Ultra High Resolution ^{1}H-^{13}C HSQC Spectra of Metabolite Mixtures using non-linear sampling and Forward Maximum Entropy (FM) Reconstruction

Sven G. Hyberts, Gregory Heffron, Nestor Tarragona, Kirty Solanky, Katie Edmonds, Harry Luithardt, Jasna Fejzo, Michael Chorev, Husseyin Aktas, Kenneth Falchuck, Jose Halperin and Gerhard Wagner*

Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, Boston, MA 02115.

*To whom correspondence should be addressed:
Gerhard Wagner
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
240 Longwood Avenue
Boston, Massachusetts 02115
USA
Tel: 617-432-3213
Fax: 617-432-4383
Email: gerhard_wagner@hms.harvard.edu

KEYWORDS. NMR, non-linear sampling, maximum entropy reconstruction, data processing, metabolomics

ABSTRACT.

To obtain a comprehensive assessment of metabolite levels from extracts of leukocytes we have recorded ultra-high-resolution ^{1}H-^{13}C HSQC NMR spectra of cell extracts, which exhibit spectral signatures of numerous small molecules. However, conventional acquisition of such spectra is time consuming and hampers measurements on multiple samples, which would be needed for statistical analysis of metabolite concentartions. Here we show that the measurement time can be dramatically reduced without loss of spectral quality when using non-linear sampling (NLS) and a new high-fidelity Forward Maximum-entropy (FM) reconstruction algorithm. This FM reconstruction conserves all measured time domain data points and guesses the missing data points by an iterative process. This consists of discrete Fourier transformation of the sparse time-domain data set, calculating the spectral entropy, determination of a multidimensional entropy gradient, and calculation of new values for the missing time domain data points with a conjugate gradient approach. Since this procedure does not alter measured data points it reproduces signal intensities with high fidelity and does not suffer from a dynamic-range problem. We show that a high-resolution ^{1}H-^{13}C HSQC spectrum with 4k complex increments recorded within four days can be reconstructed from $1/7^{th}$ of the increments with nearly identical spectral appearance, indistinguishable signal intensities and comparable or even lower root mean square (rms) and peak noise pattern. Thus, this approach allows recording of high-resolution ^{1}H-^{13}C HSQC spectra in $1/7^{th}$ of the time needed for recording linearly sampled spectra.

Abbreviations: NMR - nuclear magnetic resonance; non-linear sampling –NLS; HSQC – heteronuclear single-quantum coherence;

Introduction

Comprehensive measurements of the concentrations of large numbers of metabolites can provide detailed insights into the state of cells (Lindon et al., 2004; Nicholson et al., 1999). This has the potential of being used to diagnose disease, to follow the effect of drug treatment, or to study toxicity. Comparison of metabolite samples from different groups, such as healthy and sick individuals, or normal and transformed cells may lead to identification of biomarkers that can be invaluable for targeted disease diagnosis or for understanding metabolic pathways and mechanisms of disease. Metabolic profiling of cell lines may lead to new insights into metabolic pathways and their alteration in disease.

Assessment of metabolite levels in metabolomics studies has primarily relied on mass spectroscopy, NMR spectroscopy, chromatographic separation techniques and various multivariate data analysis techniques (El-Deredy, 1997; Lindon et al., 2004; Nicholson et al., 1999). Among the NMR methods, one dimensional (1D) ^{1}H spectroscopy is most commonly used, where the spectrum is divided in a limited number of buckets, and the signal intensities of the buckets are compared between multiple samples using principle component analysis, partial least square discriminant analysis (PLS-DA) or other statistical approaches in order to separate groups of samples and to identify the metabolites (biomarkers) that are most different between the groups (Gavaghan et al., 2002). Alternatively, 1D NMR spectra have been fitted to databases of metabolite reference spectra to obtain concentrations of the most abundant molecules in a metabolite mixture (Weljie et al., 2006). Occasionally, two-dimensional NMR spectra have been used to enhance the spectroscopic resolution for more detailed metabolite identification (Liang et al., 2006; Keating et al., 2006; Dunn et al., 2005; Rhee et al., 2004; Frederich et al., 2004). Resonances of metabolites typically have very long transverse relaxation times so that 2D NMR spectra, such as ^{1}H-^{13}C HSQC or ^{1}H-^{1}H TOCSY, can be recorded at very high resolution and can resolve nearly all metabolite signals **(Heffron et al.)**. However, this requires a large number of increments, which makes data acquisition very time consuming and impractical for recording spectra from multiple samples as is necessary for statistical analysis.

It has been proposed and experimentally verified in the past that measuring times of 2D and can be reduced with non-linear sampling in the indirect dimensions (Barna and Laue, 1987; Hoch et al., 1990). Obviously, non-linearly-sampled time-domain

data require processing methods different from the discrete Fourier transform (DFT). Non-linear sampling was originally proposed by (Barna et al., 1987), and the data were processed with a maximum entropy method relying on the Burg algorithm (ref#). Subsequently, non-linear sampling was seriously pursued by Hoch and collaborators (Hoch, 1989). Processing software was developed that relied on an alternative maximum entropy (MaxEnt) reconstruction algorithm that could handle phase-sensitive data and included adjustable parameters to tune the outcome of the data reconstruction. This software is available through the Rowland NMR Toolkit (RNMRTK) (Hoch, 1989) and has been successfully applied for processing 2D non-linearly sampled COSY spectra (Schmieder et al., 1993), constant-time evolution periods of triple-resonance data (Schmieder et al., 1994), or quantification of HSQC spectra for relaxation experiments (Schmieder et al., 1997a). Development of a two-dimensional MaxEnt reconstruction procedure designed to run in parallel on workstation clusters (Li et al., 1998) has stimulated several applications to explore optimum evolution times (Rovnyak, Hoch et al., 2004), to rapidly acquire complete sets of triple resonance experiments for sequential assignments (Rovnyak, Frueh et al., 2004), to facilitate side-chain assignments (Sun, Hyberts et al., 2005), and to enable high-resolution triple resonance experiments (Sun, Frueh et al., 2005; Frueh et al., 2006).

The principle advantages of non-linear sampling are increasingly recognized (Tugarinov et al., 2005). Besides Maximum Entropy reconstruction, other methods are used for processing non-linearly recorded spectra, such as the maximum likelyhood method (MLM) (Chylla and Markley, 1995), a Fourier transformation of nonlinearly spaced data using the Dutt-Rokhlin algorithm (Marion, 2005), or multi-dimensional decomposition (MDD) (Korzhneva et al., 2001; Orekhov et al., 2001; Orekhov et al., 2003; Gutmanas et al., 2002).

A special case for reducing the sampling time sampling for 3D or higher-dimensional data is used with sampling along various angles of the indirect sampling space (Kupce and Freeman, 2004b, a), and previously developed projection-reconstruction procedures (Hounsfield, 1973) are used for processing the data. Originally, this had problems of artifacts from crossing shadows of projected lines, which may have been overcome with time-domain-directed procedures **{Kupce, private communication}**. Another strategy for minimizing sampling times is the reduced dimensionality approach (Szyperski et al., 1993), which has been further developed into the GFT method (Kim and Szyperski, 2003) and applied to proteins up to 21 kDa (Liu et

al., 2005). However, to our knowledge, the projection reconstruction method and the GFT approach have primarily been applied to rather small systems where low sensitivity is not an issue. Unlike MaxEnt reconstruction and MDD, this class of approaches is only suited for 3D and higher dimensional experiments but not applicable to shortening acquisition of 2D experiments.

In the past, we have used the MaxEnt reconstruction procedure of the Rowlan NMR Toolkit (RMNTK) (Hoch et al., 1990). It is very efficient and ideally suited for handling non-linearly sampled data with a low dynamic range, such as triple-resonance spectra where all peaks have similar intensities. However, processing of data with a high dynamic range, such as in NOESYs, TOCSYs, mixtures of metabolites, spectra with diagonals or peaks close to the noise level seems to suffer from effects of non-linearity of peak intensities. This has previously been recognized and remedies for correcting intensities have been suggested for MaxEnt reconstructions of relaxation data (Schmieder et al., 1997b). However, there remains the problem of losing weak peaks that are weak and barely above noise level, which is a significant problem in spectra of metabolite mixtures with large variations peak intensities.

As another strategy to cope with the dynamic range problem we will examine whether this issue can be eliminated by using a MaxEnt-related approach. In the classic MaxEnt algorithm (Hoch and Stern, 2001), the linearity of the reconstruction depends of the parameter λ (lambda). The parameter λ is the Lagrange multiplier, which is required to create the objective function $Q(\mathbf{f}) = S(\mathbf{f}) - \lambda C(\mathbf{f,d})$. $C(\mathbf{f,d})$ is the constraining function between the measured time domain data points and the time domain data points calculated from the inverse Fourier transformed iteratively "guessed" mock spectra. By the nature of altering the spectra in the frequency domain, C(f,d) is always > 0.

Here we present a new procedure termed **F**orward **M**aximum entropy (FM) reconstruction for processing non-linearly sampled 2D NMR data. Similarly to the MaxEnt reconstruction of Hoch and Stern (Hoch et al., 1990), it aims to minimize a target function that contains the negative entropy (thus maximizes the entropy). However, in contrast to the MaxEnt reconstruction procedure (Hoch and Stern, 2001), which allows a variation of the measured data points by minimizing the target function $Q(\mathbf{f}) = S(\mathbf{f}) - \lambda C(\mathbf{f,d})$, the FM reconstruction does not, which increases the fidelity of peak intensity reconstruction. By altering the method such that the "guessing" occurs in the time domain, the constraining function $C(\mathbf{f,d})$ can easily be set to zero – practically, alteration is only done for non-obtained data points of a non-uniformly acquired FID – the

requirement of the Lagrange multiplier thus vanishes. The objective function is now reduced to $Q'(\mathbf{f}) = S(\mathbf{f})$. The implementation hence requires the "mock" data to be constantly "guessed" in the time domain, then forward Fourier transformed, where they are scored by only calculating the implemented Entropy function. This approach doesn't seriously bias against weak signals. We validate this approach with a ^{1}H-^{13}C HSQC spectrum of metabolites from cell extracts recorded with 8k increments. Acquisition of the linearly sampled reference spectrum required four days on a 600 MHz spectrometer. We show that the same quality of spectrum can be obtained within fourteen hours using non-linear sampling. This makes possible recording of multiple ultra high-resolution ^{1}H-^{13}C HSQC spectra needed for statistical analysis of metabolomics data.

Materials and Methods

Principles of the Forward Maximum Entropy (FM) Reconstruction

We pursue to reconstruct a time-domain data set $F(d) = \{d_i; i = 1.. N\}$ where a subset of points $\{d_k; k = 1 ..M\}$ ($M < N$) has been measured experimentally but all other points are unknown. We guess the unknown data points in the time domain, calculate the spectrum with fast Fourier transform (FFT) algorithm and calculate its entropy:

$$S(f) = -\sum_{k=1}^{N} f_k \log f_k$$

Because many of the data points are fixed and given by the experimental data set d, we may use a simplistic approach of the Entropy $S(f)$ for complex data points f_k, namely

$$S(f) = -\sum_{k=1}^{N} |f_k| \log |f_k| .$$

We pursue to minimize $Q(f) = - S(f)$ iteratively using the Forward Maximum Entropy (FM) reconstruction method proposed here that maintains the experimental time domain data points unaltered. $Q(f)$ is related to the total intensity of a spectrum, and minimizing it reduces the total signal and noise subject to maintaining the measured time-domain data points.

The FM reconstruction program starts by setting the missing time-domain data points to zero. The data are then transformed with the fast Fourier transform (FFT) algorithm, and the entropy of the resulting spectrum is calculated. Next, the guessed time-domain data points are changed by ± d, followed by FFT and calculation of the entropy. From the resulting entropies a multidimensional gradient of Q_F, $\vec{\nabla}Q_F$, is calculated with respect to the missing data points d_j. Using this gradient we maximize the entropy with the Polak-Ribiere conjugate gradient method **(Gingras et al.)** to gradually change the guesses of the missing time-domain data points d_j. The data points in the time domain data set, are separated into two independent variable sets $\hat{r}^+$ and $\hat{i}^+$, where $\hat{d}_j^+ = (\hat{r}_j^+ , \hat{i}_j^+)$. This creates a total of $2*(N-M)$ free vectors of $\vec{\nabla}Q_F$. For initialization purpose, all $\hat{r}_j^+ , \hat{i}_j^+$ are set to zero. The gradients are achieved numerically by independently varying the values $\hat{r}_j^+$ and $\hat{i}_j^+$ by a small value δ. The individual gradients can hence be written as:

$$\frac{\partial Q_F(\hat{d})}{\partial[\hat{r}:\hat{i}]_j} = \frac{-S(FFT\{\hat{d},[\hat{r}:\hat{i}]_j+\delta\}) + S(FFT\{\hat{d},[\hat{r}:\hat{i}]_j-\delta\})}{2\delta}, \text{ where } M < j \le N .$$

Minimization is iterated until a cut-off criterion is reached, such as that the value of the target function Q_F doesn't decrease by more than the cut-off parameter.

The final result of the FM reconstruction procedure is a time-domain data set where the missing data points are filled in with the best guesses maximizing the entropy or minimizing Q_F. This time-domain data set can now be transformed with any available data processing software and can be manipulated with window functions, zero filling or linear prediction.

Practical implementation of FM reconstruction:

The package FFTW (http://www.fftw.org), version 3.0.1 was used as a C library for computing discrete Fourier transforms. FFTW is distributed under the GNU software licenses and is free software. The Polak-Ribiere conjugate gradient method,

gsl_multidim_fdsminimizer_conjugate_pr of the package GSL, vers 1.5 (GNU Scientific Library, http://www.gnu.org/software/gsl), was used for minimization purposes. For data handling, we used the open NMRPipe data format was used for describing NMR data. See http://spin.niddk.nih.gov/bax/softgware/NMRPipe for a description of NMRPipe. The resulting C program was compiled using gcc and the i686 instruction set on a Dell dual Xenon 3.0 GHz computer operating under Fedora Core 1 Linux OS.

Preparation of the metabolomics sample

A metabolite sample from the aqueous phase of cell extracts from mouse BaF3 cells carrying the gene for the Bcr-Abl kinase (Daley and Baltimore, 1988) was prepared as described elsewhere in more detail (Tarragona et al., 2006). Approximately 900 million cells were used.

NMR spectroscopy

NMR spectra were recorded on a Bruker Avance 600 spectrometer equipped with a cryogenic triple-resonance probe. A set of seven ^{1}H-^{13}C HSQC spectra were recorded with 4k increments (complex) and 16 scans per increment. A relaxation delay of 1.2 seconds was used between scans. Each of the seven spectra was recorded in 14 hours; the total measurement time for all seven 2D spectra was 4 days.

Results

Recording an ultra high resolution (UHR) ^{1}H-^{13}C HSQC spectrum

As a first step towards identifying metabolites in BaF3 cells we have recorded a 1D spectrum of the aqueous phase of cell extracts in $^{2}H_2O$, which is shown in **Fig. 1A**. Identification and measurement of concentrations of metabolites is severely hampered by spectral overlap. Similarly, 2D NMR spectra recorded with the typically 100 to 200 increments suffer from low resolution in the indirect dimension. Thus, we recorded, an ultra high resolution (UHR) ^{1}H-^{13}C HSQC spectrum with 4k increments (complex data points) (**Figure 1B**). The signal separation obtained with 4k complex increments resolves essentially all overlap as is shown with the expansion of the most crowded

region in **Fig. 1D**. The dispersion of the UHR spectrum promises to provide a tool for assigning nearly all metabolites that are present in sufficiently high concentrations and for measuring concentrations of the individual metabolites.

The 4k complex increments result in a maximum t_1 value of **4 s**, which is close to the estimated average T_2 value of metabolite carbon coherence. As stated previously, it is desirable to sample to around that long an evolution times to obtain optimal resolution and sensitivity (Rovnyak, Hoch et al., 2004). To further examine the optimal number of increments we transformed the data set using 4k, 2k, 1k and 512 complex t_1 values. **Figure 2** shows the effect on a small crowded region that is indicated with a box in **Fig. 1D**. The 4k and 2k data clearly resolve all peaks but the transformations of only 1024 and 512 increments do not. A cross section drawn through the strongest peak demonstrates that increased resolution also results in larger peak height. Thus, the relative height of the tallest peak is 3.9 : 3.0 : 2.0 : 1.0 in the 4k, 2k, 1k and 512 point transforms, respectively (**Fig. 2)**. Although the apparent resolution doesn't improve much by going from 2k to 4k complex data points in t1 the peak height increases by approximately 30%, which is close to $\sqrt{2}$ (41%) and consistent with the expectation (**Fig. 2**). Obviously, doubling the number of t_1 values from 2k to 4k doubles the measuring time, and the spectrum shown here required a total of four days of instrument time. This is undesirably long if one wants to measure multiple samples for determining statistically significant differences of metabolite concentrations between cell types. It is indeed the long-term goal of our research to identify metabolites with concentrations that differ between cell types. This includes comparison of normal and malignant cells, cells before and after treatment with inhibitors, or any other pairs of distinct cell types. Measurement of multiple samplesfor each cell type would be difficult with measurement times of four days as was used for the spectrum in **Fig. 1.**

To reduce the measurement time we have explored the non-linear sampling approach with the Forward Maximum Entropy (FM) reconstruction for data processing. We hypothesized that this approach allows recording of high-resolution 2D spectra within a reasonably short time, and developed the FM with the goal to avoid bias against weak peaks.

To test this approach we have recorded seven identical linearly sampled high resolution ^{1}H-^{13}C HSQC spectra acquired over a period of 14 hours each. Addition of the seven spectra and standard FFT yields the spectrum shown in **Fig. 1**. Recording all seven spectra required four days of instrument time. The high resolution achieved is

demonstrated by showing an expansion of a small portion of the most crowded spectral region (**Fig. 1C**). Numerous ^{1}H-^{13}C cross peaks are visible and all are very well resolved.

Impact of non-linear sampling with FM reconstruction on resolution and S/N

Recording spectra for as long as four days is rather impractical for metabolomics studies. Therefore, we tested whether non-linear sampling and a suitable processing routine would allow shortening the measurement time. Since spectra of mixtures of metabolites have a large variation of intensities we employed the FM reconstruction approach outlined above. Toprovide a basis for testing the FM algorithm we recorded seven identical data sets measured within 0.6 days each. This allowed us to compare spectra recorded linearly within 0.6 days with data recorded non-linearly within the same amount of time, with only 1/7th of the increments but seven times the number of scans per increment.

Figure 3 shows different versions of a representative cross section along the carbon dimension of the ^{1}H-^{13}C HSQC spectrum of **Figure 1**. **Figure 3A** provides transformations of all linearly acquired data points. The lower three traces are from spectra 2, 4 and 6 out of the seven spectra recorded for 14 hrs each, using 16 scans per increment. The top spectrum is the sum of all seven spectra, it represents a measuring time of four days with 7 x 16 scans per increment. It demonstrates the gain in signal to noise by a factor of $\sqrt{7}$ compared to the individual spectra, such as 2, 4 or 6 shown in the lower part of the figure. The top spectrum of **Fig. 3B** is the same as in **Fig. 3A**. The lower three traces, however, are FM reconstructions using only one 1/7th of the increments but 7 x 16 scans per increment. Thus each of the lower three traces represents the same total measuring time of 14 hours, equal to each of the lower traces of **Fig. 3A**. Three different sampling schedules, S1, S2 and S3 were used to pick increments from the total data set. For S1, increments were picked randomly and with constant probability along t1, S2 used an exponentially decreasing probability, and S3 applied a linearly decreasing pick rate. As can be seen, the quality of the three non-linearly sampled spectra, transformed with FM, is comparable to the top trace, which represents a seven-fold longer measurement time. There is no obvious bias in favor of strong peaks or against weak peaks. The three sampling schedule exhibit similar results but the schedule with linearly decreasing weight seems to have the best signal-to-noise ratio with a small margin.

To further assess the quality of the NLS spectra processed with FM reconstruction we analyzed the apparent noise in the linearly and non-linearly sampled spectra. **Figure 4** shows a small representative section of the 1H-13C HSQC spectrum that contains strong and weak peaks. And a cross sections through the strongest peak along the carbon dimensions were plotted so that the strongest signals had equal height (**Fig. 4 A**). The noise was measured in a region outside the range that contains signals. Both the root mean square (rms) noise and the peak noise were measured for the three linearly sampled spectra 2, 4 and 6, as well as the average of all seven linearly recorded spectra. As expected, both the rms and peak noise are approximately $\sqrt{7}$ lower when averaging the seven linearly sampled spectra.

Importantly, the FM reconstruction of the NLS data picked from the averaged data set have roughly the same peak noise as the full averaged data; it is lowest for the exponential schedule S2. For all three sampling schedules, the rms noise is approximately two-fold lower than in the transform of the full linear averaged data set. Thus, the FM processing of the NLS data sets, which can be acquired in 1/7th of the measuring time, yield high quality spectra comparable in quality to the DFT of the full linearly sampled averaged data set.

To analyze whether NLS with FM reconstruction affects the relative intensities of signals we measured peak heights of all detectable signals in the carbon cross section for 16 ^{13}C traces. In **Fig.5**, the peak heights in the linear averaged data set are plotted against the corresponding values in the NLS (random samplisng schedule S1) data set processed with FM reconstruction. Clearly, there is an excellent correlation, and there is no bias against weak peaks.

To further analyze this, we compare a small section of the ^{1}H-^{13}C HSQC spectrum shown in Fig. 1 obtained from the average of all seven linearly sampled spectra (**Fig. 6A**) and compare it with the NLS data selection of 1/7th the increments and processed with FM reconstruction (Fig. 6B).. Both the contour plot and the cross section are essentially identical although the NLS FM spectrum corresponds to 1/7th of the measuring time.

Discussion

NMR spectra of mixtures of metabolites as obtained from cell extracts contain a large number of signals. Since the metabolites have all narrow line widths the individual signals are resolvable in 2D NMR spectra but only when the experiments are recorded at very high resolution. Previously, we have argued signals should be recorded to about T_2* in order to obtain good resolution without significantly deteriorating signal to noise. (Rovnyak, Hoch et al., 2004). For spectra of metabolites with narrow line widths this requires sampling to long evolution times and needs very long measuring times. We have recorded a ^{1}H-^{13}C HSQC spectrum with 8k by 4k (real) data points in the t_1 and t_2 dimensions, respectively. Estimating that the spectrum contains as many as 1000 cross peaks, each peak is defined by 32k data points in the average. Thus, the signals are largely over-determined, and it should be possible to extract spectral parameters from a reduced data set, such as obtained with NLS.

We have developed a simple new algorithm that estimates missing time-domain data points of non-linearly sampled data by conjugate gradient minimization of the target function *Q(f)*, which is the negative entropy of the frequency spectrum, - *S(f)*. Obviously, *Q(f)* is a measure of the total amount of intensity and the target function would be minimal with all frequency data points equal to zero. Previously, an efficient Maximum Entropy reconstruction routine has been written that can handle linearly and non-linearly sampled data (Hoch and Stern). For handling of linearly sampled data it is necessary to allow for small variation of measured data points, and the balance between maintaining measured data points and maximizing the entropy is guided by setting adjustable parameters, def and lambda. Depending on the choice of these parameters, weak peaks may be de-emphasized.

The FM reconstruction algorithm presented here does not allow for variation of measured time-domain data points. Thus, it does not de-emphasize weak signals as long as they are represented in the measured data points. FM reconstruction does not require setting of parameters for the reconstruction. The only operator decision to be made is when to terminate the iterative minimization of the target function *Q(f)*. On the other hand, FM reconstruction is not suitable for and cannot be applied to linearly sampled data sets. The final FM result is a reconstructed time-domain data set that can be transformed with any of the available processing programs. Here we have used NMRPipe (Delaglio et al.) for all processing.

So far, we have used FM reconstruction for processing spectra with only one non-linearly sampled indirect dimension. We have used a standard conjugate gradient

minimizer, and processing is relatively slow as reconstruction of one 4k complex time-domain data set with $6/7^{th}$ of the data points missing takes about 3 and 5 hours on a Optron xx computer. Obviously, the processing time depends on the number of missing points, and shorter time domain data can be processed significantly faster. Processing of 2D spectra benefits for farming out the reconstruction of FIDs to processors of PC clusters. FM reconstruction is significantly slower than the MaxEnt routine used in the Rowland NMR Toolkit (RNMRTK), which uses an analytically calculated gradient for minimizing the target function.

In principle, it should be possible to apply FM reconstruction to higher-dimensional spectra with more than one non-linearly sampled dimension. This will require, however, exploring more efficient minimizer routines, such as the Monte Carlo method.

ACKNOWLEDGMENT:

This research was supported by NIH (Grants GM47467, DK020299 and EB002026). We thank Dr. Jeffrey Hoch for stimulating discussions on the topic of this publication.

FIGURE CAPTIONS.

Figure 1: NMR spectra of the aqueous fraction of cell extracts from mouse BaF3 cells in $^{2}H_2O$, pH 6.5, 25 degrees. A. One-dimensional ^{1}H NMR spectrum. B. ^{1}H-^{13}C HSQC spectrum recorded with 4k complex points in the indirect dimension. C and D. Expansion of the section indicated with the box in B and corresponding 1D spectrum.

Figure 2: Comparison of a small spectral region indicated with a box in Fig. 1D transformed with different numbers of increments. Clearly, only 2k and 4k complex points can resolve all peaks. While 2k complex increments seem to resolve all peaks, going to 4k complex points sharpens the peaks and increases peak hieight by approximately 30% (see arrows).

Figure 3: Comparison of representative cross sections along the ^{13}C direction at a proton frequency of 3.76 ppm. (A) Use of the full linearly sampled data. The bottom three traces are from the full linearly sampled data sets 2, 4 and 6. The top trace is from the average of all seven linearly sampled data sets. (B) The top trace is the same as in A. The three traces at the bottom, however, are obtained by selecting $1/7^{th}$ of the increments of the averaged data set and transformed with FM reconstruction. In the sampling schedule S1 $1/7^{th}$ of the increments were picked randomly from the averaged data set with equal density along t1. In schedule S2, $1/7^{th}$ of the increments were picked with exponentially decreasing density, and in schedule S3, $1/7^{th}$ of the increments were picked with linearly decreasing density.

Fig. 4: Comparison between the linearly sampled 4-day experiment and the NLS/FM reconstruction of a 14-hrs subset of increments. Both the 2D plot of a small portion of the spectrum and the cross section along the ^{13}C direction at the position of the strongest peak are nearly indistinguishable. This demonstrates the high fidelity of the FM reconstruction.

Figure 5: Comparison of rms noise (blue) and peak noise (red) of the cross sections along the 13C dimension. Both rms and peak noise are measured outside of the region that contains signals. The columns labeled 2, 4 and 6 represent three of the seven linearly sampled spectra of 14 hours duration. The column ave shows rms and peak noise for the average of the seven linearly sampled spectra and corresponds to four

days of data acquisition. As expected, the noise levels decrease by $\sqrt{7}$. S1, S2 and S3 show the measured noise values for three sampling schedules of the averaged spectra but using only 1/7th of the increments. S1 is a randomly distributed schedule, S2 is sampled with exponentially decreasing sampling frequency, and S3 corresponds to a linearly decreasing ramp.

Fig. 6: Comparison of peak heights between the linearly sampled 4-day experiment and the randomly sample 14-hour experiment. The latter provides a high-fidelity reproduction of the peak heights in the full linearly sampled spectrum.

References

Barna, J. C. J. and Laue, E. D. (1987). *J. Magn. Reson.* **75**, 384-389.

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

Chylla, R. A. and Markley, J. L. (1995). *J Biomol NMR* **5**, 245-258.

Daley, G. and Baltimore, D. (1988). *Proc Natl Acad Sci U S A.* **85**, 9312-9316.

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

Dunn, W. B., Bailey, N. J. and Johnson, H. E. (2005). *Analyst* **130**, 606-625.

El-Deredy, W. (1997). *NMR Biomed.* **10**, 99-124.

Frederich, M., Cristino, A., Choi, Y. H., Verpoorte, R., Tits, M., Angenot, L., Prost, E., Nuzillard, J. M. and Zeches-Hanrot, M. (2004). *Planta Med* **70**, 72-76.

Frueh, D. P., Sun, Z. Y., Vosburg, D. A., Walsh, C. T., Hoch, J. C. and Wagner, G. (2006). *J Am Chem Soc* **128**, 5757-5763.

Gavaghan, C. L., Wilson, I. D. and Nicholson, J. K. (2002). *FEBS Lett* **530**, 191-196.

Gingras, A. C., Raught, B., Gygi, S. P., Niedzwiecka, A., Miron, M., Burley, S. K., Polakiewicz, R. D., Wyslouch-Cieszynska, A., Aebersold, R. and Sonenberg, N. (2001). *Genes Dev* **15**, 2852-2864.

Gutmanas, A., Jarvoll, P., Orekhov, V. Y. and Billeter, M. (2002). *J Biomol NMR* **24**, 191-201.

Heffron, G., Solanky, K., Tarragona, N., Luithardt, H., Colson, K., Maas , W., Hyberts, S. G., Fejzo, J., Edmonds, K., Falchuk, K., Chorev, M., Aktas, H., Halperin, J. A. and Wagner, G. (2006). in preparation.

Hoch, J. C. (1989). *Methods Enzymol* **176**, 216-241.

Hoch, J. C. and Stern, A. S. (2001). *Methods Enzymol* **338**, 159-178.

Hoch, J. C., Stern, A. S., Donoho, D. L. and Johnstone, I. M. (1990). *J. Magn. Reson.* **86**, 236-246.

Hounsfield, G. N. (1973). *Brit. J. Radiol.* **46**, 1016.

Keating, K. A., McConnell, O., Zhang, Y., Shen, L., Demaio, W., Mallis, L., Elmarakby, S. and Chandrasekaran, A. (2006). *Drug Metab Dispos* **34**, 1283-1287.

Kim, S. and Szyperski, T. (2003). *J Am Chem Soc* **125**, 1385-1393.

Korzhneva, D. M., Ibraghimov, I. V., Billeter, M. and Orekhov, V. Y. (2001). *J Biomol NMR* **21**, 263-268.
Kupce, E. and Freeman, R. (2004a). *J Am Chem Soc* **126**, 6429-6440.
Kupce, E. and Freeman, R. (2004b). *J Biomol NMR* **28**, 391-395.
Li, K. B., Stern, A. S. and Hoch, J. C. (1998). *J Magn Reson* **134**, 161-163.
Liang, Y. S., Kim, H. K., Lefeber, A. W., Erkelens, C., Choi, Y. H. and Verpoorte, R. (2006). *J Chromatogr A* **1112**, 148-155.
Lindon, J. C., Holmes, E. and Nicholson, J. K. (2004). *Curr Opin Mol Ther* **6**, 265-272.
Liu, G., Aramini, J., Atreya, H. S., Eletsky, A., Xiao, R., Acton, T., Ma, L., Montelione, G. T. and Szyperski, T. (2005). *J Biomol NMR* **32**, 261.
Marion, D. (2005). *J Biomol NMR* **32**, 141-150.
Nicholson, J. K., Lindon, J. C. and Holmes, E. (1999). *Xenobiotica* **29**, 1181-1189.
Orekhov, V. Y., Ibraghimov, I. and Billeter, M. (2003). *J Biomol NMR* **27**, 165-173.
Orekhov, V. Y., Ibraghimov, I. V. and Billeter, M. (2001). *J Biomol NMR* **20**, 49-60.
Rhee, I. K., Appels, N., Hofte, B., Karabatak, B., Erkelens, C., Stark, L. M., Flippin, L. A. and Verpoorte, R. (2004). *Biol Pharm Bull* **27**, 1804-1809.
Rovnyak, D., Frueh, D. P., Sastry, M., Sun, Z. Y., Stern, A. S., Hoch, J. C. and Wagner, G. (2004). *J Magn Reson* **170**, 15-21.
Rovnyak, D., Hoch, J. C., Stern, A. S. and Wagner, G. (2004). *J Biomol NMR* **30**, 1-10.
Schmieder, P., Stern, A. S., Wagner, G. and Hoch, J. C. (1993). *J Biomol NMR* **3**, 569-576.
Schmieder, P., Stern, A. S., Wagner, G. and Hoch, J. C. (1994). *J Biomol NMR* **4**, 483-490.
Schmieder, P., Stern, A. S., Wagner, G. and Hoch, J. C. (1997a). *J Magn Reson* **125**, 332-339.
Schmieder, P., Stern, A. S., Wagner, G. and Hoch, J. C. (1997b). *J Magn Reson* **125**, 332-339.
Sun, Z. J., Hyberts, S. G., Rovnyak, D., Park, S., Stern, A. S., Hoch, J. C. and Wagner, G. (2005). *J, Biomol. NMR* **32**, 55-60.
Sun, Z. Y., Frueh, D. P., Selenko, P., Hoch, J. C. and Wagner, G. (2005). *J Biomol NMR* **33**, 43-50.
Szyperski, T., Wider, G., J.H., B. and Wuthrich, K. (1993). *J Am Chem Soc* **115**, 9307-9308.
Tarragona, N., Heffron, G., Solanky, K., Luithardt, H., Colson, K., Maas , W., Hyberts, S. G., Fejzo, J., Edmonds, K., Falchuk, K., Chorev, M., Aktas, H., Halperin, J. A. and Wagner, G. (2006). in preparation.
Tugarinov, V., Kay, L. E., Ibraghimov, I. and Orekhov, V. Y. (2005). *J Am Chem Soc* **127**, 2767-2775.
Weljie, A. M., Newton, J., Mercier, P., Carlson, E. and Slupsky, C. M. (2006). *Anal Chem* **78**, 4430-4442.

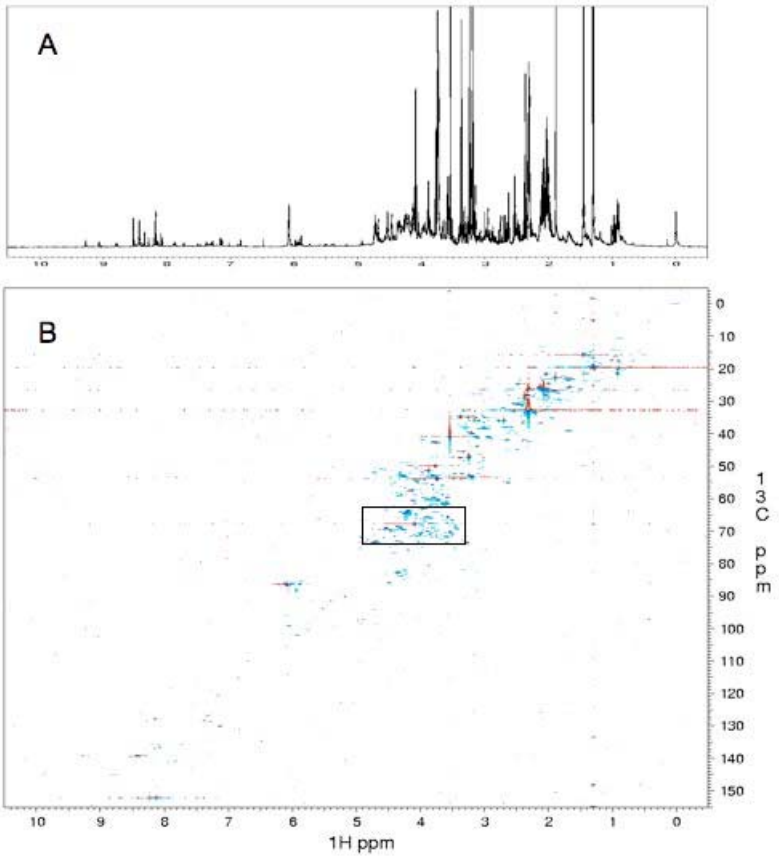

A
B
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
13C ppm
10
9
8
7
6
5
4
3
2
1
0
1H ppm

Figure 1. A, B

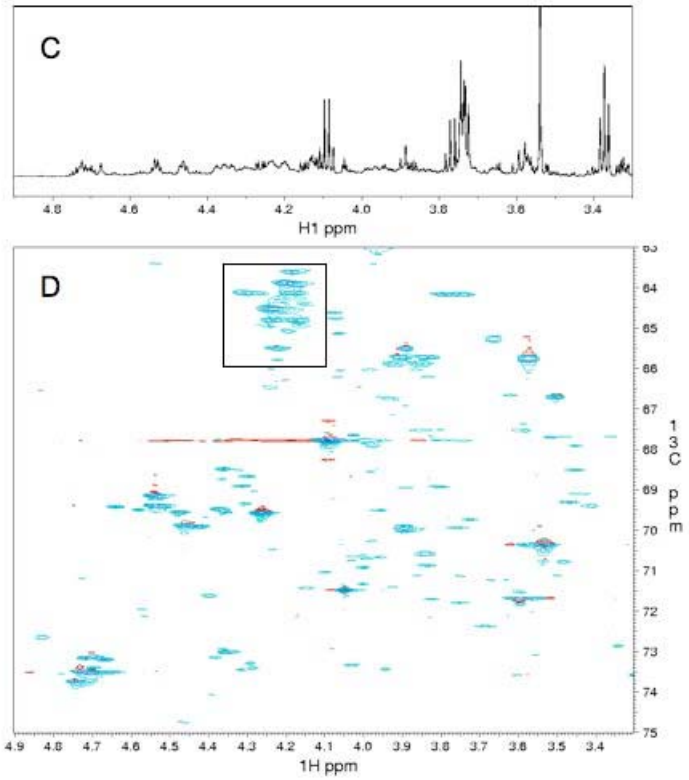
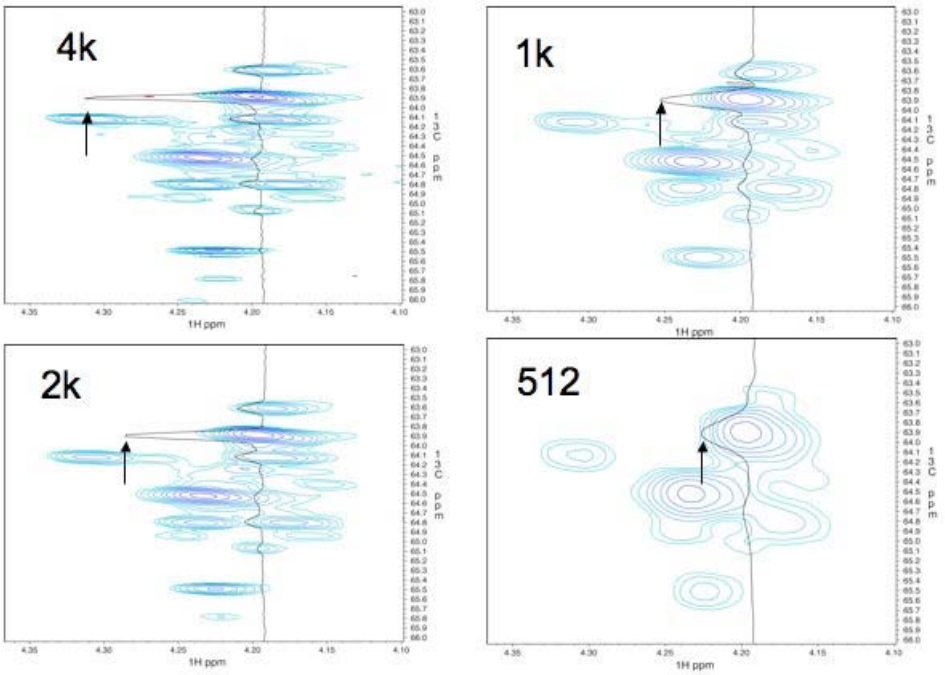
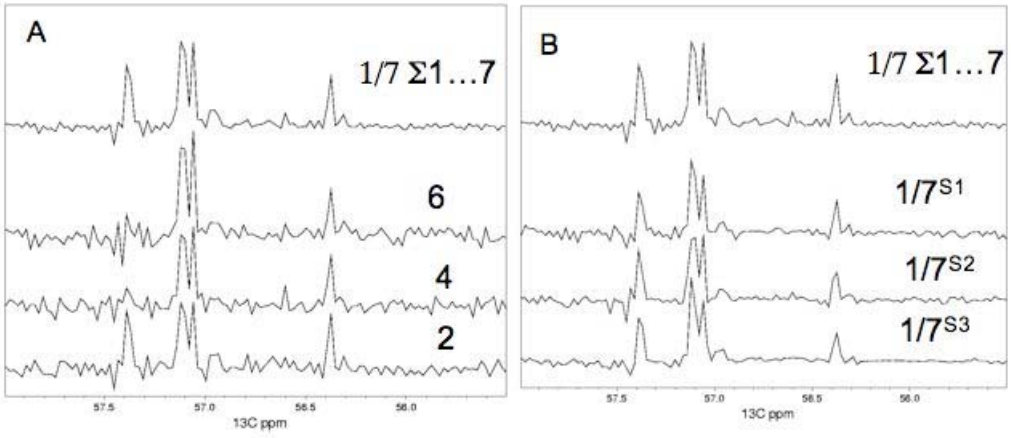

C
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
H1 ppm
D
64
65
66
67
68
69
70
71
72
73
74
75
13C ppm
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
3.9
3.8
3.7
3.6
3.5
3.4
1H ppm

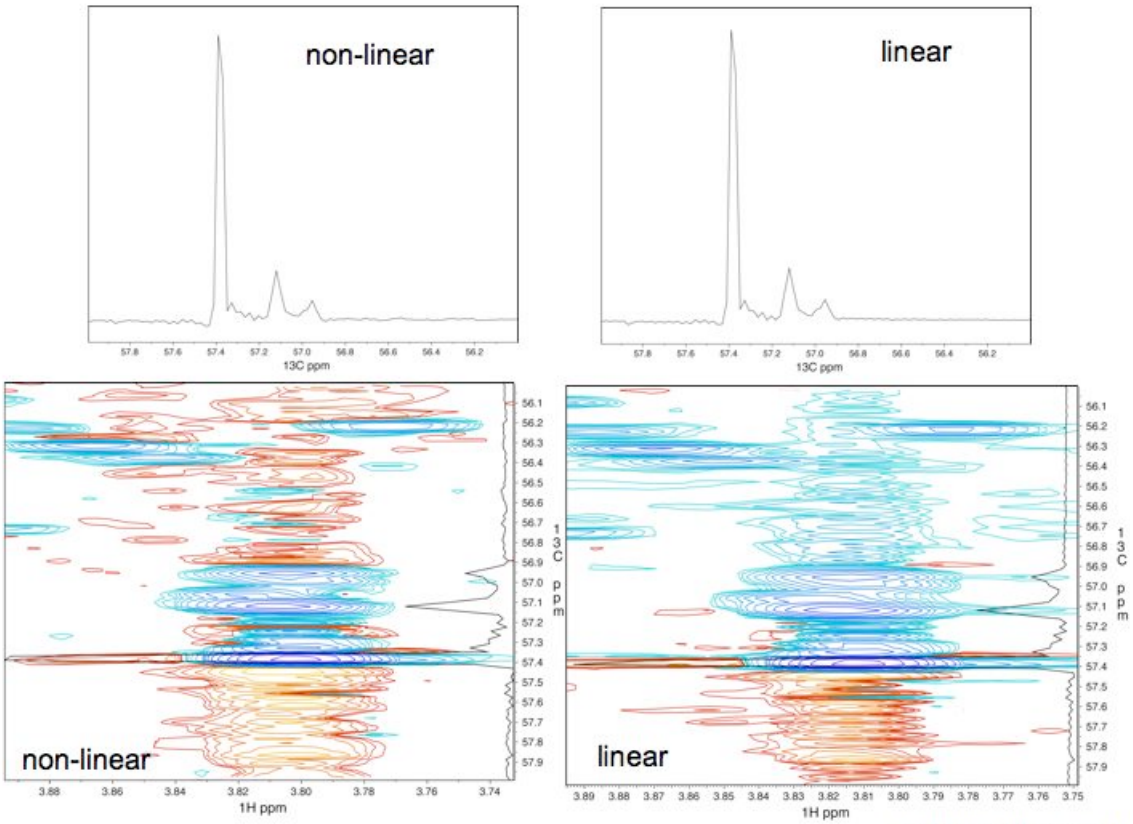
Figure 1. C,D



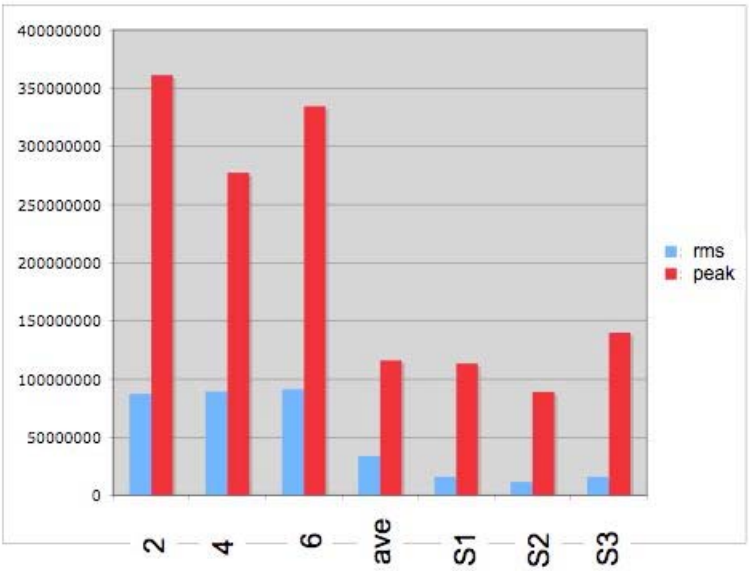
Hyberts et al. Figure 2



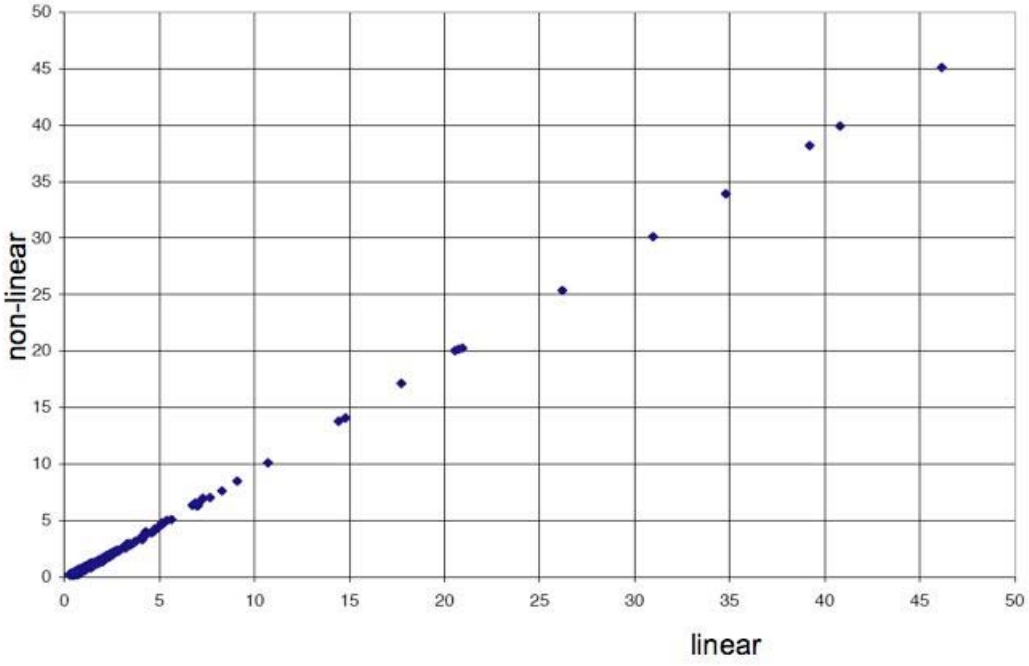
Hyberts et al. Figure 3



Hyberts et al. Figure 4



Hyberts et al., Figure 5



Hyberts et al. Figure 6