J|A|C|S
ARTICLES
Published on Web 09/27/2006

Targeted Acquisition for Real-Time NMR Spectroscopy

Victor A. Jaravine and Vladislav Yu. Orekhov*

Contribution from the Swedish NMR Centre at Gothenburg University, Box 465, 40530 Gothenburg, Sweden

Received March 29, 2006; Revised Manuscript Received August 15, 2006; E-mail: Vladislav.orekhov@nmr.gu.se

Abstract: A target-oriented approach for the acquisition of information in biomolecular NMR spectroscopy is being developed. This approach combines concurrent data accumulation, processing, and monitoring of spectral quality. Real-time estimation of parameters allows acquisition to be stopped when results are complete and have a specified precision. The technique is based on multidimensional decomposition, which can process incomplete data. An incremental nonuniform sampling scheme ensures the optimization of resolution sensitivity. To validate this method, 3D HNCO spectra of three biomolecular systems (8 kDa ubiquitin, 22 kDa barstar−barnase complex, and 82 kDa malate synthase G) are processed incrementally at small acquisition time steps. The range of molecular sizes illustrates applicability in both sample- and sensitivity-limited regimes. In each case, the target was to acquire all backbone resonances in the spectra. For the three systems, the targets are achieved after 4.5 min, 1.6 h, and 22 h of acquisition time, respectively. A number of other targets that can be similarly monitored as a function of time are discussed.

Introduction

All aspects of contemporary biomolecular NMR spectroscopy are focused on obtaining relevant spectroscopic information while eliminating noise and artifacts. This is the common aim of pulse sequence and hardware development as well as multinuclear isotope labeling.[1] However, the precision of each signal and the completeness of results are rarely addressed on a methodological level. The focus of this work is to develop techniques for monitoring the quality and content of acquired information concurrently with data collection. First, such integration should provide a lock-on-target, i.e., to continuously control progress toward the acquisition of signals of interest. Subsequently, it becomes a time optimization tool for obtaining complete and precise data. This approach is illustrated on three protein systems of increasing size and complexity: 8 kDa ubiquitin, 22 kDa barstar−barnase complex, and 82 kDa malate synthase G.

A precise judgment of the experiment's time allocation and selection of acquisition parameters can only be done from the spectra themselves. In conventional approaches, a whole spectrum must be recorded before processing, quality validation, and analysis. This often results in recording either insufficient or highly redundant data. The problems are well known and are sometimes referred to as "sensitivity-limited" and "sampling-limited" regimes. The former may be alleviated, for example, by increasing measurement time; the latter has recently received a lot of attention and is addressed by several "fast" approaches.[2−10] A promising solution in both cases is given by an incremental acquisition and concurrent data analysis. So far, this approach has been demonstrated only in the context of reduced dimensionality techniques[11,12] for acquiring an optimized set of spectral projections for signal identification. An alternative to reduced dimensionality is a scheme often referred to as nonuniform sampling in the time domain.[13] This scheme does not limit measurements to projections and gives greater flexibility in optimizing spectral sensitivity and resolution. Nonuniform sampling, however, requires special processing methods, e.g., nonuniform Fourier transform,[14,15] maximum entropy,[13] or multidimensional decomposition (MDD);[9,16,17] the latter is applied in this study. Notably, these processing techniques deal with data sampled in arbitrary (nonuniform) fashion and, as such, are also applicable for reconstructing spectra from the reduced dimensionality data. Recently, a method related to MDD was applied to the projection data and could be considered for incremental acquisition in the frequency domain.[18]

(1) Kay, L. E. *J. Magn. Reson.* **2005**, *173*, 193−207.
(2) Freeman, R.; Kupce, E. *Concepts Magn. Reson.* **2004**, *23A*, 63−75.
(3) Schanda, P.; Kupce, E.; Brutscher, B. *J. Biomol. NMR* **2005**, *33*, 199−211.
(4) Zhang, F. L.; Bruschweiler, R. *J. Am. Chem. Soc.* **2004**, *126*, 13180−13181.
(5) Rovnyak, D.; Frueh, D. P.; Sastry, M.; Sun, Z. Y. J.; Stern, A. S.; Hoch, J. C.; Wagner, G. *J. Magn. Reson.* **2004**, *170*, 15−21.
(6) Kupce, E.; Nishida, T.; Freeman, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *42*, 95−122.
(7) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125*, 1385−1393.
(8) Frydman, L.; Lupulescu, A.; Scherf, T. *J. Am. Chem. Soc.* **2003**, *125*, 9204−9217.
(9) Orekhov, V. Y.; Ibraghimov, I.; Billeter, M. *J. Biomol. NMR* **2003**, *27*, 165−173.
(10) Mandelshtam, V. A.; Taylor, H. S.; Shaka, A. J. *J. Magn. Reson.* **1998**, *133*, 304−312.
(11) Eghbalnia, H. R.; Bahrami, A.; Tonelli, M.; Hallenga, K.; Markley, J. L. *J. Am. Chem. Soc.* **2005**, *127*, 12528−12536.
(12) Hiller, S.; Fiorito, F.; Wuthrich, K.; Wider, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10876−10881.
(13) Hoch, J. C.; Stern, A. S. *NMR data processing*; Wiley-Liss: New York, 1996; xi + 196 p.
(14) Marion, D. *J. Biomol. NMR* **2005**, *32*, 141−150.
(15) Kazimierczuk, K.; Kozminski, W.; Zhukov, I. *J. Magn. Reson.* **2006**, *179*, 323−328.
(16) Tugarinov, V.; Kay, L. E.; Ibraghimov, I.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2005**, *127*, 2767−2775.
(17) Jaravine, V.; Ibraghimov, I. V.; Orekhov, V. Y. *Nat. Methods* **2006**, *3*, 605−607.
(18) Malmodin, D.; Billeter, M. *J. Am. Chem. Soc.* **2005**, *127*, 13486−13487.

Table 1. Spectral Parameters of the 3D HNCO Experiments

	ubiquitin[a]		barstar–barnase complex[a]		malate synthase G[b]
Samples					
sample conc (mM)	1.7		1.2		0.5
MW (kDa)	8		22		82
Spectra					
reference in the text	Ubi1	Ubi2	BBc2	BBc4	MSG
no. of transients, nt	1	2	2	4	8
interscan delay, D1 (s)	0.7	1.2	0.75	0.75	1.5
spectral width (Hz)					
F3(^{1}H)	12 500		12 500		9565
F1(^{13}C)	2000		2350		1506
F2(^{15}N)	2000		2800		1700
time domain size (points)					
F3(^{1}H)	1024		1250		613
F1(^{13}C)	32		64		28
F2(^{15}N)	32		64		54
digital resolution,[c] δ^H, δ^C, δ^N (ppm)					
F3(^{1}H)	0.015		0.0125		0.026
F1(^{13}C)	0.310		0.183		0.357
F2(^{15}N)	0.771		0.540		0.519
frequency domain size (points)					
F3(^{1}H)	240		292		153
F1(^{13}C)	64		128		64
F2(^{15}N)	64		128		128
^{1}H spectrometer frequency (MHz)	800		800		600
full measurement time (h)	1.0	3.0	8.4	16.8	22.0

[a] For ubiquitin and barnase−barstar complex, two spectra were recorded (with twice the number of transients). Only parameters that differ between the two are presented in the second columns; peak identification was performed using ^{1}H−^{15}N assignments. [b] ^{1}H, ^{13}C, ^{15}N assignment from the BMRB entry 5471 was used for peak identification. Out of 653 BMRB assignments, 638 peaks were found in the full spectrum and used as targets. [c] The digital resolution is defined as the inverse of the acquisition time.

In this work, we introduce a method of targeted acquisition (TA) that involves concurrent data accumulation using incremental nonuniform sampling (INUS), data processing using MDD, and monitoring of spectral quality. A choice of quality scores and a set of signals for monitoring constitute user-defined targets, which depend on the type of experiment and a priori knowledge about the system under study. For example, in this work, TA aims for an optimal measurement time allocation to detect the backbone signals and is used to monitor the precision of signal positions and intensities. The corresponding scores are defined for a set of a priori determined lists of signals with known assignments.

Experimental Section

Table 1 shows details of the NMR experiments performed. The proteins were ^{13}C,^{15}N-labeled (except for barstar): ubiquitin (Ubi), barstar−barnase complex (BBc), and malate synthase G (MSG), with the last also fully deuterated. NMR experiments were performed at 25 °C on Varian spectrometers with Larmor frequencies of 800 MHz (Ubi and BBc spectra) and 600 MHz (MSG), equipped with a room-temperature pulsed-field gradient triple-resonance probe. The spectrum of MSG was measured in the group of Professor L. Kay (Toronto University). For Ubi and BBc, 3D HNCO spectra were recorded using gradient sensitivity-enhanced pulse sequence from the BioPack library (Varian Inc.); for MSG, a TROSY version of the experiment from the Toronto NMR library was used.

In this work, the INUS schedule described below was implemented off-line, i.e., quartets of 1D FIDs representing hypercomplex data points in the two indirect dimensions were extracted in accordance with an INUS schedule from a prerecorded complete reference data set for each system. The R-MDD procedure[17] (eqs 1−3, below) was implemented in a home-built software, *mddNMR*. This software is available from the authors.

Theoretical Basis

The MDD Model. The multidimensional decomposition (MDD) model 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.[19] When applied to NMR spectra, the MDD can be formulated as follows. Given a matrix $\boldsymbol{S}$ with sizes N_m of its M dimensions ($m = 1, ..., M$) and elements $S_{n1,n2,...,nM}$, the algorithm finds scalar numbers ${}^\beta a$ and normalized vectors ${}^\beta \boldsymbol{F}^m$, with elements ${}^\beta F^m(n_m)$ ($n_m = 1, ..., N_m$), such that the following norm becomes minimal:[20]

$$|\boldsymbol{G}\cdot[\boldsymbol{S} - \sum_\beta ({}^\beta a\, {}^\beta \boldsymbol{F}^1 \otimes {}^\beta \boldsymbol{F}^2 ... \otimes {}^\beta \boldsymbol{F}^M)]|^2 + \lambda \sum_\beta ({}^\beta a)^2 \quad (1)$$

Here, the symbol "⊗" denotes the tensor product operation and the matrix $\boldsymbol{S}$ corresponds to an experimental M-dimensional NMR spectrum in time and/or frequency domain representation. In the case of sparse (nonuniform) sampling, only a fraction of the elements in $\boldsymbol{S}$ is measured, and the matrix $\boldsymbol{G}$, which contains elements $g_{n1,n2,...,nM} \in \{0,1\}$, 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 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.

The original MDD method assumes no restriction on a particular form of the signal line-shape vectors ${}^\beta \boldsymbol{F}^m$. In all practical cases, the theory predicts the NMR time domain signal as a superposition of a relatively small number of sine waves that can be amplitude-modulated by relaxation or by other factors. This autoregressive property can be incorporated into the MDD model.[17] The model defines a family of possible line shapes and includes decaying sine waves as a subclass. In particular, each of the time domain shapes ${}^\beta \boldsymbol{F}^m$ of length N_m is represented as a product of K_m vectors of length d_m, so that $N_m = (d_m)^{K_m}$:

$${}^\beta \boldsymbol{F}^m = {}^\beta \boldsymbol{V}^1 \otimes {}^\beta \boldsymbol{V}^2 ... \otimes {}^\beta \boldsymbol{V}^K \quad (2)$$

This equation defines a second decomposition to the shapes obtained in the first decomposition (eq 1). The second MDD decomposition reduces the number of unknowns in the least-squares minimization of eq 1. After replacing vectors ${}^\beta \boldsymbol{F}^m$ for all β in eq 1 by the right-hand side of eq 2, M-dimensional decomposition becomes an $(M - 1 + K_m)$-dimensional one.

(19) Kruskal, J. B. Rank, decomposition, and uniqueness for 3-way and N-way arrays. In *Multiway data analysis*; Coppi, R., Bolasco, S., Eds.; North-Holland Elsevier Science Pub.: Amsterdam/New York, 1989.
(20) Ibraghimov, I. *Num. Linear Algebra Appl.* **2002**, *9*, 551−565.

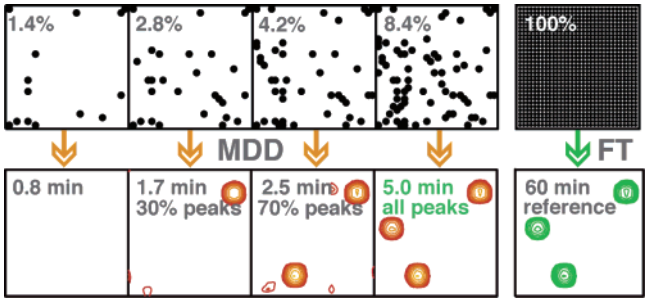

Figure 1. Targeted acquisition procedure. Actual data from several consecutive TA steps (columns) for Ubi1 exemplify a direct connection between the amount of sampled data (indicated as percentages in the upper panels) and the fraction of detected target peaks in the MDD reconstructions (indicated as percentages and elapsed measurement times in the bottom panels). The INUS schedule, i.e., pairs of time domain increments in the $^{13}C'$,^{15}N dimensions acquired at each TA step, are shown as dots in the upper panels. The rightmost column is given for comparison and represents a conventional reference spectrum.

Subsequently, the vector ${}^{\beta}\boldsymbol{F}^{m}$, with N_m unknown elements, is defined by a much smaller number of parameters, $d\,\mathrm{Log}_d(N_m)$.

Although input spectrum $\boldsymbol{S}$ may have a significant fraction of missing data, the output shapes ${}^{\beta}\boldsymbol{F}^{m}$ are complete, and the complete spectrum reconstruction $\boldsymbol{S}^{\mathrm{R}}$ can be calculated:

$$\boldsymbol{S}^{\mathrm{R}} = \sum_{\beta}({}^{\beta}a\,{}^{\beta}\boldsymbol{F}^{1} \otimes {}^{\beta}\boldsymbol{F}^{2} \ldots \otimes {}^{\beta}\boldsymbol{F}^{M}) \qquad (3)$$

The overall procedure can be understood as filling gaps in the original matrix $\boldsymbol{S}$ using an interpolation model. The complete spectrum is further processed as usual with Fourier transform, linear prediction, etc.

Algorithm for the Generation of the INUS Schedule. Figure 1 (top panels) illustrates the idea of the INUS schedule. At each step, a small fraction of data points from a putative complete uniform sampling schedule is randomly selected and measured. To enhance the sensitivity of the experiment, the sampling schedule is matched to the decay of spin coherences.[21] Occasionally, owing to random selection, the measurement of some data points may be repeated. Thus, the INUS schedule can contain any number of points, which naturally allows for the reduction or extension of time allocation for multidimensional experiments, depending on spectra sensitivity.

The schedule matches a two-dimensional probability density function for the indirect evolution dimensions. The function is defined on a two-dimensional grid (t_1,t_2) determined by spectral widths and maximal acquisition times $(t_{1_{max}},t_{2_{max}})$ in the dimensions. The distribution is obtained as a product of the two envelopes, $P(t_1,t_2) = P1(t_1)\cdot P2(t_2)$. The envelope functions $P1(t_1)$ and $P2(t_2)$ are devised to match the signal coherences in the indirect dimensions for a particular system and experiment. Because a ^{15}N chemical shift evolves in all our experiments during a constant delay, the envelope in this dimension is a constant function, $P2(t_2) = 1$. The envelope of the signals in the $^{13}C'$ dimension, $P1(t_1) = \exp(-t_1/T_2)$, is the exponential decay with transverse relaxation time T_2, which is an input parameter for the procedure and should be roughly estimated for each protein system (Table 1). However, unless the $^{13}C'$ acquisition time exceeds T_2, the results are insensitive to this parameter. For a given probability distribution, we use the following procedure to generate the INUS schedules. First, a pair of integer indices is randomly selected that corresponds to the acquisition times (t_1,t_2). The pair is then added to the sampling schedule table if the corresponding value of the probability distribution $P(t_1,t_2)$ is larger than another randomly generated number ranging between 0 and 1; otherwise, the index pair is discarded. This process is repeated until the sampling table contains the requested number of data points for each step. In this work, we used an additional optional condition in which the points are not repeated. This condition ensures that at the last INUS step the sampling table corresponds to a re-ordered but complete reference data set without repetitions. For all spectra we had 128 steps, and the numbers of points generated for each step were 8, 32, and 12 for Ubi1,2, BBc2,4, and MSG, respectively. Thus, an INUS schedule is a table of evolution delays (t_1,t_2) spanning maximal acquisition times and spectral widths in the indirect dimensions and grouped sequentially into a certain number of steps. The algorithm is implemented as a stand-alone program and is available from the authors. It is also included in a standard BioPack package (Varian Inc.), which has a corresponding interface that allows setting up an INUS schedule for any multidimensional experiment.

Procedure for Targeted Acquisition with INUS. The targeted acquisition procedure is illustrated in Figure 1. At each step, an incrementally increasing data set is processed using the MDD algorithm to obtain the reconstructed spectrum. The data set is then evaluated in the frequency domain to monitor completeness and quality of signal content. The general procedure is described below, along with additional/alternative procedures that were performed in this work for method validation and comparison with the reference spectra.

1. Define *targets*, e.g., number of anticipated signals and scores for monitoring their quality. To define the target on number of signals, count the peaks in a 2D heteronuclear single quantum coherence (HSQC) or 2D projection spectrum, or estimate the number of signals from a protein amino acid sequence. In general, targets are determined by a problem in consideration and depend on available routines for evaluating regular multidimensional spectra, e.g., peak-pickers. Targets can use available a priori knowledge about a protein system, e.g., known peak positions in some of the dimensions. They can be defined for selected signal subsets and can be based on various quality scores and peak parameters. The precision of parameters of individual signals, e.g., peak positions and intensities, can be obtained using the Jackknife procedure,[22] which basically compares results obtained with partially different input data. In this work, we monitored the differences in values of the signal parameters obtained on consecutive TA steps. Also in this study, the MDD reconstructions were evaluated by comparison with the conventionally obtained reference spectrum. We monitored the following parameters: total number of peaks in the raw peak list provided by the automatic peak-picker, number of peaks from the target list containing all assigned backbone signals, number of false peaks, accuracies of the peak intensities and positions in the $^{13}C'$ dimension, and precision of the $^{13}C'$ frequencies.

2. Prepare a table with the *incremental* INUS schedule (see above), a list of pairs of indices in the indirect dimensions grouped sequentially into a number of steps (128 in this work).

(21) Barna, J. C. J.; Laue, E. D.; Mayger, M. R.; Skilling, J.; Worrall, S. J. P. *J. Magn. Reson.* **1987**, *73*, 69−77.

(22) Efron, B. *The jackknife, the bootstrap, and other resampling plans*; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1982; vii + 92 pp.

The following procedure is performed for each TA step:

3. A new portion of the data comes from a spectrometer that is running an experiment without interruption in accordance with the INUS schedule. In this work, the INUS schedule was implemented off-line; i.e., data were extracted from complete uniformly sampled reference spectra rather than actually being measured.

4. The time domain of the directly detected dimension (t_3) is converted into the *nmrPipe*[23] format and processed in the conventional way, i.e., Fourier-transformed using a 30°-shifted squared sine-bell window function and zero-filling to 2K points.

5. To organize fast parallel calculations, the spectrum is split into ~20 overlapped segments in the directly detected $\omega 3$ dimension (e.g., amide range 6.05−9.7 ppm for Ubi). The segment data are converted into the format required for the MDD software. To account for possible incomplete reconstruction of signals on the segment borders (in ω_3), only the central 28-point parts (of 32) were merged to obtain the whole spectrum reconstruction, while the overlapping four-point margins were discarded.

6. The number of components is determined for each spectral segment by counting the approximate number of peaks in the 2D ^{1}H−^{15}N projections. To account for the presence of possible minor peaks and some large noise features or spectral artifacts, the number was increased by 30%. Unless the number of components is too small, the results are rather insensitive to this parameter. If estimates for individual segments are not available, a maximal value can be used as a default, e.g., one-quarter of the total number of expected peaks in the spectrum.

7. The minimization of eq 1 is performed to obtain shapes ${}^{\beta}\boldsymbol{V}^i$ and amplitudes ${}^{\beta}a$ for all components β. Segment minimization is performed as independent calculations, which naturally allows parallelization. To speed up the convergence of the algorithm, a solution from the previous TA step is used as an initial approximation, except for the first step, in which random numbers are used instead. Convergence of the minimization routine is established by extending the computation for an additional period with no further decrease in the residual.

8. The shapes ${}^{\beta}\boldsymbol{F}^i$ (eq 2) are calculated. These are time domain (t_1,t_2) and frequency domain (ω_3) profiles.

9. The reconstructed spectrum is calculated from the shapes and amplitudes for each segment using eq 3. The segments are merged into a regular 3D data set (in the *nmrPipe* format), and then the time domain signal in t_1 and t_2 is multiplied with a squared cosine-bell window function, zero-filled to double size, and Fourier-transformed.

10. The peak-picking routine *pkFindROI* from the *nmrPipe* package is used on the reconstructed 3D spectrum to obtain the *raw* peak list and count the signals. The exact peak positions in the peak lists were refined as a maximum in three-point interpolation.

11. Target scores are calculated and the decision is made whether the experiment can be finished.

Results and Discussion

Figure 2 presents the results of the incremental acquisition as a function of time. Completeness of the reconstructed spectrum (Figure 2A) is defined as a fraction of the target peak list detected at each TA step. The target lists for Ubi, BBc, and MSG contain 70, 101, and 638 signals, respectively, as shown in Table 2. All these signals correspond to assigned backbone amide peaks found in the reference spectra. For peak identification, the same automatic procedure from *nmrPipe*[23] was used in both the reference and MDD reconstructions. This was done for consistent and reproducible comparison. The lists of peaks for the reference spectra were manually verified. Raw 3D peak lists were matched with the target list. The sizes of the raw peak lists exceeded the target number by 5−10% because of signals from the side chains and unassigned minors. However, the lists contained no false peaks except for a few verified peak-picker failures. The curves in Figure 2A are the sigmoid functions of acquisition time. Initially, a few strong peaks appear, followed by detection of the bulk of peaks with average intensity, and trailed by a few weak peaks. For ubiquitin, the large sensitivity allows the detection of all HNCO backbone correlations with only one scan in 4.5 min. It takes 3 times longer for the spectrum with double the number of scans and a 50% longer interscan delay. This situation is a clear sample-limiting regime. In other words, the same number of data points is required regardless of sensitivity. For BBc, the situation is sample-limited for the majority of peaks and sensitivity-limited for a few of the weakest signals. The latter group determines the total time (about 2 h) required to obtain all target correlations in both BBc spectra. The MSG curve exemplifies the sensitivity-limited regime. Parts B and C of Figure 2 illustrate a gradual

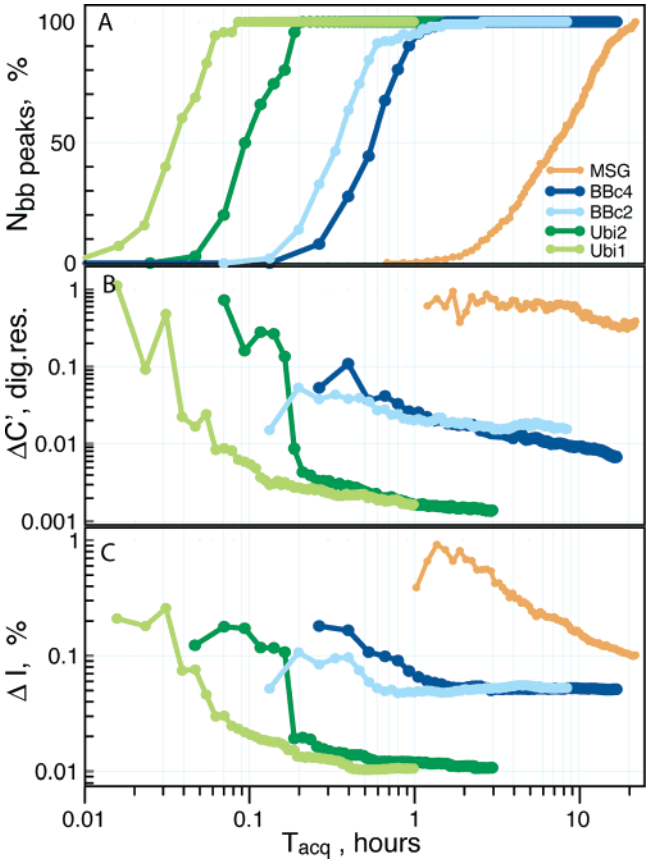

Figure 2. Evaluation of the spectral content and quality of the HNCO spectra for the 128 INUS steps: (A) percentage of detected backbone peaks; (B) mean accuracies of the chemical shifts C′ in units of digital resolution; (C) accuracies of the peak intensities, normalized to the median peak amplitude. Color code: light green, Ubi1; green, Ubi2; light blue, BBc2; blue, BBc4; and red, MSG.

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

Table 2. Results of the MDD Calculations for Five 3D HNCO Spectra

	Ubi1	Ubi2	BBc2	BBc4	MSG
			INUS		
sampling bias in $^{13}C'$, T_2 (ms)	50		30		25
			MDD Calculations		
average CPU time[a] for ~20 regions	0.0026		0.0064		0.0511
(min, max) per one TA step (h)	(0.0006, 0.0070)		(0.0018, 0.0135)		(0.0017, 0.1242)
the same, relative to the experimental	0.327	0.112	0.097	0.050	0.297
time, per step	(0.083, 0.900)	(0.028, 0.294)	(0.028, 0.206)	(0.014, 0.101)	(0.010, 0.723)
			Peak Number Targets		
target no. of peaks,[b] N_{peaks}	70 (72)		101 (106)		638 (699)
reference, raw N_{peaks}	70	70	108	114	733
target 50%					
step[c]	5	5	5	4	39
N_{peaks}, raw/target/false[d]	49/48/1[e]	57/56/1[e]	49/47/1[f]	48/48/0	338/316/0
time[g] (h) (% total time)	0.04 (3.9%)	0.12 (3.9%)	0.33 (3.9%)	0.53 (3.1%)	6.70 (30.5%)
target 95%					
step[c]	9	8	13	8	85
N_{peaks}, raw/target/false[d]	68/68/0	68/68/0	97/95/1[i]	100/99/0	639/606/0
time[g] (h) (% total time)	0.07 (7.0%)	0.19 (6.2%)	0.85 (10.2%)	1.05 (6.2%)	14.61 (66.4%)
target 100%					
step[c]	11	9	30	12	128
N_{peaks}, raw/target/false[d]	70/70/0	70/70/0	103/101/0	104/101/0	732/638/0
time[g] (h) (% total time)	0.09 (8.6%)	0.21 (7.0%)	1.97 (23.4%)	1.58 (9.4%)	22.0 (100%)
$S/N_{full\ spectrum}$ ($S/N_{50\%}$)[h]	333 (62)	610 (107)	85 (17)	126 (23)	24 (15)

[a] The computation times are those required on a single-processor iMac PowerPC G5 1.8 GHz. Calculations for MSG were performed on the HPC Linux cluster (ww.nsc.liu.se) with Intel Xeon processors (2.2 GHz), where each of the segments was given a separate processor. The times for MSG are scaled by a factor of 0.812 to closely correspond to the iMac performance. [b] Number of amide signals corresponding to published assignments: Ubi;[24] BBc, BMRB entry 7126; and MSG, BMRB entry 5471. The figures in parentheses give the total number of backbone amide HN groups in the amino acid sequence.[c] Step number (of 128 total) when 50% (95% and 100%) of the peaks from the target peak list is obtained. [d] "raw" is the number of all peaks detected by an automated peak-picker, "target" is the number of raw peaks which have a backbone assignment, and "false" is the number of peaks not found in the corresponding reference (within tolerances of digital resolution). The raw peaks can include several attenuated side-chain peaks (e.g., three peaks for BBc2 or BBc4). [e] The peak is a fault of the peak-picker: it is a part of the t_1 noise tail of the correct signal in several early reconstruction steps; it did not appear in subsequent steps with more input data (i.e., above 5% data). [f] The peak is a fault of the peak-picker: it was centered directly on the border of the C′ region, which caused flickering of the peak position between the upper and lower borders of the spectrum. [g] Measurement time required to record data used for the TA step presented. The number in parentheses is the time relative to the complete experiment. [h] Signal-to-noise ratio (S/N) in the full reference spectrum for a peak with median intensity. In parentheses are the S/N values extrapolated from the reference using square-root dependence on measurement time corresponding to a 50% target on the number of peaks.

improvement in accuracy of the positions and intensities of the detected peaks. After about 90% of the peaks are acquired, the accuracies of the chemical shifts and the peak intensities level off below digital resolution and fraction of a percent, respectively.

The use of the reference spectrum and a well-defined target set of signals allowed us to validate the subject method by demonstrating the gradual improvement in completeness and accuracy of the signals of interest. In practical applications, a target peak list can be a set of peaks picked in a 2D correlation spectrum that is recorded and analyzed in a few minutes before setting up a series of multidimensional experiments. The precision of the spectral parameters can be obtained concurrently with data accumulation using a Jackknife/Bootstrap algorithm. Figure 3 shows how the precision of carbonyl chemical shifts of individual peaks, defined as the difference in values at two successive steps, gradually improves as the number of measurements used for the BBc4 spectral reconstructions increases. Good precision is obtained after the majority of peaks are described by the MDD model.

Similar to a conventional experiment, where peak line shapes and intensities do not change with an increasing number of transients, in the INUS, additional measurements only improve signal precision, while noise level gradually decreases. When all expected peaks are obtained with a requested precision, there is no need to continue acquisition; and vice versa, it makes sense to continue the experiment if some expected peaks are still not detected. It must be kept in mind, though, that some signals

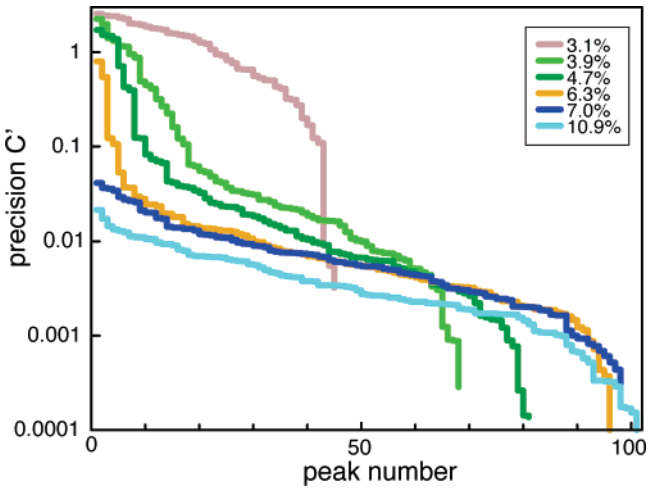

Figure 3. Precision of the chemical shifts (C′) of the individual backbone resonances (sorted in descending order) detected at various TA steps in the BBc4 spectrum. A version of the Jackknife procedure determines precision as the difference in C′-shift estimation between two successive steps, normalized to acquisition digital resolution (refer to Table 1). The precision values for TA steps 4, 5, 6, 7, 9, and 14, corresponding to 3.1, 3.9, 4.7, 6.3, 7.0, and 10.9% of the complete data, are shown with colored lines.

could be very weak and require too much time. The sigmoid shape of the curves in Figure 2A with a steep slope at the half height permits an optimal measurement time to be estimated at the early stages of data acquisition. Indeed, 50% (95%) peaks are detected at 3.9 (7.0), 3.9 (6.2), 3.9 (10.2), 3.1 (6.2), and 30.5% (66.4%) of the total experimental time for the five graphs

in Figure 2A. Thus, doubling the time taken to obtain the first half of the signals is sufficient to obtain about 95% of the peaks. The estimates are rather tolerant of the errors in target estimation; values close to these were obtained using the target peak numbers inferred from the amino acid sequences. Additional time allocation to detect the remaining weak signals largely depends on their sensitivity.

The MDD calculations are computationally demanding. However, when the directly detected dimension was split into a number of spectral regions, the parallel (on a Linux cluster) calculation times for each TA step were always shorter than the corresponding measurement times. This observation proves the feasibility of concurrent real-time acquisition and analysis of INUS data. Without a computer cluster, a single up-to-date computer (PowerPC G5 1.8 GHz) is adequate to assess the quality of several of the most difficult, crowded regions in real time. Segmentation does not affect algorithm convergence or quality of reconstruction. It is done only to save computational time by parallel processing, since segments can be processed independently.

Here, for the first time, we demonstrate incremental acquisition and concurrent analysis using nonuniform sampling. Although we use MDD in this study for TA, similar results perhaps could be obtained using other methods capable of dealing with nonuniformly sampled data, e.g., maximum entropy.

Conclusions

The results of our experiments demonstrate a novel concept in multidimensional NMR spectroscopy—incremental nonuniform data collection targeted at acquisition of biomolecular information of interest. The examples illustrate that the approach is applicable to both small and large protein systems, as it offers optimization of measurement time allocation, depending on experiment sensitivity. The use of matched sampling offers sensitivity enhancement per unit time. Targeted acquisition can considerably enhance the performance of automated and integrated protocols that have exploded in the second phase of structural genomics[11,12,25−28] because such protocols depend critically on quality and completeness of input data. For example, when acquisition is coupled with concurrent sequential assignment, more data and improved precision of peak positions can progressively reduce ambiguity until assignment is accomplished.

Acknowledgment. This work was supported by Swedish national research grants SSF A3 04:160d, VR 2005-2951, and SNIC 3/04-44, and a visiting research fellowship from the Wenner-Gren foundation. We are grateful to S. Grzesiek, D. Korzhnev, and L. Kay for providing us with the Ubi and BBc samples and the spectrum of MSG.

JA062146P

(24) Wang, A. C.; Grzesiek, S.; Tschudin, R.; Lodi, P. J.; Bax, A. *J. Biomol. NMR* **1995**, *5*, 376−382.
(25) Huang, Y. J.; Moseley, H. N.; Baran, M. C.; Arrowsmith, C.; Powers, R.; Tejero, R.; Szyperski, T.; Montelione, G. T. *Methods Enzymol.* **2005**, *394*, 111−41.
(26) Medek, A.; Olejniczak, E. T.; Meadows, R. P.; Fesik, S. W. *J. Biomol. NMR* **2000**, *18*, 229−238.
(27) Sun, Z. Y.; Hyberts, S. G.; Rovnyak, D.; Park, S.; Stern, A. S.; Hoch, J. C.; Wagner, G. *J. Biomol. NMR* **2005**, *32*, 55−60.
(28) Masse, J. E.; Keller, R.; Pervushin, K. *J. Magn. Reson.* **2006**, *181*, 45−67.