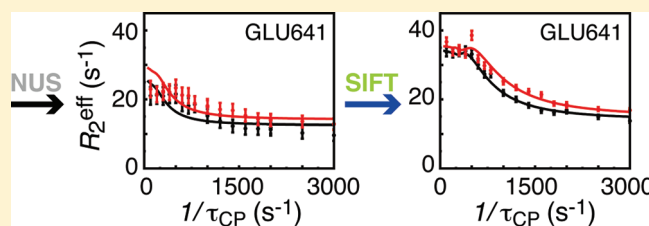


# Boosting Protein Dynamics Studies Using Quantitative Nonuniform Sampling NMR Spectroscopy

Yoh Matsuki,<sup>†</sup> Tsuyoshi Konuma,<sup>‡</sup> Toshimichi Fujiwara,<sup>†</sup> and Kenji Sugase<sup>\*,‡</sup><sup>†</sup>Institute for Protein Research, Osaka University, Osaka, Japan,<sup>‡</sup>Bioorganic Research Institute, Suntory Foundation for Life Sciences, Osaka, Japan

S Supporting Information

**ABSTRACT:** NMR spectroscopy is uniquely suited to study protein dynamics over a wide range of time scales at atomic resolution. However, existing NMR relaxation methods require highly serial, lengthy data collection, ultimately limiting their application to short-lived samples, such as proteins in living cells. In recent years, the utility of nonuniform sampling (NUS) NMR methodologies has been increasingly recognized, but their application has been rare in relaxation measurements where highly accurate spectral quantification is demanded. Recently, Matsuki et al. developed a new NUS-processing method, SIFT (Spectroscopy by Integration of Frequency and Time domain information), which is highly robust and faithful in reproducing signals. In this work, we demonstrate the gains that are possible with more aggressive use of frequency domain information than was employed previously. This improvement is crucial for SIFT to be used in accelerating relaxation measurements while preserving full analytical accuracy. By taking the KIX domain of mouse CREB-binding protein (CBP) as an example, we demonstrate that this quantitative NUS processing method enables total 10-fold expedition of the  $R_2$  relaxation dispersion measurements. The advanced SIFT processing should be equally useful for other NMR relaxation measurements.



## INTRODUCTION

Protein dynamics plays an essential role in many biological processes, including ligand binding, enzyme catalysis, molecular recognition, and protein folding.<sup>1</sup> Nuclear magnetic resonance (NMR) spectroscopy is uniquely suited to investigate protein dynamics over a wide range of time scales at atomic resolution. Accordingly, a variety of methods are available for the measurements of nuclear relaxation rates ( $R_1$ ,  $R_2$ , heteronuclear NOE)<sup>2</sup> for elucidating local molecular flexibility, relaxation dispersion profiles for (conformational) exchange parameters,<sup>3</sup> and  $^1\text{H}$  exchange rates (using CLEANEX-PM,<sup>4</sup> and/or  $^1\text{H}/^2\text{H}$  exchange<sup>5</sup>) for solvent accessibility. A common drawback with these methods, however, is their lengthy data collection, which ultimately limits the application of the methods to short-lived or low-sensitivity dilute samples typical of proteins that aggregate quickly or are difficult to express in large quantities. One of the most striking examples is an NMR study using living-cells, or in-cell NMR. In-cell dynamics studies are currently prohibited due to the short lifetime of proteins (<6 h for a putative heavy-metal binding protein, TTHA1718, at 37 °C) in the living cells.<sup>6</sup>

There are four reasons for the lengthy data collection required for the aforementioned NMR analysis. First, they require collecting typically dozens of two-dimensional (2D) spectra with a systematic change of the length of the relaxation or build-up delay in the radio frequency (RF)-pulse sequences.<sup>2,4,7,8</sup> Second, the series of measurements are often to be repeated at different sample temperatures, concentrations, external field strengths

( $B_0$ ) or orders of the coherence (single- or multiquantum).<sup>3</sup> Third, each 2D spectrum needs very high resolution (a large number of data points in the indirect dimension) to resolve most of the peaks in each 2D plane for site-specific information and accurate peak quantification. A serial 2D measurement is equivalent to collecting a 3D data set, and forces a similar compromise between desired spectral resolution and available instrumental time. And last, the data are to be collected with a relatively long recycle delay (>3 s) to enhance the accuracy of measured peak intensities. The NMR-based dynamics studies use peak intensities as quantitative sources of information on protein motions, thus they are more stringent about sample stability than other classes of measurements such as those for signal assignments, for which a bare detection of peaks suffices, and partial degradation or precipitation during the measurement does not affect the assignments for a protein consisting of the major part of the sample. Due to these challenges, most of the successful protein dynamics studies so far have been limited to constructs optimized to provide enough sensitivity and stability.

The approach that extensively reduces the bandwidth beyond the Nyquist bandwidth to speed the data collection is not generally applicable here, although it has been successfully demonstrated for small molecules.<sup>9</sup> In this approach, the avoidance of the peak

Received: August 23, 2011

Revised: October 11, 2011

Published: October 12, 2011

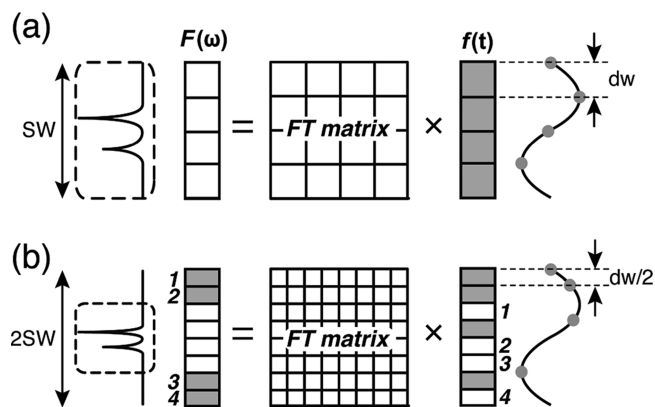
overlap due to aliasing is not always guaranteed because of the random nature of the peak positions, and will be particularly problematic for large protein systems. Moreover, the smaller bandwidth linearly degrades the sensitivity of the measurement, and is thus difficult to apply to dilute protein samples, which is the aim in this work.

In recent years, the utility of nonuniform sampling (NUS) for faster acquisition of NMR data has been increasingly recognized with the advent of a number of efficient NUS-processing methods, such as nonlinear Fourier transform (NLFT),<sup>10,11</sup> maximum entropy (MaxEnt)<sup>12,13</sup> and multi-dimensional decomposition (MDD).<sup>14,15</sup> Although the use of NUS NMR is promising, its application has been limited for protein dynamics studies. This is partly because the success of the studies highly hinges upon accurate quantification of peak intensities, a situation that has long been under discussion for the existing NUS-processing tools. Although NUS with MaxEnt or MDD has been used in a semiquantitative manner for protein structure determination,<sup>6,16</sup> quantification of NOE peaks is one of the examples where only rough quantification suffices without serious degradation of the final structures. This is in contrast to protein dynamics measurements that entail direct and precise fitting of peak intensity variations along the serial measurements with theoretical curves to determine dynamics parameters and their confidence limits.

Recently, Matsuki et al. developed a new NUS-processing method, SIFT (Spectroscopy by Integration of Frequency and Time domain information),<sup>17,18</sup> which is highly robust and faithful in reproducing signals. NUS together with SIFT-processing has been shown to achieve for a decaying signal either higher sensitivity in a given measurement time or faster data acquisition for the same sensitivity.<sup>18</sup> While the sensitivity gain has been demonstrated in previous applications of SIFT,<sup>17,18</sup> we here show that significant acceleration of relaxation measurements is possible with NUS and SIFT. For this new application of SIFT, we demonstrate a more aggressive use of frequency domain information for SIFT so that it can process NUS data even when the data are recorded without oversampling in the indirect dimension. This improvement is crucial in accelerating the measurement while preserving the quantitative accuracy and reliability of the processing result. The reliability of the measurement is of central importance in the relaxation measurements. With the advanced SIFT processing, we demonstrate a total 10-fold expedition of  $R_2$  relaxation dispersion measurements without appreciably compromising the analytical accuracy. The acceleration is afforded by minimizing the number of data samples using NUS (2.5-fold) and the number of scans per FID (4-fold) utilizing the robustness of SIFT against noise. The new SIFT procedure is equally valid for other NMR relaxation measurements.

## MATERIALS AND METHODS

**New SIFT Procedure.** SIFT is based on the classical error-reduction (Gerchberg–Papoulis) algorithm<sup>19,20</sup> that has been extensively studied and used in picture processing and medical imaging area.<sup>21–23</sup> The algorithm is an iterative technique based on Fourier/inverse-Fourier transformations, while imposing a priori constraints at each iteration. In this way, information across the domains is integrated without any biasing model or parameters. In the application to NUS NMR spectroscopy, the prior information is about “dark” spectral points, or regions known to be devoid of any signal, and the NUS time-domain data experimentally collected. The SIFT cycle can be concisely formulated for a one-



**Figure 1.** A cartoon illustrating time- and frequency-domain data,  $f(t)$  and  $F(\omega)$ , respectively, paired with the Fourier transform matrix in the cases of Nyquist sampling (a) and twice-oversampling NUS (b). Gray rectangles represent known or measured data points either in the frequency or time-domain whose information content can be integrated via SIFT. In panel b, four dark spectral points are produced at the spectral edges due to the twice-wider bandwidth (on the left-hand side of the equation), which are used to compensate the information deficiency due to NUS in the time domain (on the right-hand side of the equation).

dimensional (1D) case as

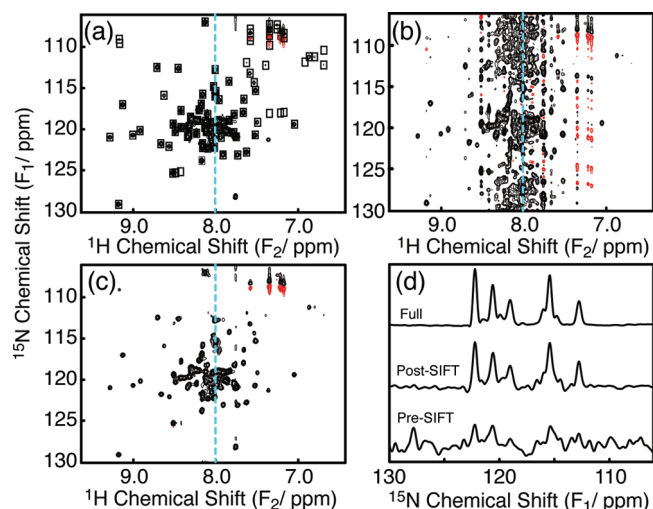
$$f^{k+1}(t) = f^0(t) + \mathbf{P}(t)\mathbf{F}^{-1}\mathbf{D}\mathbf{F}f^k(t) \quad (1)$$

where  $f^p(t)$  with  $p = 0$  or  $p > 0$  denotes the NUS data collected in an experiment, or the data in the  $p$ th SIFT cycle, respectively.  $\mathbf{F}$  and  $\mathbf{F}^{-1}$  are the Fourier and inverse-Fourier transform operators,  $\mathbf{D}$  is the spectral masking operator that zeros all the dark spectral points, and  $\mathbf{P}(t)$  is the NUS operator, whose element is 0 or 1 for the sampled or unsampled time grid, respectively. The step-by-step description of the data treatment is given in a previous publication.<sup>18</sup>

SIFT reinstates unsampled time-domain points as much as the information given by “dark” spectral points by integrating this frequency-domain information back into the time domain. Thus, the processing is most accurate and reliable when the number of dark spectral points exceeds the number of unmeasured elements in the NUS time domain. In this most conservative sampling regime, called the above-critical sampling condition, improvement of overall sensitivity of NMR measurements has been shown previously.<sup>17,18</sup>

In the previous applications of SIFT to NMR data, dark spectral regions were produced by an intentional oversampling in the indirect acquisition dimension,<sup>18</sup> or existed due to oversampling for other experimental necessity.<sup>17</sup> However, as long as SIFT relies only on the dark information produced by oversampling, the experimental time can never be saved as compared with the experiments conducted without oversampling as illustrated in Figure 1. Oversampled wider bandwidth produces *known* dark spectral points, but produces also the same number of *unknowns* in the time domain due to a smaller Nyquist grid. Namely, overall number of time-domain measurements, shown by gray circles in Figure 1, is invariant. To expedite the measurement, which is the primary goal in this work, one is required either to further thin out the sampling schedule, venturing into the subcritical sampling condition as opposing to the above-critical sampling, or to recruit dark information in a way other than oversampling.

Here, we extended the original SIFT program<sup>18</sup> so that it can find dark points within a fully native spectrum taken at the Nyquist bandwidth, i.e., without oversampling in the indirect



**Figure 2.** The reference spectrum of KIX. Fourier transform of fully acquired data (a), and Fourier transform of NUS data before (b) and after (c) SIFT processing. The rectangles in (a) define the peak-containing “bright” spectral region. The lowest contour level is set to 6% of the highest peak in each panel. The negative contours are shown in red. A representative slice taken at  $F_2 = 8.005$  ppm (indicated by a dashed blue line in a, b, and c) is shown in panel d for the full (top), post-SIFT (middle), and pre-SIFT (bottom) spectra in the same vertical scale.

dimension, by adaptively defining the peak-containing regions as many as the number of peaks in the spectrum. A key idea is that in  $R_2$  dispersion measurements and many other NMR experiments for dynamics studies, one is interested in accurately measuring the change in peak intensities relative to those in the reference spectrum. Because signal positions are known from the reference spectrum taken in the beginning (Figure 2a), and invariant over the serial measurements, one could use all the dark spectral points existing between NMR peaks for SIFTing to accelerate all the subsequent measurements. This is a demonstration of the scout-and-SIFT approach delineated in the original paper.<sup>18</sup> We note that the same expedition is achievable for the measurement regardless of the number of 2D data sets involved in the series. We also emphasize that our method has an important difference from the spectral thresholding method:<sup>24</sup> SIFT does not risk throwing away the peak intensity smaller than an arbitrarily set threshold.

**Sample Preparation.** Uniformly  $^2\text{H}$ - and  $^{15}\text{N}$ -labeled KIX domain (residues 586–672) of mouse CREB-binding protein (CBP) was expressed in BL21-DE3 cells grown in M9 minimal medium and purified by reverse-phase high-performance liquid chromatography (HPLC).<sup>25</sup> The protein was dissolved in NMR buffer [95%  $\text{H}_2\text{O}$ /5%  $\text{D}_2\text{O}$ , 20 mM Tris- $d_{11}$ -acetate- $d_4$  (pH 5.5 at 25 °C), 50 mM NaCl, 2 mM  $\text{NaN}_3$ ] and concentrated to 500  $\mu\text{M}$ . Protein concentration was determined from the absorbance at 280 nm, using an extinction coefficient of 12.95  $\text{mM}^{-1} \text{cm}^{-1}$ .

**NMR Measurements.**  $^{15}\text{N}$   $R_2$  relaxation rates were measured on Bruker BioSpin 600 and 750 MHz spectrometers at 25 °C using relaxation-compensated constant-time Carr–Purcell–Meiboom–Gill (rcCPMG) pulse sequences with a constant relaxation time of  $T_{\text{CPMG}} = 40$  ms.<sup>26,27</sup> The spectral width of the indirect dimension was set to a conventional Nyquist width of approximately 25 ppm so that the backbone amide peaks are not folded, and no oversampling was applied to the indirect acquisition dimension.  $R_2$  dispersion profiles were generated by measuring effective transverse relaxation rates  $R_2^{\text{eff}}$  as a function of  $\tau_{\text{CP}}$ , the delay time between two successive  $180^\circ$  pulses in the CPMG pulse

train. 2D data sets with  $1024 \times 64$  ( $t_2 \times t_1$ ) complex points were acquired at  $B_0 = 14.1$  and 17.6 T, and at  $\tau_{\text{CP}} = 10, 5, 3.33, 2.5, 2.0, 1.66, 1.43, 1.25, 1.0, 0.83, 0.71, 0.63, 0.55, 0.50, 0.4$ , and 0.33 ms. Reference spectra were acquired by omitting the CPMG period. Each 2D spectrum was collected with eight scans per FID unless otherwise stated.

**SIFT Processing and Analysis of  $R_2$  Dispersion Profiles.** Resonance peaks were automatically picked in the reference spectrum using the program Sparky, followed by manual removal of obvious junk peaks. The peak list was imported into MATLAB and used during SIFT processing to define the peak positions with a fixed box size ( $\pm 0.94$  by  $\pm 0.07$  ppm along the  $^{15}\text{N}$  and  $^1\text{H}$  axes, respectively) around the peak-top. The input for SIFT processing was the NUS time-domain NMR data, the NUS schedule, the peak list, and specification of the SIFT cycle number (= 8 in this study). All steps, including the specification of bright regions, SIFT cycle, final processing with apodizations and FT, and compilation of the peak intensities in 42 spectra, were fully automated, and involved absolutely no user intervention. This entire process took 0.6 min on two 2.26 GHz quad-core Intel Xeon processors.

The integral peak intensities in all spectra were obtained using MATLAB as a sum of the intensities at  $3 \times 3$  grid points centered on the picked peak top. The compiled peak intensities were converted into effective  $R_2$  relaxation rates ( $R_2^{\text{eff}}$ ) using the equation  $R_2^{\text{eff}}(\tau_{\text{CP}}) = (-1/T_{\text{CPMG}}) \ln(I(\tau_{\text{CP}})/I_0)$ , where  $I(\tau_{\text{CP}})$  represents the peak intensity at a particular  $\tau_{\text{CP}}$  delay, and  $I_0$  represents the intensity in the reference spectrum. An error in peak intensity,  $\varepsilon_I$ , was evaluated from the standard deviation of the noise intensity in signal-containing  $F_1$  slices. Thus, in pre-SIFT spectra, the errors were dominated by strong sampling noise due to NUS. Errors in the peak intensities were then propagated to those of  $R_2^{\text{eff}}$  as:  $\varepsilon_R = \varepsilon_I / (T(\tau_{\text{CP}}) \times \tau_{\text{CP}})$ .<sup>28</sup>

The  $R_2$  dispersion profiles of 55 nonoverlapping peaks were fit to the Carver and Richards equation<sup>29</sup> derived for a two-state conformational exchange model ( $A \leftrightarrow B$ ) using the program GLOVE.<sup>8</sup> The fitting parameters were the population-average intrinsic relaxation rate,  $R_2^0$ , the exchange rate constants,  $k_{\text{ex}}$  ( $k_{\text{ex}} = k_{\text{AB}} + k_{\text{BA}}$ ), the chemical shift differences between states,  $|\Delta\omega|$ , and the product of the populations of states A and B,  $p_A p_B$ . The population of state B, designated as the lower-populated state, was calculated according to the formula  $p_B = (1 - (1 - 4p_A p_B)^{1/2})/2$ . The  $R_2$  dispersion profiles recorded at the two external fields were fit simultaneously for each residue. Fits were initially performed for individual residues, and the goodness-of-fit was assessed by the reduced  $\chi^2$  value ( $\chi^2$  divided by the degrees of freedom). The resulting  $k_{\text{ex}}$  and  $p_A p_B$  values were consistent for all residues with the two-state conformational exchange model, indicating that all sites experience an identical kinetic process.<sup>30</sup> The  $R_2$  dispersion curves were therefore fit with the global parameters  $k_{\text{ex}}$  and  $p_A p_B$  for all residues. Uncertainties in the parameters were estimated by Monte Carlo simulations, using 100 synthetic data sets generated on the basis of the experimental uncertainties.<sup>28</sup>

Root-mean-square (rms) deviation of the  $R_2$  dispersion profile obtained from NUS data from the full data “true” profile was calculated as

$$\Delta R_2^{\text{eff}} = \sum_{B_0} \left\{ \sqrt{\frac{\sum_{\tau_{\text{CP}}} (R_{2,\text{Full}}^{\text{eff}}(\tau_{\text{CP}}, B_0) - R_{2,\text{NUS}}^{\text{eff}}(\tau_{\text{CP}}, B_0))^2}{n}} \right\} / 2 \quad (2)$$

where  $n$  represents the number of  $R_2^{\text{eff}}$  rates in a profile.



**Nonuniform Sampling.** We collected 64 uniform  $t_1$  samples for all the spectra, including the reference spectrum, and then omitted the samples to emulate NUS. This allowed us to assess the discrepancy of the exchange parameters obtained from NUS data in comparison to the “true values” derived from the full data set. In practical applications, all spectra would be acquired with NUS to expedite data collection; however, for the most accurate specification of peak-containing “bright spectral regions”, it is recommended to use uniform sampling for the reference spectrum.

We generated a NUS schedule as described previously,<sup>18</sup> using a standard NUS generator.<sup>10</sup> The NUS generator uses a stratified jittered algorithm to constrain the average distances between random sample points, and Gaussian weighting function for the probability density of the samples. The constraint tries to reduce the NUS-noise before any processing is applied, thus it has no bias for SIFT or other NUS processing methods. We used  $\sigma = 0.6$  for the width of the Gaussian probability distribution; i.e., the inflection point for the decaying probability occurs at  $\sim 60\%$  of the maximum  $t_1$  evolution length.

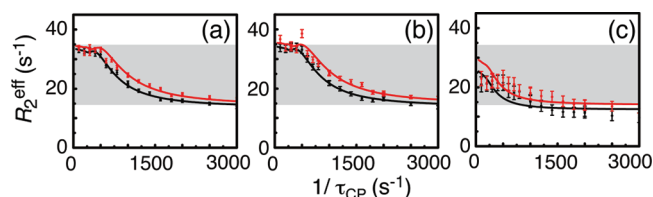
We used a NUS distributing 25  $t_1$  samples over 64 Cartesian grids in the time domain with Gaussian probability density; i.e.,  $\sim 60\%$  of the observations were omitted. The sparse level of NUS was decided from the following consideration. The proportion of the spectral dark regions in each slice along the  $F_1$  axis, or the spectral “darkness”, varied from 60% to 95% for our KIX data (Figure S1a, Supporting Information). Since SIFT can restore a NUS data set most accurately when the spectral darkness is greater than sparseness of the NUS,<sup>18</sup> the most conservative approach would be to adopt a NUS whose sparseness is matched to the available minimum spectral darkness ( $\sim 60\%$  in this case).

## RESULTS AND DISCUSSION

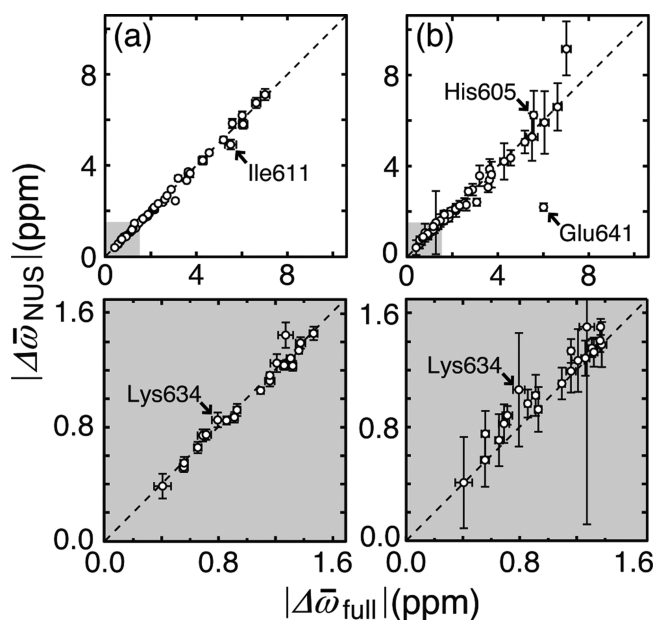
To validate the new SIFT protocol, we applied it to  $R_2$  relaxation dispersion measurements for KIX, which is a modular transcriptional coactivator known to undergo a functionally important two-state conformational exchange between the native and non-native forms.<sup>30</sup>  $R_2$  dispersion spectroscopy, one of the most powerful tools in the field of protein dynamics, can elucidate the kinetics and thermodynamics of protein motions,<sup>3</sup> as well as conformational information on weakly ( $>0.5\%$ ) and transiently ( $\sim$ ms) populated states.<sup>31</sup>

The reference  $^{15}\text{N}$  HSQC spectrum of KIX is shown in Figure 2a. All regions outside the peak positions (rectangles in Figure 2a) were treated as dark regions for SIFT processing. With this exhaustive use of information in a native spectrum, oversampling for the indirect acquisition dimension was not necessary in this work, although it is another valid way for sourcing dark information when peak positions are not known.<sup>18</sup> One may also combine the new procedure with oversampling.

The data shown in Figure 2 demonstrate the power of SIFT processing. For the serial 2D measurements, we used a NUS that randomly distributes 25  $t_1$  samples over 64 Cartesian grids in the time domain; i.e.,  $\sim 60\%$  of the observations were omitted. The spectrum before SIFT, which is equivalent to that obtained with NLFT,<sup>10</sup> was highly corrupted with NUS noise (Figure 2b). The peak intensities were not only severely suppressed due to the sparse time-domain data, but their relative intensities to each other were also upset due to the overlapping NUS noise (Figure 2d, bottom). Therefore, this spectrum is clearly not suited for the present application. After SIFT, the NUS noise was removed (Figure 2c), and the peak intensities were accurately



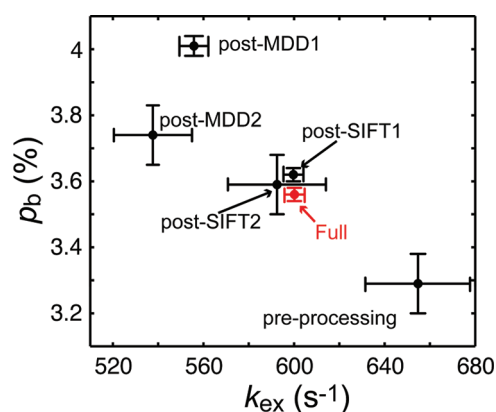
**Figure 3.**  $R_2$  dispersion profiles obtained for Glu641 at  $B_0 = 14.1$  T (black) and 17.6 T (red) from the full (a), post- (b), and pre-SIFT spectra (c). The best-fit profiles are superposed in the same color. The rms deviations from the true profile,  $\Delta R_2^{\text{eff}}$ , were  $7.9 \text{ s}^{-1}$  and  $1.1 \text{ s}^{-1}$  for the pre- and post-SIFT profiles, respectively. A part of the plot between the highest and lowest  $R_2^{\text{eff}}$  values of the full data is shaded in each panel to facilitate comparison. More examples are shown in Figure S3.



**Figure 4.**  $^{15}\text{N}$  chemical shift differences ( $|\Delta\omega|$ ) for two interconverting states extracted from NUS data (vertical axis) with (a) and without (b) SIFT-processing are plotted against “true” values from the full data (horizontal axis). The coefficient of determination was 0.873 for pre- and 0.993 for post-SIFT data. The shaded regions in panels a and b are expanded in the panels below. For the data points labeled with a residue number, the corresponding  $R_2$  dispersion profiles are shown in Figures 3 or S3.

restored (Figure 2d, middle). Indeed, peak intensities measured in the post-SIFT spectra exhibited an excellent correlation to the “true” values observed in the full spectrum (Figure S2). SIFT took only 0.6 min for processing 42 input 2D spectra together with a full compilation of 2310 peak intensities. In addition to this unparalleled computational efficiency, the SIFT process requires neither user intervention nor parameter tweaking for the process. The total data acquisition time was compressed to 7.2 h with a recycle delay of 3 s, down from the 18 h conventionally needed on each magnetic field, with little loss in spectral resolution and accuracy of observed peak intensities.

The accurate peak intensities measured in the post-SIFT spectra were translated into accurate  $R_2^{\text{eff}}$  values, hence the accurate  $\tau_{\text{CP}}$ -dependence of  $R_2^{\text{eff}}$ , or the  $R_2$  dispersion profiles. As expected, the rms deviations of the  $R_2$  dispersion profiles from the “true” profiles given by the full spectra,  $\Delta R_2^{\text{eff}}$  (eq 1), were



**Figure 5.** Comparison of the exchange rate and population obtained from NUS data set processed in various ways, relative to the “true” values obtained from the full data set shown in red. For the data labeled as “post-SIFT 1” and “post-SIFT 2”, the number of scans was eight and two, respectively. For the data labeled as “post-MDD 1” and “post-MDD 2”, the parameter sets #1 (number of scans = 8, number of component = 20, subregion size = 0.3 ppm,  $\lambda$  = 0.05) and #2 (number of scans = 8, number of component = 12, subregion size = 0.3 ppm,  $\lambda$  = 0.005) were used, respectively.

generally smaller for the post-SIFT profiles than for the pre-SIFT profiles (Table S1). Figure 3 shows representative  $R_2$  dispersion profiles obtained for Glu641.

Fitting of the accurate post-SIFT profiles resulted in accurate exchange parameters. Figure 4 shows plots of chemical shift differences  $|\Delta\bar{\omega}| (= |\Delta\omega/2\pi B_0|)$  for the two interconverting states obtained by fitting the pre- or post-SIFT profiles against the “true” values. Similar correlations were obtained for the intrinsic relaxation rates  $R_2^0$  (Figure S4). For both  $|\Delta\bar{\omega}|$  and  $R_2^0$ , uncertainties and deviations from the “true” values were larger with the pre-SIFT data. Two global parameters extracted from the post-SIFT profiles (the exchange rate,  $k_{\text{ex}} = 599.72 \pm 4.18 \text{ s}^{-1}$ , and the population of the state B,  $p_B = 3.62 \pm 0.02\%$ ) were in good agreement with the “true” values ( $k_{\text{ex}} = 600.17 \pm 4.18 \text{ s}^{-1}$ ,  $p_B = 3.56 \pm 0.02\%$ ), while those from the pre-SIFT profiles ( $k_{\text{ex}} = 654.74 \pm 22.9 \text{ s}^{-1}$ ,  $p_B = 3.29 \pm 0.09\%$ ) were inaccurate and imprecise (Figure 5).

The excellent robustness of SIFT in processing noisy data as shown before<sup>17</sup> should allow for accurate relaxation measurements in a noisier data set. With this in mind, we analyzed data recorded with only two scans per FID instead of eight, as typically employed for moderately concentrated samples or for a standard phase cycle. Despite a factor of 2 deterioration in the signal-to-noise (S/N) ratio compared with the eight-scan data shown above, the S/N ratio for the weakest signals found in the spectra taken at the longest  $\tau_{\text{CP}}$ , the delay time between two successive  $180^\circ$  pulses in the CPMG pulse train, was more than  $\sim 10$  (median S/N was 35) with the  $500 \mu\text{M}$  sample. This was high enough for accurate peak quantification, and resulted in accurate  $R_2$  dispersion profiles after SIFT processing (Figure S5), with fairly accurate dynamics parameters:  $k_{\text{ex}} = 592.50 \pm 21.5 \text{ s}^{-1}$ ;  $p_B = 3.59 \pm 0.09\%$  (Figure 5). Indeed, SIFT has been shown to remain quantitative for data with even lower sensitivity,<sup>17</sup> thus it is compatible with relatively dilute samples. The approach requires the use of field gradient pulses and precisely calibrated RF pulses to suppress experimental artifacts, but yields another factor of 4 acceleration of the measurement, leading to the total measurement time of  $< 2 \text{ h}$ , or allows for a half dilution of the

sample (i.e.,  $250 \mu\text{M}$  KIX) while minimally affecting the accuracy of the measurement. To our best knowledge, the lowest protein concentration used for  $R_2$  dispersion analysis is  $300 \mu\text{M}$ <sup>32</sup> (typically  $\sim 1 \text{ mM}$ ), using the conventional uniform data sampling. With an introduction of NUS and SIFT, our result establishes accurate and significantly faster data acquisition at the lowest ever protein concentration, thus paves the way to successful analyses of dilute and unstable proteins.

The S/N ratios of the signals used in the present analysis ranged between 10 and 430. The high accuracy of the extracted relaxation parameters confirms the quantitative and unbiased nature of SIFT for signals with very low to high S/N ratios, with a dynamic range of more than 40.

There are a few applications of Forward MaxEnt and MDD to quantitative analyses. Since Forward MaxEnt, a quantitative version of MaxEnt, still awaits a breakthrough in the computational efficiency before being routinely applied to cases similar to the present study (for example, it takes 1–5 h for processing individual 2D data),<sup>12,33</sup> we here compared SIFT to MDD. It is noteworthy that MDD so far has not been applied to relaxation measurements where data are recoded in NUS mode. Figure S6 and Table S1 demonstrate that most of the post-MDD profiles resulted in larger  $\Delta R_2^{\text{eff}}$  than those for the post-SIFT profiles. The inaccuracy is partly due to the NUS noises remaining after processing, which could not be removed with an extensive parameter optimization, by recursive processing<sup>15</sup> or by coprocessing of  $F_1$  and  $F_2$ . Different sets of MDD parameters resulted in different values for the extracted exchange parameters, but they were similarly distant from the “true” values (Figure 5). In each MDD trial, processing time was typically  $\sim 1.5 \text{ min}$  per spectrum, and took  $\sim 1 \text{ h}$  for processing 42 spectra. MDD is primarily based on the recorded sparse NUS data for modeling a spectrum, while SIFT uses explicit information on the location of dark frequency points found independently after the experiment to restore the missing time points, and leaves existing time-domain data unchanged. Thus, when the number of measurements is small relative to the number of peaks to model, typical for low-dimensional spectra of complex molecules, SIFT is compared favorably with MDD for a quantitative analysis.

A preliminary test suggested that the present approach is also valid for larger systems. For a protein with 283 residues, we observed the minimum spectral darkness of  $\sim 50\%$  as compared with  $\sim 60\%$  for KIX (Figure S1). Since signals tend to cluster, rather than neatly filling up the spaces between peaks, the spectral darkness seems to decrease only slowly with the number of residues.

## CONCLUSION

We have introduced NUS and advanced SIFT processing into a protein dynamics study, which represents a novel application of NUS to cases where highly accurate peak quantification is demanded. To validate the method, we applied it to  $R_2$  relaxation dispersion measurements for the KIX domain (87 residues) of mouse CBP. A total 10-fold reduction in the measurement time afforded by minimizing the number of scans per FID (4-fold) using the robustness of SIFT in processing noisy data, and the number of data samples via NUS (2.5-fold) enabled us to complete  $R_2$  dispersion measurements within 2 h with a recycle delay of 3 s at each  $B_0$ , while preserving analytical accuracy. The approach is also equally valid for other NMR measurements described in the Introduction for probing protein motions,<sup>2,4,5</sup> and is considered to be valid for much larger systems. Not only

does this method shorten the measurement time, the ability to collect data rapidly may also obviate time- and labor-intensive sample screening/modification processes currently required for finding stable and soluble constructs. Alternatively, the saved instrumental time can be used to increase the number of scans per FID for improving sensitivity. Altogether, NUS and SIFT would significantly widen the applicability of NMR-based dynamics studies to more difficult systems with limited sensitivity and stability, enabling these studies to encompass a greater diversity of biological processes than are currently accessible. An attractive example would be a protein dynamics study in living cells,<sup>6</sup> which we are considering for our future work.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Reference spectrum of KIX and its “darkness”; correlation plots for peak intensities and intrinsic relaxation rate,  $R_2^0$ , obtained from the post-SIFT and pre-SIFT spectra; additional  $R_2$  dispersion profiles obtained with two scans per FID; comparison of  $R_2$  dispersion profiles; and rms errors from true profiles obtained with SIFT- and MDD-processed spectra. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [sugase@sunbor.or.jp](mailto:sugase@sunbor.or.jp); phone: +81-75-962-7484; fax: +81-75-962-2115.

## ■ ACKNOWLEDGMENT

This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas and a Grant-in-Aid for Young Scientists (B) for K.S. from the MEXT of Japan.

## ■ REFERENCES

- (1) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679–6685.
- (2) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Formankay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984–6003.
- (3) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.
- (4) Hwang, T. L.; van Zijl, P. C. M.; Mori, S. *J. Biomol. NMR* **1998**, *11*, 221–226.
- (5) Dempsey, C. E. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *39*, 135–170.
- (6) Sakakibara, D.; Sasaki, A.; Ikeya, T.; Hamatsu, J.; Hanashima, T.; Mishima, M.; Yoshimasu, M.; Hayashi, N.; Mikawa, T.; Walchli, M.; et al. *Nature* **2009**, *458*, 102–106.
- (7) Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. *Nature* **2004**, *430*, 586–590.
- (8) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021–1025.
- (9) Jeannerat, D. *J. Magn. Reson.* **2007**, *186*, 112–122.
- (10) Kazimierczuk, K.; Zawadzka, A.; Kozminski, W.; Zhukov, I. *J. Magn. Reson.* **2007**, *188*, 344–356.
- (11) Stanek, J.; Kozminski, W. *J. Biomol. NMR* **2010**, *47*, 65–77.
- (12) Hyberts, S. G.; Heffron, G. J.; Tarragona, N. G.; Solanky, K.; Edmonds, K. A.; Luithardt, H.; Fejzo, J.; Chorev, M.; Aktas, H.; Colson, K.; et al. *J. Am. Chem. Soc.* **2007**, *129*, 5108–5116.
- (13) Mobli, M.; Maciejewski, M. W.; Gryk, M. R.; Hoch, J. C. *Nat. Methods* **2007**, *4*, 467–468.

- (14) Hiller, S.; Ibragimov, I.; Wagner, G.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2009**, *131*, 12970–12978.
- (15) Jaravine, V.; Ibragimov, I.; Orekhov, V. Y. *Nat. Methods* **2006**, *3*, 605–607.
- (16) Tugarinov, V.; Choy, W. Y.; Orekhov, V. Y.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 622–627.
- (17) Matsuki, Y.; Eddy, M. T.; Griffin, R. G.; Herzfeld, J. *Angew. Chem., Int. Ed.* **2010**, *49*, 9215–9218.
- (18) Matsuki, Y.; Eddy, M. T.; Herzfeld, J. *J. Am. Chem. Soc.* **2009**, *131*, 4648–4656.
- (19) Gerchberg, R. W. *Opt. Acta* **1974**, *21*, 709–720.
- (20) Papoulis, A. *IEEE Trans. Circuits Syst.* **1975**, *22*, 735–742.
- (21) Diffrancia, G. T. *J. Opt. Soc. Am.* **1955**, *45*, 497–501.
- (22) Jorge, P.; Ferreira, S. G. *IEEE Trans. Signal Process* **1994**, *42*, 2596–2606.
- (23) Plevritis, S. K.; Macovski, A. *Magn. Reson. Med.* **1995**, *34*, 686–693.
- (24) Stern, A. S.; Donoho, D. L.; Hoch, J. C. *J. Magn. Reson.* **2007**, *188*, 295–300.
- (25) Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741–752.
- (26) Tollinger, M.; Skrynnikov, N. R.; Mulder, F. A.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 11341–11352.
- (27) Loria, J. P.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1999**, *121*, 2331–2332.
- (28) Ishima, R.; Torchia, D. A. *J. Biomol. NMR* **2005**, *32*, 41–54.
- (29) Carver, J. P.; Richards, R. E. *J. Magn. Reson.* **1972**, *6*, 89–105.
- (30) Tollinger, M.; Kloiber, K.; Agoston, B.; Dorigoni, C.; Lichtenecker, R.; Schmid, W.; Konrat, R. *Biochemistry* **2006**, *45*, 8885–8893.
- (31) Korzhnev, D. M.; Religa, T. L.; Banachewicz, W.; Fersht, A. R.; Kay, L. E. *Science* **2010**, *329*, 1312–1316.
- (32) Kay, L. E.; Sprangers, R.; Gribun, A.; Hwang, P. M.; Houry, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16678–16683.
- (33) Hyberts, S. G.; Frueh, D. P.; Arthanari, H.; Wagner, G. *J. Biomol. NMR* **2009**, *45*, 283–294.