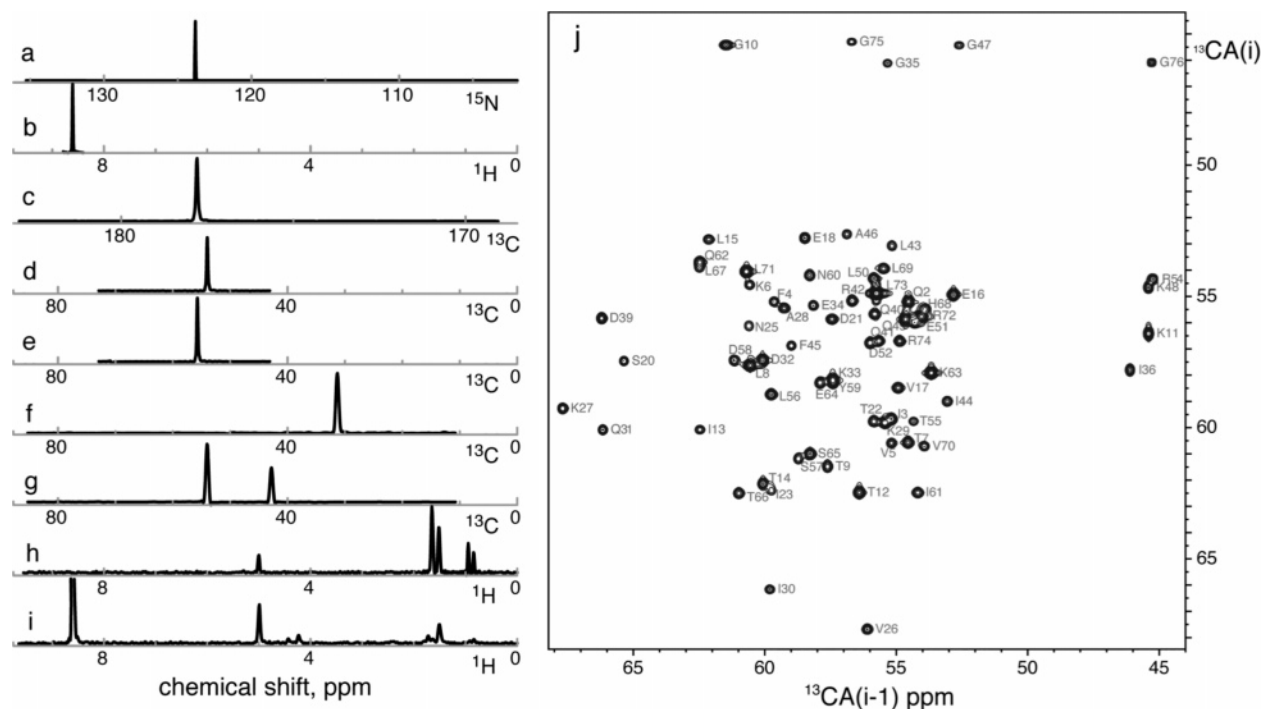


Figure 2. Example of MDD HD analysis for ubiquitin: (a–i) HD component (R72); the 1D line shapes are from the 3D experiments (a–c) HNCO, (d) HNCO, (e) iHNCA, (f) iHN(ca)CB, (g) CBCA(co)HN, (h) H(cco)NH, and (i) ^{15}N NOESY-HSQC; (j) 2D CA(i)-CA(i-1) projection from the HD spectrum; the correlations are labeled with ubiquitin sequence assignments; data of step 30.

Table 1. Experiments on the Proteins and Results of Spectral Data Analysis

	ubiquitin (8 kDa, globular)	ζ_{cyt} (13 kDa, disordered)
	NMR spectra	
^1H spectrometer frequency, ^1H MHz	800	900
number of H^{N} signals detected (sequence) ^a	70 (72)	116 (107)
3D experiments ^b	HNCO ^c , HNCO, CA, CBCA(co)NH, iHNCA, iHN(ca)CB, (NOESY-HSQC, H(cco)NH) ^d	HNCO ^c , HNCO, CA, CBCA(co)NH, iHNCA, iHN(ca)CB, HN(ca)CO
	Calculations & analysis	
measurement time per step, h	0.144	1.21
average CPU time ^e per one step, h	0.1	0.2
MDD, peak-picking, assignment time points “90% peaks” ^f	0.01, 0.08, 0.01	0.07, 0.10, 0.01
TA step number	- ^c , 2, 3, 2, 4	- ^c , 3, 5, 2, 2, 7
ϕ experiment, %	-, 0.4, 1.0, 0.5, 1.3	-, 0.25, 0.8, 0.8, 0.5, 2.7
ϕ theory (eq 3), %	-, 0.8, 1.5, 0.7, 0.7	-, 0.3, 0.7, 1.0, 0.5, 3.0

^a The number of amide signals observed in the amide area of 3D HNCO. The figures in parenthesis are the numbers of backbone H^{N} groups in the amino acid sequences. For ubiquitin, the signals of M1, E24, and G53 were not identified; for ζ_{cyt} , G1 and S2 were not identified, and there were about a dozen minors. ^b The list of experiment names. For each protein, parameters for individual spectra are given as comma separated lists with the order corresponding to the names on the list of experiments. ^c The HNCO spectra were recorded in one step as 5% NLS data prior to recording the other spectra, and they are excluded from the TA procedure and time calculations. ^d H(cco)NH and NOESY-HSQC experiments for ubiquitin were recorded (25% NLS) after the backbone set; they were not used for the automated assignment, and thus, they are not counted in the measurement and processing timing for the assignment. The parameters for these experiments are given in parentheses. ^e The computation times are on a Linux workstation with one Dual Core Opteron processor (2×2.0 GHz). ^f The point during acquisition when at least 90% of the peaks are obtained. The numbers are given in each spectrum: the step number in the TA schedule, experimental and theoretically predicted ϕ , and values as fractions of the full data matrix.

matching probability distribution to the signal decay in the time domain, the NLS schedules are optimized for better sensitivity. The number of entries in the schedule does not change with each step but ranges from 8 to 34 for individual experiments (see Table 1). The numbers define the time resolution of the incremental procedure (Figures 3–5) and partitioning between different experiments in the set. The steps were

short enough to closely monitor the completion of the assignment and the stability of the solution around and after this point.

Real-Time Assignment with Targeted Acquisition (TA). The main principles of TA are described in our previous publication.⁵⁰ In short, TA is stepwise incremental data acquisition using a NLS schedule, which is performed concurrently with MDD spectra processing and automated analysis. The approach provides time optimization in reaching predefined project targets. While originally TA was applied to a single experiment, here the concept is extended to incremental collection and real-time processing of several experiments. Since each

(51) Hyberts, S. G.; Heffron, G. J.; Tarragona, N. G.; Solanky, K.; Edmonds, K. A.; Luthardt, H.; Fejzo, J.; Chorev, M.; Aktas, H.; Colson, K.; Falchuk, K. H.; Halperin, J. A.; Wagner, G. *J. Am. Chem. Soc.* **2007**, *129* (16), 5108–5116.

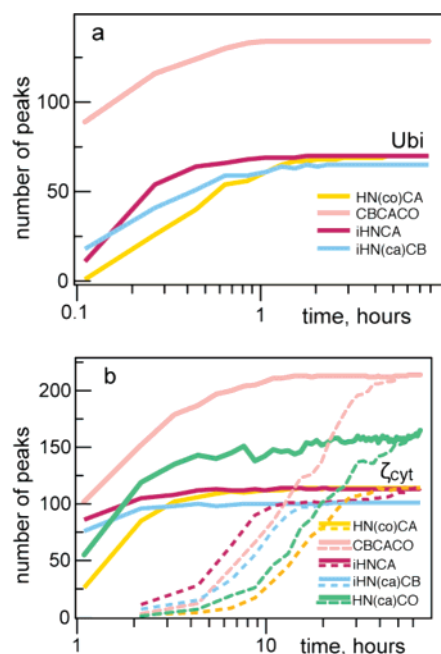


Figure 3. Numbers of peaks detected using the HD approach (solid lines) as a function of acquisition time (step) for (a) ubiquitin and (b) ζ_{cyt} . The color codes are as follows: (yellow) HN(cco)CA; (pink) CBCA(co)NH; (magenta) iHNCA; (blue) iHN(ca)CB; and (green) HN(ca)CO. The dashed lines in panel (b) give the number of signals obtained from the same data but without the HD approach, i.e., using MDD reconstructions and peak-picking of individual 3Ds.

experiment carries an essential part of information for the real-time assignment procedure, the experiments must be recorded in parallel, that is, in the interleaved way. Direct outputs of the procedure are spectral quality scores for each step. In this work, the scores reflect the number of detected peaks as well as completeness and precision of the assignments. In HTP applications, the experiments can be stopped when the scores reach target values, for example, 95% of assignment, thus dealing with the problem of incomplete and redundant data collection. In this methodological work, however, the experiments continued far beyond the time allocation needed for the assignments. This was done to demonstrate the clear plateau for the scores and to confirm that no new information was obtained after a certain point. This point defines the amount of measurement time needed to reach the targets and is reported in this study for each of the proteins as a formal time needed to obtain signal assignments. The timing can serve as a measure of efficiency for different protocols of data collection and analysis. Thus, here the TA scores on the number of detected peaks were used to qualitatively measure the specific advantage of the HD analysis in comparison to the approach, where individual experiments of the set are processed separately.

Here, we give a short description of the TA assignment procedure. The detailed description of the procedure for the data collection and analysis can be found in the Supporting Information. The data set accumulated by the end of each TA step is subjected to the HD MDD decomposition (eq 2). Gradually, a more refined MDD model of the HD spectrum is constructed as more data points are measured. For each amide group, the processing gives a list of hyperdimensional components, which are composed of the 1D shapes. The time domain shapes are then Fourier transformed, and the positions of the signals are determined using *nmrPipe* software.⁵² The chemical shift values obtained for each amide represent an unassigned spin system. Sequence-specific assignments for backbone atoms are obtained automatically

from the list of spin systems using the program AutoAssign.⁵³ This program is invoked for each step using a batch macro with the adapted execution schedule. The accuracy and precision of the assignment obtained at each step are quantified using the routine *compare_bmrbl* from “Assignment Validation Suite”.⁵⁴ The assignments of the last step were submitted to the BMRB database. (Data deposition note: the backbone assignments for ubiquitin and ζ_{cyt} are deposited as BMRB entries 15410 and 15409, respectively; to be released upon publishing of the paper.)

Results and Discussion

In this work, the concept of HD spectroscopy, which was originally introduced for radial sampling,²⁹ is demonstrated in combination with NLS, MDD signal processing, and targeted acquisition. Experimental results and theoretical considerations show that the HD approach allows significant reduction of measurement time compared to non-HD analysis. In the TA procedure, we construct gradually improved models of the high-resolution HD spectra. The signals in the spectra are described completely by the sets of the HD components grouped by amide frequencies (Figure 2a–i). While illustrating the method’s effectiveness and robustness, we obtained complete backbone resonance assignments for the two proteins in a very short time (Figure 5). The two systems represent low and high spectral complexity. Ubiquitin (8 kDa) is a small globular protein. The cytoplasmic domain of the T cell receptor ζ_{cyt} (13 kDa) represents a naturally disordered protein.^{45–49}

Interleaved Recording of HD Spectra and Real-Time Signal Detection. A standard triple-resonance set (Figure 1) of short experiments is acquired using nonlinear sampling schedules and analyzed in equal steps. A varying number of transients and time increments define time allocations for the individual experiments (Table S1 in the Supporting Information). These are chosen to roughly compensate for the differences in the sensitivities of the spectra. Ideally, a comparable percentage of signals is detected in the spectra at all steps to ensure balanced input to the automated assignment program. For example, the least sensitive HN(ca)CO experiment was recorded with 4 times more transients than other experiments in the ζ_{cyt} set. Figure 2 exemplifies a HD component (R72 of ubiquitin, step 30), which is the direct output of the MDD calculations. The sets of components for all amides constitute complete representations of the hyperspectra. Any correlations between the spins or even the full 9D (10D) hyperspectrum for ubiquitin (ζ_{cyt}) can be reconstructed from the components. As an example, a 2D reconstruction of the CA(*i*)-CA(*i* – 1) projection for ubiquitin is given in Figure 2j. The multidimensional reconstructions, however, were unnecessary for our analysis, since the signals were easily and reliably detected in the 1D shapes. Furthermore, MDD efficiently deconvolutes peaks belonging to different components.³⁹ Detection of signals in the reconstructed *n*D projections is more difficult, because signals from different components may overlap.

The first seven (eight) shapes for ubiquitin (ζ_{cyt}) were obtained for each amide group at every step of the TA procedure. The peaks obtained in these shapes were used for the automated assignments. The two last NOESY and TOCSY H(cco)NH

(52) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. J. *Biomol. NMR* **1995**, *6* (3), 277–293.

(53) Moseley, H. N.; Monleon, D.; Montelione, G. T. *Methods Enzymol.* **2001**, *339*, 91–108.

(54) Moseley, H. N.; Sahota, G.; Montelione, G. T. *J. Biomol. NMR* **2004**, *28* (4), 341–355.

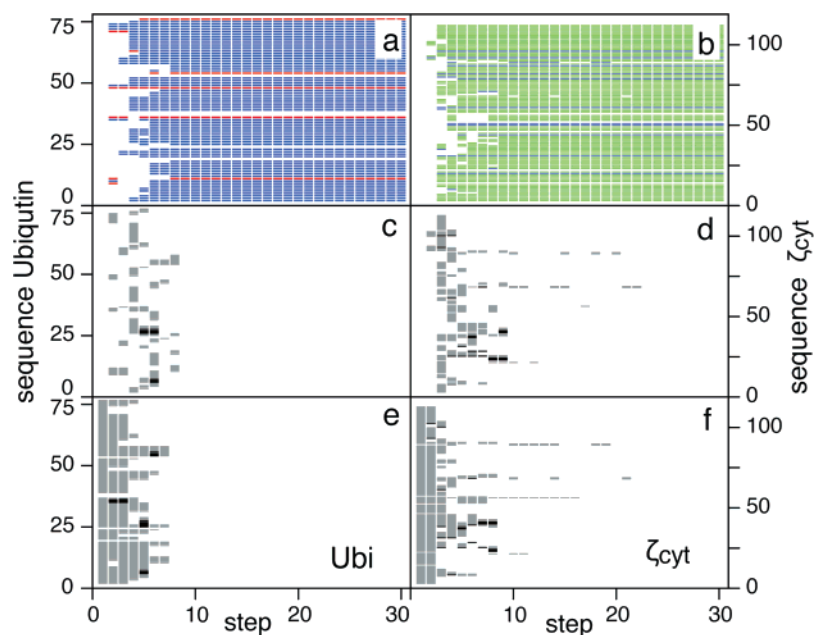


Figure 4. Content and quality of the backbone chemical shift assignment tables for ubiquitin (left) and ζ_{cyt} (right) as a function of the targeted acquisition step. The number of matching shifts (or sequential connectivities) for each residue (panels a and b) is indicated by colored dashes (red, 1; blue, 2; green, 3). The *precision* (c and d) and *accuracy* (e and f) of the assignment are defined as maximum absolute difference of the chemical shifts at current versus previous and last TA steps, respectively. A gray dash indicates that the assignment for one of the atoms for the residue is missing in one of the tables; a black dash indicates that at least one of the chemical shift differences exceeds a threshold value (0.1 ppm for ^{13}C and ^{15}N or 0.01 ppm for ^1H).

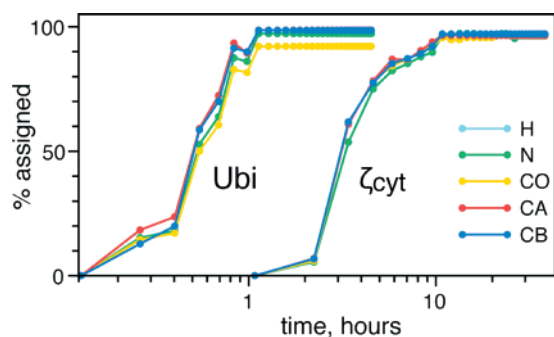


Figure 5. Progress of backbone assignments for ubiquitin (left) and ζ_{cyt} (right). The percentages of assigned backbone atoms (H, cyan; N, green; CO, yellow; CA, red; CB, blue), relative to the totals in the amino acid sequences of the proteins, are given as a function of experimental time. The statistics are derived from the assignment tables (in BMRB 3.0 format) on each step using the procedure *missing_shifts.pl* from “Assignment Validation Suite”.

shapes in Figure 2h and i illustrate the perspective of the HD approach for protein side chain assignments and structure determination. These two shapes were not used for the backbone assignment and were obtained using experiments recorded after the backbone set.

Figure 3a and b shows the direct (unassigned) results of the HD processing and peak-picking in the 1D shapes. The peak count curves are sigmoid functions of the acquisition time. Initially, a few of the strongest peaks appear, followed by detection of bulk peaks with average intensity, trailed by a few weak peaks. For ubiquitin, the high sensitivity and relatively small number of amide groups allow for detection of nearly all the expected correlations in the set of four experiments in ~ 1 h. For the ζ_{cyt} spectra, it takes 1 order of magnitude longer to reach the plateau of the peak counts. More time is needed for ζ_{cyt} due to lower sample concentration, higher molecular weight, larger number of experiments in the set, and more severe signal overlap for the disordered protein. The number of detected peaks

is a complex function of sensitivity and the ratio of the number of collected data points to the number of the MDD model parameters. We previously showed⁵⁰ that peak detection could be limited both by low sensitivity and by underdetermination of the MDD model. The 1D shapes are obtained at all steps, but with few measured data points the shapes contain too much noise and only few, if any, peaks of acceptable quality are detected. The latter situation may happen at the very first TA steps.

Saving Measurement Time. Experimentally observed time saving factors ϕ (eq 3) for individual experiments recorded for the two protein systems are presented in Table 1. The approach makes it feasible to use a small fraction of measurements. For example, 90% of the peaks in the HN(co)CA spectrum of ζ_{cyt} are detected with only 0.2% of the full uniformly sampled matrix. On average, the factor ϕ is around 1% at the level of 90% peak detection. This translates into a reduction of the total measurement times from 54 and 546 h to 0.4 and 5.2 h for ubiquitin and ζ_{cyt} , respectively. The latter times correspond to time allocations for the experiment set, when each experiment is stopped separately after 90% of peaks are detected (more details on peak monitoring are below). Note that we recorded and analyzed the HNCO spectra prior to the accumulation of the sets. Along with other necessary procedures such as temperature setup and shimming, the HNCO can be considered as a prerequisite in our procedure. Alternatively, it can be rapidly recorded⁵⁰ at the first stage, which would add the time equivalent of one or two steps of the TA procedure. Good correlation between the experimental and theoretical ϕ values (Table 1) indicates general applicability of eq 3 as well as the validity of the assumption of the sampling limiting regime used in the derivation of the equation. Note, however, that sensitivity in the ζ_{cyt} experiments was enhanced by recording two to 8 times more transients in the ζ_{cyt} spectra than in the corresponding ubiquitin spectra.

The net speedup of NMR measurements is the result of several synergetic contributions: NLS, TA, R-MDD, and HD analysis. While the former three methods were thoroughly addressed in our previous publications, it is instructive to highlight the specific contribution of the HD approach, which is the main focus of this work. Figure 3b shows that, to detect the same number of peaks, the HD analysis allows ~ 4 times less measurements in comparison to processing and analyzing of the same spectral data individually. Although, the procedures of peak identification (peak-picking) in 1D shapes and 3D spectra are inherently different, the improvements judged by comparison of the curve timings in Figure 3 are in line with our previous results. While the ϕ values of 1% and less are demonstrated in this work, those obtained previously for individual spectra were larger than 5%. The time saving factors of around 4 for the HD analysis are also predicted from the theoretical consideration (see the Supporting Information). Time saving in obtaining the resulting peak lists reflects clear improvement of the spectra due to the use of HD analysis. To the best of our knowledge, we show the shortest total measurement time for a protein backbone assignment data set, although a close result has been recently reported for ubiquitin.³⁰ Notably, the HD approach was used in both cases, indicating a significant added value of this novel type of analysis.

The demonstrated short measurement time can be appealing for HTP applications, such as structural genomics. In this work, however, it serves as a benchmark value and quantitative measure to justify the merit of the HD approach in general and, notably, characterizes our particular implementation of it. When comparing time savings of different fast techniques, it is important to look at the total measurement time needed to detect all required signals and to take into account the dimensionality of the experiments used. Generally, the higher the spectrum dimensionality, the longer the full experiment will be, and, consequently, the smaller fraction of measurement will be needed. In this work, the time saving is obtained for a set of conventional triple-resonance 3D experiments and thus can be directly compared with traditional analysis.

Computation. The original MDD algorithm is computationally demanding. However, the HD MDD procedure with fixed H^N and N shapes takes considerably less time. For example, at step 30 of the TA schedule, the computations for the HN(ca)-CO spectrum of ζ_{cyt} converged in 0.5 min (118 iterations), which should be compared with 40 min of the same calculations (2560 iterations) without the shape fixation. Although all calculations are conducted using a single modern workstation with two Dual Core CPUs, the calculation times per step are always smaller than the corresponding measurement times. Table 1 gives separate computational times for the MDD calculations, processing and analysis of shapes, and assignments. Note that most of the time is spent on shape processing and peak-picking, using routines from *nmrPipe*. Thus, we show that, with a modern computer and the HD MDD algorithm used in the current real-time implementation, the calculations do not represent a time-limiting step.

Monitoring Sequential Connectivities and Obtaining Resonance Assignments. The process of automated resonance assignment has been evolving for several years.^{6,55–58} Here, we use the external program AutoAssign53 to perform backbone assignment for the two proteins. First, we identify sequential

connectivities for individual amide groups (Figure 4a and b) of the two protein systems as a function of acquisition step (see the Experimental section). Two connectivities through C^α and C^β frequencies are rapidly obtained for the majority of ubiquitin amides. Similarly, up to three contacts via C^α , C^β , and C^O signals are established for ζ_{cyt} . A lack of sequential contacts for several residues is due to prolines (3 for ubiquitin, 6 for ζ_{cyt}) and glycines (6 for ubiquitin, 13 for ζ_{cyt}), which do not have either H^N or C^β atoms, respectively. Second, AutoAssign uses the connectivities on each step to produce resulting assignments (Figure 5). After the backbone assignment curves reach a plateau close to 100%, the acquisition is continued to verify the stability of the established solution. The progress of assignment clearly correlates with the increases of the numbers of detected peaks (Figure 3). Thus, the main and obvious factor that improves the assignment with additional measurements is the number of detected peaks. More data also improve the accuracy of the signal positions (Figure S1 in the Supporting Information) by reducing the noise level in the 1D shapes used for peak detection.

During progression of the real-time procedure, the extrapolation of the assignment curves permits estimation of the completion time. For example, 50% (95%) peak assignment levels are achieved roughly at 0.5 (1.0) h for ubiquitin and 3 (10) h for ζ_{cyt} . Thus, we can generalize that the time taken to obtain the first half multiplied by a factor of 2–3 can give an approximate time needed to assign 95% of the peaks. Note that the above measurement time estimates are on the upper side for the suggested approach, since a number of obvious optimizations were deliberately avoided in this work. This is done for simplicity of presentation and to highlight the value of the HD approach in combination with NLS and MDD analysis. In particular, the time allocation for several of the most sensitive experiments in the set can be reduced (Table 1). This allows shortening of the total spectrometer time by an additional 20–30%. Moreover, we did not use time optimized BEST experiments,¹² which would have additionally reduced the measurement time at least by half.

Extra time allocation to detect/assign any of the remaining 5% weak signals depends largely on their sensitivity. Some signals could be very weak and thus require too much recording time. Detection of the ^{13}C correlations and their correct assignments is achieved for most spin systems after step 10 (Figure 4). Longer spectra acquisition for ζ_{cyt} , up to step 24, was required for assignment of several weak HN(ca)CO correlations. The percentages of the assigned H^N , N, CO, CA, and CB correlations relative to the corresponding totals in the amino acid sequences (Figure 5) are close to 100% at about 1 and 10 h for ubiquitin and ζ_{cyt} , that is, with the exception of correlations that were not observed (Table 1).

Assignment Accuracy and Precision. In the NLS experiment, additional measurements increase the number of observed signals and improve the precision of signal positions and intensities. The effect is similar to that of conventional experiments, where the signal-to-noise ratio gradually increases with

-
- (55) Hyberts, S. G.; Wagner, G. J. *Biomol. NMR* **2003**, *26* (4), 335–344.
(56) Malmodin, D.; Papavoine, C. H.; Billeter, M. J. *Biomol. NMR* **2003**, *27* (1), 69–79.
(57) Baran, M. C.; Huang, Y. J.; Moseley, H. N.; Montelione, G. T. *Chem. Rev.* **2004**, *104* (8), 3541–3556.
(58) Masse, J. E.; Keller, R. J. *Magn. Reson.* **2005**, *174* (1), 133–51.

more transients. Besides, an increasing number of measurements can be thought of as an implementation of the jackknife/bootstrap statistical analysis,⁵⁹ which allows estimating the precision of parameters. We define the precision of the assignments as the difference between the chemical shifts obtained at two consecutive acquisition steps. In a realistic real-time procedure, the correct values of the chemical shifts are not known *a priori*, thus making it impossible to estimate the accuracy. It is instructive, however, to see the similarity of the patterns of precision (Figure 4c and d) and accuracy (Figure 4e and f). The reference assignment table was manually checked and verified against published assignments for ubiquitin⁴⁴ and ζ_{cyt} .⁴⁹ For the individual residues of the two protein systems, the errors in the chemical shifts corresponding to precision and accuracy drop below the cutoffs (0.1 ppm for any ^{13}C) at about the same acquisition step. It is worth noting that the assignment achieved at the intermediate steps is mostly correct and is revised at the following stages only for a few residues (black dashes, Figure 4c–f).

The assignment procedure using the software AutoAssign is computationally fast (Table 1). The positive results of the automated analysis are due to the reduction of combinatorial complexity in the HD MDD approach. Namely, the protocol produces peaks already grouped into spin systems defined by the amide groups. Furthermore, the NLS spectra can be recorded with maximal resolution limited only by the relaxation properties of the spin coherences and the duration of the constant-time evolution periods in the pulse sequences. High-resolution translates into low-signal overlap and high precision of the chemical shift values, which provide accurate connectivities between the spin systems. The ζ_{cyt} spectra (Table 1) were acquired with up to 200 and 224 complex data points in the ^{13}C and ^{15}N dimensions, respectively. The automated analysis of signals is further simplified by using the intraresidual versions of the backbone experiments that produce a single peak for each amide.⁶⁰ In such experiments, the accuracy of identification of a single peak is not compromised by an occasional overlap, which can be heavy for CA signals, for example, in HNCACB or HNCA experiments. It certainly helps if all experiments in a set are optimized for detection of a single amide peak; however, HD spectroscopy does not set such a requirement, and the selection of experiments for a set can be based on their higher sensitivity or on any other criteria. Finally, deconvolution of the overlapped signals in the HD MDD protocol and the high digital resolution affordable for the 1D shapes increase the reliability of frequency estimation compared to regular 3D peak-picking.

The Intrinsically Unstructured Protein and the Problem of Overlap. The heavy overlap exacerbated by homologous sequence repeats is not entirely unexpected for an unfolded protein with a small content of residual secondary structure. This is the reason why we selected such a protein system for testing the HD approach with 3D data. Most chemical shifts of ζ_{cyt} are random coil values. There is a large number of residues having close values of C^α and C^β chemical shifts, in particular, the cluster of 14 arginines, 8 lysines, 8 glutamic acids, and 9 glutamines. The consequence is that many amino acid segments have similar chemical shift patterns. The procedure *typing_de-*

generacy.pl from “Assignment Validation Suite”,⁵⁴ which reports statistics on degenerate or similar segments in a protein amino acid sequence, found degeneracy in 55 pairs of four, 27 pairs of five, 7 pairs of six, and 1 pair of eight residues. The latter pair of stretches starts at G30 and G50. Due to the high degeneracy of C^α and C^β chemical shifts for ζ_{cyt} , additional connectivities through C^O were essential and sufficient for the sequential disambiguation.

As it is stated in the theory, we could expect wrong correlations in the HD spectra in the case of severe overlap in the amide dimensions. With the used algorithm of the HD analysis, this effect could reveal itself as an “exchange” of 1D carbon shapes between the overlapped components (or spin systems). With the notations from the Theory section, it might happen that, instead of the correct correlations A-C and A'-C', the wrong ones A-C' and A'-C are observed in the HD spectrum. The effect is noise dependent and would reveal itself as changes in the C-frequencies in the individual HD components from one TA step to another. ζ_{cyt} has a 4.25 times more crowded ^1H dimension than the average in the *refDB* protein database. Thus, the H^N chemical shift dispersion of 0.16 ppm for ζ_{cyt} corresponds to the overlap complexity of a 50 kDa globular protein. In the ^1H - ^{15}N amide correlation spectrum of ζ_{cyt} , the 2D distance (^1H ppm scale) between a peak and its closest neighbor ranged from 0.005 ppm for two residue pairs (Q22, D102 and G60, G81) to 0.42 ppm with the median value of 0.03 ppm. Although, for $\sim 10\%$ of the amides, the distances between the peaks were close and, for some experiments, below the physical spectral resolution, we did not observe a correlation between the number of changes in the peak positions and the degree of overlap (Figure S1 in the Supporting Information). Most of the peaks were detected, and their positions became stable by step 10 of the TA schedule, although the final solution is reached in step 21. The overlap did not pose a problem in our analysis, because NLS enabled high spectral resolution in all the indirect dimensions and MDD provided efficient peak deconvolution.

If peaks overlap severely in all dimensions of the 3D HNCB spectrum, they are indistinguishable and can be effectively treated as one component. However, there were no 3D overlaps for any of the components for the high-resolution HNCB. Consequently, a fixed number of fully separated components identified in the 3D HNCB spectra is used for the 3D deconvolutions in all other experiments. Knowing the number of components and their H–N shapes considerably facilitates 3D decompositions in the remaining spectra. Thus, 3D-based HD analysis is largely unaffected by H–N overlaps. The TA procedure provides an internal statistical monitor for potentially problematic HD components. These are revealed directly as changes in the peak lists from one step to another (Figure S1 in the Supporting Information) and as changes in the assignment tables (Figure 4). In this demonstration, even for the crowded spectra of ζ_{cyt} , information about amide line shapes resulted in efficient deconvolution of the overlaps and unambiguous binding of signals into spin systems. For large systems with naturally broad lines and heavy overlap, we envisage the use of a suite of 4D experiments,³⁶ where components are identified in four dimensions and bound via three dimensions.

Conclusion

The dramatic speedup of NMR measurements and analysis demonstrated here is the result of several synergetic contribu-

(59) Efron, B. *The jackknife, the bootstrap, and other resampling plans*; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1982.

(60) Permi, P. *J. Biomol. NMR* **2002**, *23* (3), 201–209.

tions. NLS avoids large redundancy in the measurements while providing adequate resolution and preserving or improving sensitivity per unit of instrument time. Targeted acquisition allows for real-time optimization of time allocation for individual experiments in the data set. The HD analysis finds common information for the signals in different spectra, which allows further reduction of experimental data needed for the analysis. The HD MDD analysis provides deconvolution of overlapped signals and the possibility to work with the 1D representations of multidimensional spectra. This dramatically simplifies identification of the signals, which is important for the success of automated spectra analysis.

The HD approach for efficient handling of raw experimental data fits well into the context of recent advances in computational techniques for automated signal assignments,^{56,57} rapid protein structure determination, and macromolecular complex characterization.^{61–66} Together, the methods should enable one to determine spatial structures and characterize protein dynamics

and interactions with higher accuracy and much more rapidly than is possible at present, thereby enhancing the value of NMR spectroscopy in HTP applications such as structural genomics⁶⁷ and in conventional “hypothesis driven” structural biology projects.

Acknowledgment. This work was supported by research grants from the Swedish Foundation for Strategic Research A3 04 160d; the Swedish Research Council 2005-2951; EU FP6 LSHG-CT-2005-018988; and the Academy of Finland 106852 and by the “ARTS” award from Apple Inc. We are grateful to Dr. A. B. Sigalov for providing us with the sample of the ζ_{cyt} domain.

Supporting Information Available: Extended version of Table 1; theoretical derivation of the equation for time saving in the sampling limiting case; description of the algorithm of phase sensitive (States) detection in R-MDD; detailed description of the stepwise data acquisition and analysis protocols; figure showing the precision of carbon frequencies of the ζ_{cyt} components versus TA step; and full version of ref 3. This material is available free of charge via the Internet at <http://pubs.acs.org>

- (61) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615–9620.
(62) Lopez-Mendez, B.; Guntert, P. *J. Am. Chem. Soc.* **2006**, *128* (40), 13112–13122.
(63) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15* (5), 563–570.
(64) Bonvin, A.; Boelens, R.; Kaptein, R. *Curr. Opin. Chem. Biol.* **2005**, *9* (5), 501–508.
(65) Kontaxis, G.; Delaglio, F.; Bax, A. Molecular fragment replacement approach to protein structure determination by chemical shift and dipolar homology database mining. In *NMR of Biological Macromolecules, Part C*; James, T. L. Ed.; Elsevier Academic Press: San Diego, 2005; Vol. 394, p 42.

JA077282O

- (66) Grishaev, A.; Llinas, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (10), 6707–6712.
(67) Chandonia, J. M.; Brenner, S. E. *Science* **2006**, *311* (5759), 347–351.