J Biomol NMR (2014) 58:83–99
DOI 10.1007/s10858-013-9807-x

ARTICLE

SARA: a software environment for the analysis of relaxation data acquired with accordion spectroscopy

Bradley J. Harden · Dominique P. Frueh

Received: 25 September 2013 / Accepted: 18 December 2013 / Published online: 10 January 2014

Abstract We present SARA (Software for Accordion Relaxation Analysis), an interactive and user-friendly MATLAB software environment designed for analyzing relaxation data obtained with accordion spectroscopy. Accordion spectroscopy can be used to measure nuclear magnetic resonance (NMR) relaxation rates in a fraction of the time required by traditional methods, yet data analysis can be intimidating and no unified software packages are available to assist investigators. Hence, the technique has not achieved widespread use within the NMR community. SARA offers users a selection of analysis protocols spanning those presented in the literature thus far, with modifications permitting a more general application to crowded spectra such as those of proteins. We discuss the advantages and limitations of each fitting method and suggest a protocol combining the strengths of each procedure to achieve optimal results. In the end, SARA provides an environment for facile extraction of relaxation rates and should promote routine application of accordion relaxation spectroscopy.

Keywords Accordion · Relaxation rate · Software · Analysis · SARA

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9807-x) contains supplementary material, which is available to authorized users.

B. J. Harden · D. P. Frueh (✉)
Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 701 Hunterian, 725 North Wolfe St., Baltimore, MD 21205, USA
e-mail: dfrueh@jhmi.edu

Introduction

Over the years, nuclear magnetic resonance (NMR) has become a workhorse for studies of protein dynamics, owing in great part to the atomic level description it provides. NMR has been used to probe modulation of protein dynamics during enzymatic reactions (Eisenmesser et al. 2005; Henzler-Wildman et al. 2007), to relate protein dynamics and thermodynamics of binding (Akke et al. 1993; Wand 2013), or to provide mechanistic insights such as revealing minor conformers (Sugase et al. 2007; Korzhnev and Kay 2008). Measurement of NMR relaxation rates is a mainstay among the methods used to probe dynamics. Unfortunately, these measurements can require prohibitive acquisition times, on the order of a day or longer. Consequently, in the absence of a priori knowledge of the relevance of internal motions, the required investment in spectrometer time dampens the incentive to perform such experiments. Indeed, dynamics may prove to be important for a protein's function, but the return from lengthy measurements may be disappointing as well, for instance when only unstructured terminal regions display flexibility. To overcome costs in acquisition time, the accordion method (Bodenhausen and Ernst 1981) can be used to rapidly measure relaxation rates with accuracy indistinguishable from those measured with traditional methods (Mandel and Palmer 1994; Rabier et al. 2001; Chen and Tjandra 2009). In effect, the accordion method accelerates data acquisition by encoding two indirect dimensions simultaneously. For example, when measuring relaxation rates, the traditional method involves recording a series of 2D HSQC experiments, each with a different relaxation period, effectively resulting in a 3D experiment. In the accordion version, the relaxation period is incremented synchronously with the period encoding signals

with heteronuclear chemical shifts, reducing the dimensionality of the experiment from three to two. Whereas ten or more 2D experiments are typically recorded using the traditional method, the accordion method requires only two. Thus, relaxation rates can be estimated with a few hours of acquisition instead of days.

Despite its age, the NMR community has under-utilized accordion relaxation spectroscopy. Bodenhausen and Ernst (1981) first developed the accordion method for exchange spectroscopy in 1981, and Mandel and Palmer (1994) subsequently applied it to relaxation rates in 1994. However, to our knowledge there have been only five publications (Mandel and Palmer 1994; Carr et al. 1998; Guenneugues et al. 1999; Rabier et al. 2001; Chen and Tjandra 2009) on accordion and relaxation analysis. This scarcity is even more surprising when one considers that the application of accordion spectroscopy to NMR relaxation is intuitively simple; the relaxation decay curve is encoded into the line-shape of NMR signals. Unfortunately, the various procedures available to extract the relaxation rates can be intimidating, time consuming and/or limited to resolved signals. The lack of discussion on the advantages and disadvantages of these procedures further impedes a popularization of the accordion technique. Thus, a unified software package enabling reliable and rapid data analysis is needed to promote the application and development of the accordion method.

We present here, SARA, a graphical, user-oriented software package designed to simplify and accelerate accordion data analysis. We have implemented existing and designed novel protocols allowing users to make use of the optimal fitting procedures for a given protein, bypassing the limitations of any one technique. In an effort to make it more accessible to occasional NMR users, SARA is written in MATLAB (MATLAB 8.2, The MathWorks Inc., Natick, MA, 2013) and features an intuitive graphical user interface. In addition, we discuss the advantages and limitations of the analytical procedures implemented in SARA within the framework of their application to studies of protein dynamics. Finally, we describe a new protocol that employs all of the fitting procedures included in SARA and ensures critical inspection of the fitted data during accordion relaxation analysis. Overall, SARA provides an environment well-suited for routine analysis of relaxation rates obtained with accordion spectroscopy, a powerful technique capable of rapidly assessing protein dynamics.

Principles of accordion relaxation spectroscopy

The accordion method can be used to reduce the dimensionality of NMR relaxation experiments; while traditional methods require the collection of an entire series of 2D experiments, the accordion method requires only two measurements. The principles of the method are described in this section.

Relaxation rates are measured by monitoring the decay of NMR signals as a function of the length of a relaxation period t_r, under the influence of either longitudinal or transverse relaxation. During a traditional relaxation experiment, the relaxation delay t_r is incremented independently of the heteronuclear chemical shift evolution period t_1, resulting in a series of 2D spectra. The time evolution, I, of a signal over these two dimensions can be written as

$$I(t_1, t_r) = I_0 e^{(-R^* + i\omega)t_1} e^{-R_i t_r} \quad (1)$$

where I_0 is the signal intensity and ω is the heteronuclear chemical shift frequency. R_i is the relaxation rate of interest (either R_1 or R_2) and R^* represents any relaxation occurring during t_1, including contributions from field inhomogeneities. Analyzing relaxation data in this form is straightforward; encoding with t_1 provides signal dispersion along a second dimension, and the relaxation rates are extracted by fitting an exponential to the signal intensities along a third, distinct dimension spanned by t_r. In contrast, the analysis of accordion data is more complicated because the t_1 and t_r dimensions are no longer distinct.

In an accordion experiment, the relaxation and chemical shift evolution periods are incremented synchronously yet with different time increments Δt_r and Δt_1, respectively. As a result, the t_r and t_1 dimensions are combined to form a single dimension. The proportionality constant between time increments

$$\kappa = \frac{\Delta t_r}{\Delta t_1} \quad (2)$$

is termed the accordion factor (Guenneugues et al. 1999). If we express t_r as a function of t_1

$$t_r = \kappa t_1 \quad (3)$$

and rewrite the apparent time evolution of the signal, now as a function of t_1 *only*,

$$I(t_1) = I_0 e^{(-R^* + i\omega)t_1} e^{-\kappa R_i t_1} = I_0 e^{-(R^* + \kappa R_i)t_1} e^{i\omega t_1} \quad (4)$$

we recover an easily interpretable expression. The frequency dependence of this signal is identical to that found in the traditional experiment, leaving the position of peaks in the spectrum unchanged. The signal decay envelope, however, now reports on relaxation during both the t_r and t_1 periods, as does the line-shape of the signal in the corresponding frequency domain. We define R_{obs}, the *observed* decay constant in the indirect dimension, as

$$R_{obs} = R^* + \kappa R_i \quad (5)$$

Thus, two measurements of R_{obs} with different values of κ enable discrimination between the relaxation rate during

t_1 (R^*) and the relaxation rate during t_r (R_i). Although some protocols set constraints on the values of κ in order to cleverly simplify the analysis (Mandel and Palmer 1994), SARA has been designed to analyze pairs of spectra with any two different values of κ. Additionally, SARA does not impose any experimental constraints other than requiring that the initial amplitude in t_1 be identical for both experiments. In other words, it only requires that the user not change parameters such as the gain or the number of scans between the two experiments.

Overall, the principle of the accordion method is rather simple, yet the experimental time-savings are substantial: Relaxation rates are encoded into the decay of the indirect dimension [or equivalently into the corresponding signal line-shape after Fourier transformation (FT)] and only two experiments are needed to delineate R_i from R^*. However, the process of extracting relaxation rates using non-linear fitting procedures, particularly in the context of crowded protein spectra, is a rather complex task best performed with the assistance of interactive software.

Analysis protocols and their implementations in SARA

Many protocols have been suggested for extracting rate constants from accordion spectra. In their original accordion publications (Bodenhausen and Ernst 1981; Bodenhausen 1982) studying exchange processes, Bodenhausen and Ernst proposed obtaining exchange rates from direct fits of Lorentzians to line-shapes in the frequency domain. They also suggested extracting individual peaks from spectra and monitoring cross-peak buildup for second order exchange processes following inverse Fourier transformation (IFT). Mandel and Palmer (1994), (Carr et al. 1998) proposed fitting decaying oscillators to accordion interferograms (i.e. Fourier transformed in the directly detected ^{1}H dimension but not in the indirect dimension). Rabier et al. (2001) also fit using decaying oscillators, but they computed the residuals following Fourier transformation. Guenneugues et al. (1999) fit the line-shapes of signals in the frequency domain. Finally Chen and Tjandra (2009) measured the line-widths of signals to extract relaxation rates. We have found that, in their original implementation, specific constraints limit each protocol (see below) and they may not be applicable individually and in all circumstances. SARA gives users access to a selection of protocols spanning the breadth of those proposed by other groups in order to allow researchers to choose the optimal fitting procedures for a given protein. For each protocol, modifications have been implemented to enable a more general application. SARA also uses a graphical and user-friendly MATLAB interface. In this section, we discuss the details of each fitting procedure.

Preliminaries and general features

SARA offers the flexibility to analyze data originating from several processing software packages, but some aspects of processing must conform to a few specific guidelines. Spectra may be loaded in the nmrPipe (Delaglio et al. 1995) format or in a format adapted from NMRLAB (Gunther et al. 2000). To allow for comparison among various procedures, SARA requires two input files: a t_1 interferogram and a fully processed two-dimensional spectrum. There are no restrictions on the processing of the detected dimension (F_2), as this dimension is not part of the relaxation analysis. In contrast, no processing may be applied to the indirect time domain (t_1) of the t_1-interferogram. In the two-dimensional spectrum, only apodization, zero-filling, Fourier transformation, and phasing are permissible when transforming to the indirect frequency dimension (F_1). For convenience, we have provided nmrPipe and NMRLAB processing scripts specifically designed to output both files in a format compatible with SARA.

SARA also requires a few auxiliary files and parameters for analysis: the value of κ for each spectrum, an X-easy peak list, a file containing the amino acid sequence in single-letter code, and the values of the maximum chemical shift in each dimension (F_1 and F_2) for referencing. The latter values may be verified within SARA using a calibration dialog. The calibration dialog displays the spectrum and overlays the positions of peaks found in the X-easy peak list. The user may interactively adjust the maximum chemical shift in each dimension to ensure proper alignment. In order to perform Monte Carlo simulations the user must also specify a region of the interferogram from which SARA may calculate the standard deviation of the noise. This is also accomplished with an interactive dialog.

After loading their data, researchers may choose from three procedures, which are described in detail below. The first two procedures are variations of the protocol proposed by Mandel and Palmer (MP) (1994). The first is a modified, fully-automated version of the MP method. The second is a user-assisted version of the automated MP protocol and allows the researcher to optimize the fitting parameters returned by the stand-alone MP algorithm. The last procedure, named FT/IFT, is a user-assisted, semi-automated protocol which extracts signals from the spectrum in the frequency domain and analyzes them after IFT. It addresses the inability of the MP method to isolate the fitting of one peak from that of other peaks in the same t_1 slice. We describe the implementation of these protocols in the reminder of this section.

MP method

Mandel and Palmer proposed two different data analysis schemes, both of which analyze the data in the time-domain of the indirect dimension, and each requires two experiments with accordion factors $\kappa = +\kappa_0$ (forward experiment) and $\kappa = -\kappa_0$ (reverse experiment). In the first scheme, called the "forward-reverse" method, the observed relaxation rate, R_{obs}, is fit separately in each experiment. The relaxation rate of interest, R_i, is calculated using

$$R_i = \frac{R_f - R_r}{2\kappa_0} \quad (6)$$

where R_f is the observed relaxation rate in the forward experiment and R_r is the observed relaxation rate in the reverse experiment. In the second analysis scheme, termed the "negative-time" method, the reverse experiment is inverted in time and concatenated with the forward experiment, forming a single data set. In this case, the entire combined data set is fit at once and provides R_i and R^* directly.

Although the two analysis schemes differ in their approach to obtaining R_i, fitting of the interferogram is accomplished similarly in both. The detected ^{1}H dimension is Fourier transformed to form a t_1 interferogram, and a t_1 free induction decay (FID) is analyzed for each signal. The positions of the t_1 FIDs in the ^{1}H dimension are determined by inspecting the signals in the 2D Fourier transformed spectrum. Next, Hankel singular value decomposition (HSVD) (Barkhuijsen et al. 1987) of each t_1 FID is used to populate a list of time-domain signals putatively present in the slice. HSVD parameterizes these signals by estimating their amplitudes, phases, frequencies and decay rates. We term a collection of these four parameters an oscillator. Putative oscillators are screened by comparison to the frequencies and phases of signals that are plausibly present in the FID (given their ^{1}H frequency), and erroneous oscillators are removed. The parameters of the surviving oscillators are used as the starting point for a non-linear least-squares optimization. Both the "forward-reverse" and "negative-time" analysis methods model the FID as a sum of damped oscillators, but the target function differs slightly in each.

Although all accordion publications following those by the Palmer group calculate R_i from two separate fits of R_{obs} (i.e. using the same strategy found in the MP forward-reverse scheme), there is an advantage inherent to the MP negative-time technique that has not yet been utilized by other authors. Because the two accordion experiments are combined prior to fitting, R_i is fit directly by fitting both experiments simultaneously. This increases the fit precision and bypasses subsequent calculation of R_i and associated error propagation. However, these advantages are a result of global fitting, of which the negative-time technique is a special case, and we note that there is no inherent need to record experiments with the specific values $\kappa = +\kappa_0$ and $\kappa = -\kappa_0$. A similar advantage is obtained with two arbitrary values of κ if the experiments are fit simultaneously using a suitably constructed global non-linear optimization (Herman and Lee 2012). SARA harnesses this strategy in its implementation of the MP method.

Automated MP implementation

Our implementation of the MP protocol incorporates two approaches. The first is a fully automated algorithm following the protocol outlined in Fig. 1 and described in this section. The user only needs to specify a handful of tolerances prior to initiation. The second is a user-assisted version of the same algorithm, in which the user can modify parameters in each processing block of Fig. 1 before passing the results to the next block. The latter method is discussed in the next section.

The processing blocks themselves proceed according to the MP method as described above and can be regrouped into three steps as follows: *Step 1* (Fig. 1 *blocks 1–4*), selection of a t_1 FID; estimation of the number of signals; HSVD; and assessment of the HSVD results; *Step 2* (Fig. 1 *R_{obs} circle*), non-linear least squares fit of the individual FIDs; and *Step 3* (Fig. 1 *global fit block*), global non-linear optimization of both FIDs. A Monte Carlo error analysis can be performed once global fitting has been completed. If desired, this Monte Carlo analysis can be applied to a user-defined subset of residues. The two accordion data sets (with different values of κ) are treated independently until global non-linear fitting, Step 3, where they are analyzed simultaneously to extract R_i directly.

Step 1: Initial estimation of decaying oscillators with HSVD

The first step selects the appropriate t_1 FID and estimates the number of signals within it. The outcome is a series of amplitudes, phases, frequencies and decay rates that can be used as initial estimates for the non-linear fitting of the FID in Step 2. The FIDs in t_1 are selected with help from the X-easy peak list provided by the user. The user is responsible for verifying that the coordinates specified in the peak list correspond to signal maxima. The peak list is also used to estimate the number of signals present in the FID, a parameter required by HSVD. This number is obtained by counting all signals with proton chemical shifts falling within a user-defined range of the signal under analysis. Next, HSVD is performed and returns a list of amplitudes, phases, frequencies and decay rates defining the decaying oscillators present in each t_1-FID. HSVD provides reliable estimates of the parameters yet tends to produce spurious oscillators, in

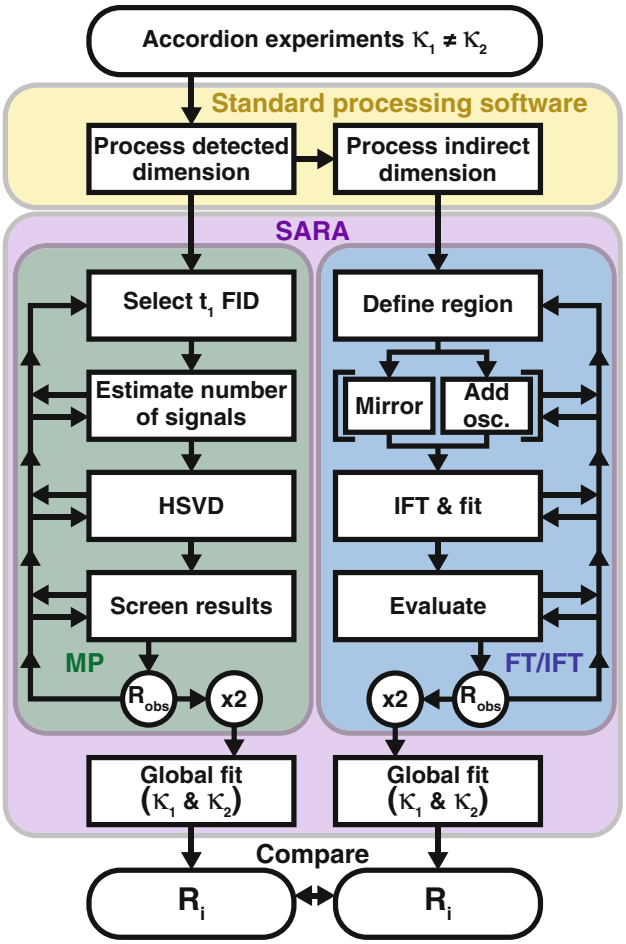

Fig. 1 Relaxation analysis flow chart. Two accordion experiments are processed with conventional NMR processing software (e.g. nmrPipe) and are loaded into SARA. The interferogram is analyzed by the MP method, while the spectrum is analyzed by the FT/IFT method. In the automated MP method, SARA completes the entire MP analysis. When using interactive methods, users follow the steps of the analysis as outlined in the text. *Arrows* indicate the flow of data analysis from one task to another. Users may return to any prior step in the analysis at any point in time. Values for R_{obs} must be obtained for each of the two accordion experiments prior to global fitting. Global fitting is accessed via dialogs separate from those of the analysis methods themselves. Results may be plotted in SARA and accessed directly via the SARA save file

particular for weak signals. To resolve this issue, we implement a strategy similar to that employed by Mandel and Palmer. First, we systematically increase the estimate of the number of oscillators provided to HSVD, ensuring that, while erroneous oscillators may be present, all oscillators corresponding to real signals are taken into account. Next, the frequencies returned by HSVD are filtered by comparison to those found in the X-easy peak list, again within the user-defined range of the signal of interest. Thus, in the end, our HSVD procedure produces a total of N oscillators, each with an estimated amplitude (A_n), phase (ϕ_n), frequency (ω_n), and observed decay rate (R_{obs_n}) for all N signals present in the FID.

Step 2: Non-linear fit of the individual FID

The next step uses the parameters returned by HSVD as the starting point for a non-linear least squares fit of the FID. The FID is fit to a sum of damped oscillators defined by the target function F(t)

$$F(t) = \sum_{n=1}^{N} A_n e^{i\phi_n} e^{(-R_{obs_n} + i\omega_n)t} \tag{7}$$

The optimized parameter χ^2 is defined as

$$\chi^2 = \sum_{k=1}^{K} Real\{I(k\Delta t_1) - F(k\Delta t_1)\}^2 + Imag\{I(k\Delta t_1) - F(k\Delta t_1)\}^2 \tag{8}$$

where I represents the experimental data and k iterates over the points in the discrete time dimension t_1 (i.e. K is the number of points acquired in t1). This step further refines the parameter estimates from HSVD prior to global fitting. Steps 1 and 2 must be performed on both accordion experiments before beginning Step 3.

Step 3: Perform global non-linear least squares optimization

While HSVD and the preliminary non-linear fits are performed on the FIDs from each accordion experiment separately, the global non-linear least squares optimization is performed on both FIDs simultaneously. The N signals identified in Step 1 and fit in Step 2 are each associated with four parameters: amplitude (A_n), phase (ϕ_n), frequency (ω_n), and observed decay rate (R_{obs_n}). Global fitting, however, parameterizes relaxation using the t_r relaxation rate (R_{i_n}) and the t_1 relaxation rate (R_n^*). For each oscillator, n, the initial values of R_i and R^* are estimated from the values of R_{obs} using the equations

$$R_i = \frac{R_{obs_2} - R_{obs_1}}{\kappa_2 - \kappa_1} \tag{9}$$

and

$$R^* = R_{obs_1} - \kappa_1 R_i \tag{10}$$

where the indices 1 and 2 correspond to the two different accordion experiments. The phase of each signal is calculated during each iteration as

$$\phi_n = \omega_n \delta \Delta t_1 \tag{11}$$

where δ parameterizes the first evolution time sampled in the experiment, $\delta\Delta t_1$. For most experiments $\delta = 0$ or ½. Initial estimates of the amplitude and frequency for each peak are calculated by averaging the respective values returned by the individual fits of the two accordion FIDs in Step 2. For this reason it is important that the two

experiments are recorded in a manner preserving this amplitude (same gain, number of transients, recycling delay, etc.). This condition is easily verified by comparing the first point in the interferogram, i.e. the 1D-proton spectrum corresponding to $t_1 = 0$. Any difference in amplitude between the two experiments will prevent the optimization from converging to the correct values. The fitted function $F_p(t)$ ($p = 1, 2$) for each of the two FIDs is a sum of N oscillators:

$$F_p(t) = \sum_{n=1}^{N} A_n e^{i\phi_n} e^{(-R_n^* + i\omega_n)t} e^{-\kappa_p R_{in} t} \quad (12)$$

The residual to be minimized, χ^2, includes the sum of the squared deviations from both FIDs:

$$\chi^2 = \sum_{p=1}^{2} \sum_{k=1}^{K} Real\{I_p(k\Delta t_1) - F_p(k\Delta t_1)\}^2 + Imag\{I_p(k\Delta t_1) - F_p(k\Delta t_1)\}^2 \quad (13)$$

where I and k are defined as in Step 2 and p iterates over the two accordion experiments. Thus, the values of R_i and R^* in $F_p(t)$ are obtained by a simultaneous fit of both experiments. As discussed, including both FIDs in a single fit rather than calculating R_i from individual fits of R_{obs} increases the precision in R_i.

Step 4: Monte Carlo error analysis

To provide a reliable error estimate, SARA includes a Monte-Carlo error analysis dialog for each method. Prior to its use, however, the level of noise in each interferogram must be established using the noise dialog accessed from the main SARA menu. Users are asked to define a region in one of the interferograms that contains only noise. Histograms of the real and imaginary values of points within the region are displayed to assist users in identifying outlier regions or artifacts. SARA then calculates the standard deviation of the points in this region for both interferograms with different values of κ to determine the t_1 noise level in each (σ_{n1} and σ_{n2}).

Within the Monte Carlo dialog, the user may specify the residues for which the Monte Carlo will be performed as well as the number of Monte Carlo iterations, N_{MC}. During each round, real and imaginary traces of pseudo-random, gaussian noise with standard deviation σ_{n1} and σ_{n2} respectively are added to the experimental data of each interferogram. Global fitting, as in Step 3, is repeated N_{MC} times. The output is an estimate of the standard deviations of both R^* and R_i. Once a Monte Carlo error analysis has been performed, error bars are automatically added when plotting results using SARA.

Interactive MP implementation

The interactive MP method follows the same organization as the automated MP method depicted in Fig. 1 but allows for user intervention at each of the individual blocks. Figure 2 shows the window used to optimize the fitting parameters and monitor the fitting quality. As in the automated method, a t_1 FID (displayed at the top left of Fig. 2) is selected from the interferogram based on the position of a residue in the X-easy peak list. This position is visualized by a blue crosshair cursor in the 2D contour plot (top right, Fig. 2). The bottom left plot contains the F_1 slice from the 2D spectrum corresponding to the vertical component of the crosshair and is the Fourier transform of the t_1 FID displayed above it. Displayed at the bottom right is a slice from the F_2 dimension corresponding to the horizontal component of the crosshair in the contour plot. The residue of interest is termed the "active" residue, and the t_1 FID should be chosen to coincide with the maximum intensity of the active residue along the detected frequency dimension (F_2), ensuring the maximum signal-to-noise ratio (SNR) possible. The F_1 and F_2 slices help the user visualize the t_1 FID selection process and are updated interactively upon changes by the user. The user positions the crosshair using the movement controls to the right of the contour plot ("Move" in Fig. 2). The additional green vertical lines displayed on the 2D contour plot represent the maximum distance in F_2 at which signals are to be considered during HSVD (as in Step 1 of the automated MP procedure). These signals are defined by the "Nearby" button in Fig. 2, and the estimated number of oscillators present in the FID is updated in an editable box ("# Osc" in Fig. 2). Vertical lines are added to the F_1 plot (bottom left) at the F_1 position of each peak identified in the search (not shown in Fig. 2). The vertical line corresponding to the active residue in this plot is marked in green (shown in the F_1 plot of Fig. 2). These lines allow the user to verify the validity of the estimated number of oscillators. Peaks may appear "nearby" according to the definition implied by the contour plot's vertical lines but may not appear substantially in the spectrum corresponding to the t_1 FID. With this knowledge the user may adjust the number of oscillators prior to HSVD. Users are reminded that it is usually preferable to over-estimate the number of oscillators, as erroneous oscillators may be removed after HSVD.

HSVD of the active residue FID is achieved by pressing the "HSVD" button. As in the automated MP method, the oscillators returned by HSVD are screened by comparison to the expected frequencies and phases of the signals identified in the search for nearby residues. The screening tolerances may be adjusted in the "Tolerance" dialog ("Tol" button in Fig. 2). The frequencies of oscillators that pass the screening process are again visualized with

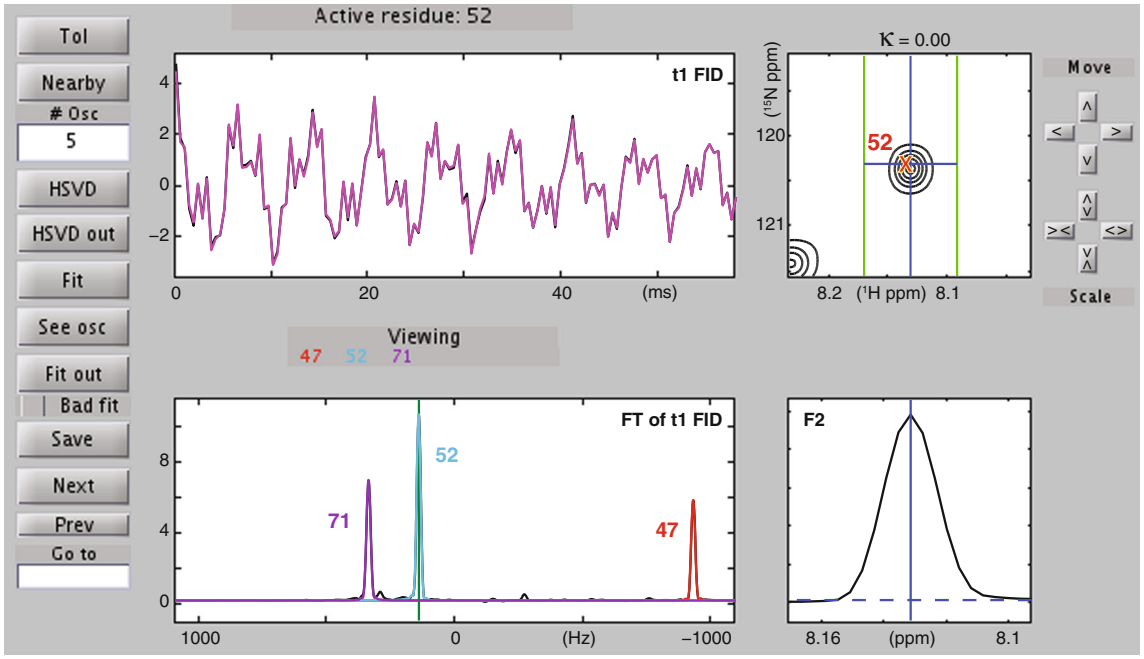

Fig. 2 The interactive MP dialog. *White boxes* represent user-editable fields. Labels have been edited for clarity. Plots, from *left* to *right* and *top* to *bottom*, contain: the t_1 FID, a contour plot of the 2D spectrum, the F_1 slice from the 2D spectrum corresponding to the t_1 FID and a slice from the F_2 dimension corresponding to the horizontal line of the *blue* crosshair in the contour plot. The value of the accordion factor (κ) for the current experiment is displayed above the contour plot (here, "$\kappa = 0.00$"). Users select the appropriate t_1 FID, with reference to the 2D spectrum, using the "Move" and "Scale" buttons. Nearby residues are identified by pressing the "Nearby" button. The number of oscillators requested from HSVD can be edited ("# Osc"), and the parameters for screening the results can be adjusted ("Tol"). The oscillators returned by HSVD may be edited ("HSVD out") and oscillators may be added or removed. Non-linear fitting is performed with the "Fit" button. The t_1 FID is fit and the resulting fitted curve is displayed in magenta on top of the *black* experimental data for comparison. The fitted curve in the time domain is Fourier transformed and overlaid on top of the spectrum slice below (not shown), and a *green vertical line* marks the F_1 position of the active residue. Rather than display the sum of all oscillators in the frequency domain, users may opt to display oscillators individually using the "See osc" feature (*red*, *cyan* and *purple* peaks). If oscillators are viewed individually, their matching residue number is displayed under "Viewing". Fitted parameters are viewed with the "Fit out" button, and poor fits may be marked with the "Bad fit" checkbox. Results must be saved ("Save") or discarded before the user may navigate through the residues of the X-easy peak list ("Next," "Prev," and "Go to")

vertical lines in the F_1 slice (not shown). Users should verify that there is a one-to-one correspondence between the peaks present in the F_1 plot and the oscillators marked by vertical lines. In the absence of any external information, HSVD provides estimates that are used as the initial values for the subsequent non-linear fit. If better estimates of these parameters have been obtained by other, independent means, the estimates provided by HSVD can be replaced prior to fitting. Additionally, users may specify that a parameter or group of parameters be held fixed during fitting, thus reducing the total number of parameters fitted. Furthermore, entire oscillators may also be added or removed from the fit. All of these changes may be made using the "HSVD out" button in Fig. 2. Such flexibility when handling fitted parameters enables users to correct inaccuracies found in the MP method, while still benefiting from its high precision (see "Discussion" section).

Until this point, the steps taken by the user have all fallen within Step 1 of the automated MP method and correspond to the blocks in the MP method of Fig. 1. The user may return to any previous step at any point in time if the parameters are not satisfactory. After editing the results of HSVD, the user presses the "Fit" button (Fig. 2) to initiate a non-linear fit as described in Step 2 of the automated MP method. The fitted data is overlaid with the experimental data in the t_1 FID plot and is Fourier transformed and overlaid with the F_1 slice below. Users may evaluate the fit using two different methods. The fitted parameters can be examined using the "Fit out" button. Alternatively, the user may display a subset of the fitted oscillators in the F_1 spectrum, rather than the entire sum, using the "See individual oscillators" feature ("See osc" in Fig. 2). Together, these methods allow the user to verify that each signal is fit appropriately and that the optimization has not converged to a local minimum. Once the fit is satisfactory, the user must save the results with the "Save" button before either proceeding to another residue in the same experiment ("Next", "Prev" or "Go To") or switching to the second accordion experiment. Global fitting (Step 3 of the automated method) may only be

performed once the same residue has been successfully fit in both accordion experiments. Both global fitting and Monte Carlo error analysis are performed as described in the automated MP method and are accessed as separate features in the main SARA window.

FT/IFT method

Although we have found the MP method to be reliable in general, severe errors may occur for special cases in crowded spectra. The combination of partial overlap and a high number of oscillators may result in a fit converging to a wrong solution (see Fig. 5 and "Discussion"). To help resolve such errors without resorting to comparison with the literature or traditional experiments (e.g. CPMG), which would defeat the purpose of performing the accordion measurement in the first place, we have designed an alternative analysis protocol which we call the FT/IFT method.

The FT/IFT protocol seeks to simplify analysis by fitting signals on an individual basis wherever possible. It is greatly inspired by the original procedure of Bodenhausen et al., in which the accordion data is Fourier transformed in both dimensions, an F_1 slice is chosen from the 2D spectrum, a peak within the slice is isolated by zeroing all points around it, the slice is inverse Fourier transformed, and finally the magnitude of the resulting time-domain FID is analyzed. Clearly this procedure can only be applied to symmetric, non-overlapped signals. To generalize the method, we have designed a procedure to extract spectral regions featuring one or potentially a few overlapping signals along F_1 while still allowing for subsequent IFT and time-domain fitting. Two solutions are presented to overcome overlap in the extracted region. The first is a simple non-linear fit of the reduced set of overlapping signals in the extracted region. The second involves mirroring half of the line-shape of a partially overlapped signal, a process we call symmetrization, resulting in a single symmetric peak which is then inverse Fourier transformed. These two solutions are not implemented as distinct procedures but rather as alternative features that can be used to assess the reliability of the rates obtained by each method.

The Fourier-transform-based methods presented here allow investigators to fit individual peaks separately. Whereas in the interferogram-based methods proposed by Mandel and Palmer, all signals in the t_1 slice must be fit simultaneously, in the FT/IFT method, signals may be extracted from F_1 slices and fit independently. Simplifying the spectrum to reduce the number of oscillators may offer significant advantages in cases of large proteins or crowded spectra.

FT/IFT implementation

The FT/IFT method requires two 2D-Fourier-transformed accordion spectra, each with a different value of κ. The steps of the FT/IFT method are summarized in Fig. 1. Only apodization, zero-filling, Fourier transformation and phasing along the *indirect* dimension are permissible. For each peak, the user defines an extraction region, isolating the signal or group of signals if there is overlap. This region may differ between the two spectra as needed (e.g. to avoid truncation of signals by the region boundaries). Next, we construct a pseudo F_1 slice, placing the extracted region in the center and zero-filling both sides to the full spectral width. Following inverse Fourier transform, the resulting time-domain FID is truncated to the number of points acquired, thus accounting for any initial zero-filling applied during spectrum processing. We refer to this FID as a "reconstructed FID," and it represents a recreation of what would have been acquired if the peak (or group of overlapped peaks) had been isolated and nearly on-resonance during acquisition.

If the extraction region contains only a single, isolated signal then the reconstructed FID is fit as a damped oscillator in a non-linear least squares optimization, bypassing block 2 in Fig. 1. If, however, the region contains two or more signals that cannot be effectively isolated, then the reconstructed FID can be analyzed in two different manners, as illustrated in Fig. 1. The first and simplest method models the reconstructed FID as a sum of damped oscillators. The second method requires further spectral manipulation prior to inverse Fourier transform. If a signal is partially resolved from others along F_1, to the extent that half of its line-shape does not contain any contribution from overlapping peaks, the F_1-slice can be mirrored around the center of the targeted peak in a process we call symmetrization. This procedure results in a single, symmetric signal that can be inverse Fourier transformed and fit using one damped oscillator. In both cases, the fitting procedure is first optimized on fits of R_{obs}, i.e. in reconstructed FIDs derived from each accordion spectrum. The final value of R_i, however, is derived by globally fitting the two reconstructed FIDs simultaneously.

The FT/IFT procedure is semi-automated. The user is required to define the extraction region position and boundaries, establish the number of signals present within the region (usually one), and verify that the non-linear optimization is able to converge. The process is discussed in detail below.

Step 1: Defining the extraction region

In the FT/IFT method the user navigates through the X-easy peak list and chooses a so-called "active residue."

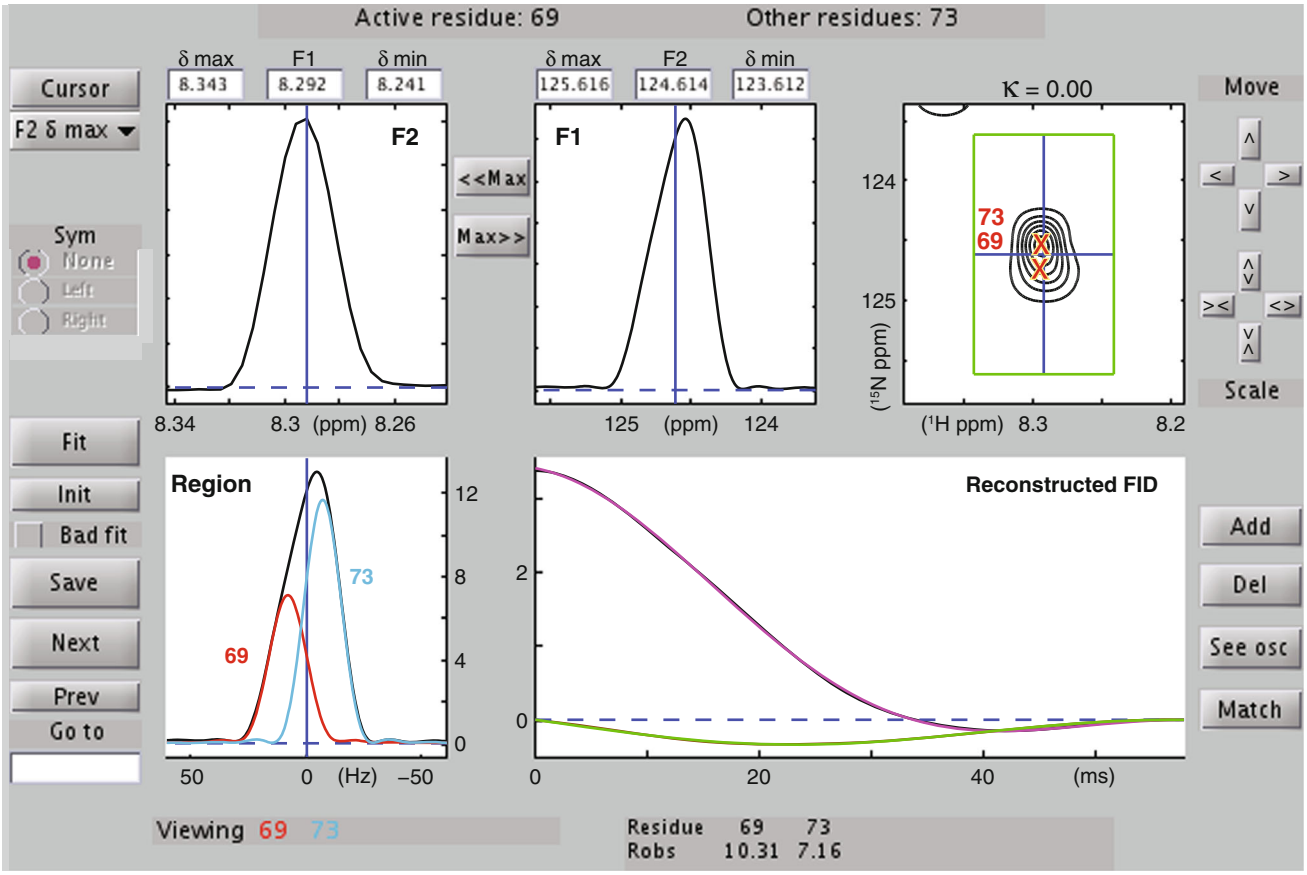

Fig. 3 The FT/IFT dialog. *White boxes* represent user-editable fields. Labels have been edited for clarity. Plots, from *left* to *right* and *top* to *bottom*, include: an F_2 slice from the spectrum, an F_1 slice from the spectrum, a contour plot of the 2D spectrum, the extracted region prior to inverse Fourier transformation and the reconstructed FID. The real (*black*) and imaginary (*brown*) parts of the reconstructed FID are both displayed. The *green box* surrounding the peak in the 2D contour plot represents the limits of the region of interest. The F_2 slice plotted at the top left corresponds to the horizontal line of the *blue* crosshair within the region-bounding box, and the F_1 slice at the top-middle corresponds to the *vertical line*. Users may "Move" and "Scale" the bounding box with the respective buttons to the *right* of the contour plot. Alternatively, users may specify the limits of the box and the positions of the slices within it using the editable fields above the F_1 and F_2 plots. Additionally, these fields can be defined interactively with a data cursor that is activated from the drop-down menu at the *left*. This menu specifies which field is being defined (here, "F2 δ max") Symmetrization may be applied using the "Sym" button group. The symmetrized peak is displayed in the "Region" plot at the *bottom left*. Alternatively, as has been done in this example, the user may add a second residue to the region using the "Add" button. Residues may be removed using the "Del" button. The active residue is listed at the top left, and any other residues added to the region are listed at the top right. Once satisfied with the reconstructed FID, the user presses the "Fit" button. The resulting fitted values of R_{obs} are displayed at the *bottom right*. The real (*magenta*) and imaginary (*green*) curves of the fit are plotted on top of the experimental reconstructed FID for comparison. The Fourier transform of the fitted curve is overlaid on the "Region" plot (not shown). Rather than plot the sum of all oscillators, individual oscillators may be Fourier transformed and overlaid on the "Region" plot (*cyan* and *red* peaks) using the "See osc" feature. The individual oscillators are displayed next to "Viewing". Oscillators are matched to their assigned residues using the "Match" dialog. The user may edit the initial fitting parameters using the "Init" button. If a satisfactory fit cannot be obtained, the residue may be marked as a "Bad fit" with the corresponding checkbox. Results must be saved ("Save") or discarded before the user may move to another residue using "Next," "Prev," or "Go to."

The X-easy peak list contains the position of each signal in the F_1 and F_2 dimensions along with the signal's corresponding residue number. This number may be arbitrary in the case of unassigned proteins, but SARA uses it as a label to differentiate signals. When establishing and testing fitting procedures, the active residue is the only residue for which the user should be concerned. While other residues may be included in the fit because of overlap, their rates will be measured from the regions in which each is the active residue. This strategy ensures that the rate measured for each residue is derived using the most optimal extraction region possible.

In SARA an extraction region is parameterized by: an active residue, a list of other residues contained within the region, boundaries of the region in F_1, boundaries in F_2, a 1D slice in F_1, and a 1D slice in F_2. The active residue is

listed at the top left of the FT/IFT dialog (Fig. 3), and any other residues that are to be considered during the fit are listed at the top right. The plots along the top row, from left to right, are: a 1D slice along F_2, a 1D slice along F_1, and a two-dimensional contour plot. The green box in the contour plot represents the boundaries of the extraction region, and the blue crosshair within it represent the positions of the 1D slices displayed to the left. The bottom two plots contain the extracted data. The frequency domain data on the left displays the data prior to IFT. If symmetrization is not used, then it is identical to the F_1 slice in the top row, otherwise it displays the symmetrized peak. The time domain data on the right is the reconstructed FID calculated from inverse Fourier transformation of the plot to its left. All of these plots are updated in real time while the user defines the extraction region.

In order to maximize the SNR, the F_1 slice must be positioned at the height of the active peak in the F_2 dimension. Additionally, the user should be sure to capture the entire line-shape within the F_1 region. Truncating the base of the peak can lead to spurious oscillations in the reconstructed FID and systematic inaccuracies in the fitted rates (see Fig. 5 and "Discussion"). The boundaries of the region along F_2 only need to be large enough for unambiguous identification of the maximal signal intensity in F_2. The region may be moved, expanded or contracted in both dimensions of the contour plot using the controls to its right ("Move" and "Scale" in Fig. 3).

In cases of partial overlap between peaks in F_1, special consideration must be taken when defining the region boundaries. The user may opt to accept some truncation of the active residue signal prior to IFT in order to separate the peaks; however, as mentioned above, this strategy can introduce systematic errors. Alternatively, the user can test if symmetrization resolves the overlap issue. SARA includes an option to reconstruct a full peak using only one half of the actual peak and can be accessed with the button group on the left side of the FT/IFT dialog ("Sym" in Fig. 3). The symmetrization option may be beneficial in cases of slight overlap, i.e. when half of the signal is not perturbed by overlapping signals. However, symmetrization can introduce inaccuracies in rate measurements when the peak maximum does not coincide exactly with a point in the spectrum. Such a problem may be minimized with extensive zero-filling prior to Fourier transformation (see Fig. 7 and "Discussion"). To accomplish this SARA selects the single t_1 FID corresponding to the signal to be symmetrized, increases the amount of zero-filling for this FID, and Fourier transforms this FID before applying symmetrization. Thus, symmetrization is not applied to signals taken from the low resolution, user-loaded spectrum but rather to signals benefiting from a much higher digital resolution. The procedure alleviates the need for extensive zero-filling of the entire dataset, thus saving disk space and accelerating calculations for other methods of analysis. The amount of zero-filling is estimated automatically to ensure that the inaccuracy introduced by symmetrization does not exceed 0.6 % for rates of 5 s^{-1} and above (see "Discussion" and Fig. 7). Clearly, this feature does not prevent inaccuracies originating from a partially overlapping peak that may contribute to the area of the signal used for symmetrization. In particular, it may be difficult to verify that half the line-shape of a signal is not perturbed by an overlapping signal when investigating two signals with a large difference in intensity. Users are advised to exercise the symmetrization option with care.

If all of these solutions fail, or if the overlap is too strong, the user may instead expand the region to contain both of the overlapping residues and fit two oscillators to the reconstructed FID instead of one. Pressing the "Add residue to region" button ("Add" in Fig. 3) will search the peak list for residues positioned within the contour plot box and prompt the user to add them to the region. When fitting in the next step, SARA will determine the number of oscillators to use based on the number of residues identified in the region. Such an implementation of the FT/IFT procedure results in a reduction in the number of oscillators that are fitted when compared to the MP method. The major advantage is that the results of the fit displayed by SARA are more simple to analyze and, hence, more likely to reveal poor fitting.

Step 2: Optimizing data fit

The data is fit in the time-domain rather than the frequency domain. To transform the F_1 slice from the frequency domain to the time domain SARA extracts the region, performs symmetrization if requested, centers it at zero frequency, zero-fills the spectrum to the full spectral width, performs an inverse Fourier transform and truncates the resulting FID to the number of points acquired K (e.g. TD1, ni, etc.).

After pressing the "Fit" button, the reconstructed FID is fit to a sum of damped oscillators. The number of oscillators N included in the fit is determined based on the number of residues that were added to the region in the previous step. Each oscillator n is defined by an amplitude (A_n), a frequency (ω_n) and a decay rate (R_{obs_n}). By fitting the frequency of signals in the reconstructed FID, even when only one signal is present, we relax any requirement that signals be perfectly centered prior to IFT. The initial amplitude of each oscillator is set to the value of the reconstructed FID's first point (i.e. at $t = 0$) divided by the number of oscillators N. The initial decay rates default to 5 s^{-1}. The initial frequencies are spread uniformly within the boundaries of the region, e.g. if there are two oscillators

in a 300 Hz region centered at zero then the initial frequencies will be set to +50 Hz and −50 Hz. As in the MP protocol, the fitted function is the sum of damped oscillators. However, it now contains a minimal number of oscillators, N, while being constrained to the same number of data points (K). In addition, the target function is now multiplied by the apodization function, G(t), used for processing the spectrum:

$$F(t) = G(t)\sum_{n=1}^{N} A_n e^{(-R_{obs_n}+i\omega_n)t} \quad (14)$$

The optimized parameter χ^2 is identical to Eq. 8 of the MP method. The resulting rate is a fit of R_{obs} in the current reconstructed FID only; the rate R_i is subsequently obtained by globally fitting both accordion spectra simultaneously (Step 3).

In cases where two or more oscillators are fit to the reconstructed FID, the sum of the oscillators is overlaid for comparison with the experimental data. Alternatively, the individual oscillators may be viewed using the "See individual oscillators" feature ("See osc" in Fig. 3). The decay rates resulting from the fit are displayed for all residues present in the region. However, the fitting procedure is such that a set of fitted parameters (A, ω, R_{obs}) may not be assigned to the correct residue. The user can assign the correct oscillator to the correct residue using the "Match residues to oscillators" feature ("Match" in Fig. 3). In cases of heavy overlap, the user should be sure to interpret the results with care, as the accuracy and precision of fitting decreases with increasing overlap.

At this stage, Steps 1 and 2 may be iterated to improve the fit. The quality of the fit is assessed by visual inspection of calculated and experimental points in the time and frequency domains (Fig. 3). If the fit is not acceptable, the user may first try to adjust the initial estimates of the parameters to be fit. Users may also opt to fix a subset of the fitting parameters using values known through other, independent measurements (e.g. peak frequency). Both of these strategies may be accessed using the initial parameter dialog ("Init" button in Fig. 3). If the fit remains unsatisfactory, the user may return to Step 1 and adjust the parameters of the extraction region. If the result still does not meet expectations, the user may mark it as a "Bad fit" before moving on.

Step 3: Global fit

Because the two accordion spectra have different values of κ, the width of signals along F_1, and consequently, the user-defined regions will be different for each spectrum. Therefore, for a given residue, the user must repeat Steps 1 and 2 on the second accordion spectrum prior to reaching step three. The third step preforms a global fit of both reconstructed FIDs for a given residue in order to extract R_i. Global fitting is initiated from the main SARA dialog. The global target function for each reconstructed FID is defined in much the same way as in the MP method, except that now it is multiplied by the apodization function G(t).

$$F_p(t) = G(t)\sum_{n=1}^{N} A_n e^{(-R_n^*+i\omega_n)t} e^{-\kappa_p R_{i_n} t} \quad (15)$$

The optimized parameter is the same as in Eq. 13, except that now the experimental data I_p is a reconstructed FID rather than a slice in the acquired interferogram. For each oscillator, the average amplitude and frequency of the two test fits performed in step 2 are used as the initial values for the global optimization. The initial values for R_i and R^* are derived from the test fits of R_{obs} and are calculated using Eqs. 9 and 10.

Step 4: Monte Carlo error analysis

SARA offers a Monte Carlo error analysis dialog for the FT/IFT as well. As with the MP methods, the user must first establish the level of noise in each experiment and may specify the number of rounds of the Monte Carlo and the residues for which it will be performed. In each round of the Monte Carlo SARA adds traces of pseudo-random noise to the reconstructed FIDs and repeats Step 3. The resulting standard deviations of R^* and R_i are stored and added as error bars when plotting results.

Discussion

When determining the ideal fitting procedures for a given protein, the user should consider the advantages and disadvantages of each fitting method. In the MP method, the absence of spectral manipulation in the t_1 dimension prevents any loss of signal or introduction of artifacts, and therefore the MP method has a higher fitting precision than the FT/IFT method (see below for an example). Nevertheless, the MP method fits many signals simultaneously and signals overlapping in t_1 may produce erroneous results that go unnoticed. The FT/IFT method provides an opportunity for users to more closely inspect their data in order to identify inconsistencies and special circumstances. Once problems occurring with the MP method are identified with FT/IFT, the user may opt to re-analyze the data with the interactive MP method, albeit with input parameters estimated with FT/IFT. Thus, inaccuracies may be identified with FT/IFT and precise and accurate rates can later be obtained with the interactive MP method. This section describes in detail the accuracy and precision of

relaxation rates obtained with the various protocols offered by SARA and describes how MP and FT/IFT can be used in concert to identify inaccuracies while maximizing precision.

Accuracy of the MP method

One example from ubiquitin captures many of the features that distinguish the MP and FT/IFT fitting methods. An F_1 slice taken from the HN spectrum of our ubiquitin sample and centered on residue 49 (the active residue) also contains four other residues. In particular residue 49 is close to the strongly overlapped residues 31 and 72. When analyzing the corresponding t_1 FID with the MP method, the estimated number of oscillators provided to the HSVD algorithm was five. However, the HSVD algorithm yielded four oscillators to describe the data; only a single oscillator was predicted to represent the two strongly overlapped peaks. Subsequent non-linear fitting using four oscillators yielded results that appeared valid. However, concerns regarding the accuracy of rates in the presence of nearby, overlapping residues prompted further investigation. Indeed, in a simulated reconstruction of the data from these residues, the rate of the active residue 49, which itself is not subject to overlap, was unexpectedly inaccurate by 3.4 % when residues 31 and 72 were fit with a single oscillator (Fig. 4a and b). Accuracy for the rate of residue 49 was recovered, however, if the number of oscillators was increased to five and the overlapped signals were fit with two oscillators rather than one (Fig. 4c and d). Hence, the combination of two oscillators more completely accounts for the presence of the overlapped signals and prevents compensation by neighboring, isolated oscillators during the fit. Even so, the rates obtained for the strongly overlapped residues themselves are extremely unreliable, as seen by a simple visual inspection of Fig. 4d. While the occurrence of such overlapping residues can be predicted in a fully assigned protein, the blind application of the MP method to unidentified overlapping residues could lead to systematic inaccuracies in the fitted parameters. Indeed, the fits of both FIDs and their Fourier transforms appear deceptively good in Fig. 4, owing to the presence of other, well-fit residues, and such erroneous results may well go unnoticed. In fact, visual inspection of individual oscillators might suggest that using four oscillators provides superior results (Fig. 4b vs. d). Clearly, when analyzing experimental data such a situation would likely go unnoticed; four oscillators would be used in the fit, and the rate of residue 49 would be inaccurate. Fortunately, the FT/IFT method provides an alternative means of analysis that reveals inconsistencies and may often provide relief, as discussed in the following section.

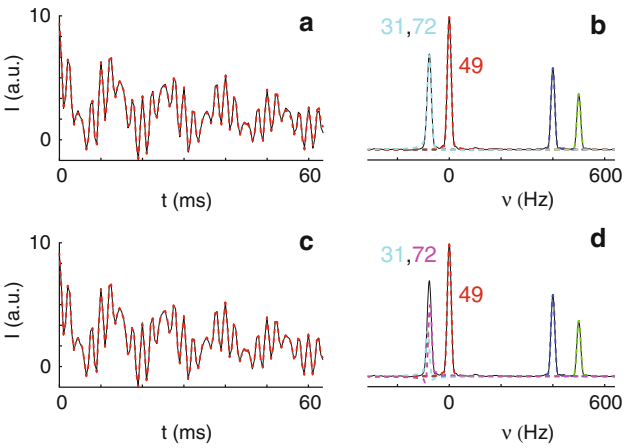

Fig. 4 Inaccuracies of the automated MP method in the presence of spectral overlap. The four panels contain simulated data (*solid*) overlaid with their corresponding fits (*dashed*). The simulated data reproduces the t_1 FID and spectrum slice of residue 49 in ubiquitin. It contains five damped oscillators corresponding to residues 4, 18, 31, 49 and 72. The frequencies of 31 and 72 are nearly degenerate. **a** Time domain data (*black*) overlaid with a fit including four damped oscillators (*red*). **b** Fourier transform of the FID in (**a**) (*black*) overlaid with the Fourier transform of each individual damped oscillator (colors). Visual inspection does not indicate inaccuracy for residue 49. **c** Same time domain data as in (**a**) (*black*) overlaid with a fit (*red*) using *five* damped oscillators. **d** Fourier transform of the data in (**c**) (*black*) overlaid with the Fourier transform of each individual oscillator of the fit (colors). Residues 31 and 72 seem to give poorer results. While visual inspection of panels (**a**)–(**d**) for residue 49 indicates similar performance between fitting with four and five oscillators, the fit in panels **a** and **b** is inaccurate by 3.4 %. This figure illustrates the importance of close inspection by the user when applying non-linear fitting methods. The FT/IFT method can be used to closely inspect and evaluate multiple fitting approaches in these types of situations. The fitting strategy and parameters can then be implemented in the interactive MP method for maximum fitting precision

Accuracy of the FT/IFT method

The FT/IFT method provides an alternative that both isolates the fitting of signals from one another and allows closer inspection by the user, facilitating the detection of poor fits. The procedure may allow users to overcome the limitations of the MP method. When analyzing the example discussed above with the FT/IFT method, residue 49 was extracted together with residues 31 and 72 and fit using only two oscillators, corresponding to the default number of four oscillators used in the MP method. Comparison between the rates obtained by MP and FT/IFT revealed a discrepancy of 0.6 s^{-1} between the two, with the rate obtained by FT/IFT inaccurate by only 1.1 % whereas that of MP suffered from 3.4 % inaccuracy. This improvement in accuracy further demonstrates that the error in the rate estimated by MP originated from the cumulative effect of numerous and overlapping oscillators in the same t_1 FID; when only two oscillators needed to be fit rather than four,

the rate of 49 became more accurate. A markedly better fit could be observed with the FT/IFT method when using three oscillators rather than two, providing a rate accurate within 0.03 %. The simplified interferogram produced by the FT/IFT method simultaneously increases the chances of identifying such an issue and provides the correct solution. Alternatively, one could isolate residue 49 using the symmetrization feature, taking advantage of the minimal overlap between 49 and 31/72. With symmetrization, the rate was accurate within 1 %, demonstrating that the rate of the symmetrized signal was closer to that obtained with three oscillators than that obtained with two oscillators using standard FT/IFT processing. Although FT/IFT, with or without symmetrization, provides a more accurate answer in this example, this trait cannot be generalized, and we seek only to highlight that FT/IFT can be used as a tool to identify discrepancies and help resolve them. Indeed, when the parameters obtained by FT/IFT were supplied to the interactive MP procedure, the estimated rate became accurate to the fourth decimal place and benefited from the increased precision of the MP method. We note that the extraction and symmetrization procedures themselves can introduce inaccuracies, and this section describes the sources of these systematic errors and strategies to mitigate them.

When determining the spectral region boundaries prior to extraction, users should be sure to capture the entire line-shape within the region. Figure 5 displays the normalized inaccuracy of a fitted rate as a function of the relative region width. A simulated spectrum was created with a signal-to-noise ratio of 20 using a relaxation rate of 30 s^{-1} and a cosine-squared window function. The FT/IFT procedure was then performed for extracted regions of various widths and repeated in a 1,000 round Monte Carlo simulation. The absolute inaccuracy δ is calculated as the difference between the mean of the measured rates (R_m) and the true rate used to simulate the signal (R_t), $\delta = R_m - R_t$. Normalized inaccuracy is calculated by dividing the absolute inaccuracy δ by the simulated rate R_t. The relative region width is defined as the extracted region width divided by the full-width at half-max (FWHM) of the apodized and Fourier transformed peak. These simulations reveal that extraction regions which truncate the line-shape lead to systematic underestimation of the relaxation rate. As expected the effect is dramatic when the extracted region approaches the FWHM, and at two FWHM the fitted rate remains inaccurate by over 15 %. At three FWHM, rates become accurate within 6 % and at five FWHM a theoretical rate of 30 s^{-1} leads to an estimated rate of 29.6 s^{-1} and, hence, an accuracy of 98.6 %. Thus, a region width of five times the FWHM ensures accurate results for rates near 30 s^{-1}. Similar simulations were performed for rates ranging from 5 to 100 s^{-1} and SNRs ranging from 5 to 300. The results show that larger rates require wider regions, while smaller rates are less prone to inaccuracies originating from truncation. Thus, at five FWHM, rates of 5, 50, and 100 s^{-1} are accurate within 1.1, 2.1, and 2.8 %, respectively. Representative examples are provided in the supplementary material for rates of 5, 10, 30, and 100 s^{-1} and for SNRs of 5 and 20 (Supplementary Fig. 1). Inspection of Fig. 5 reveals that the error obtained by the Monte Carlo analysis does not increase as the peak is truncated. Thus, rates extracted from substantially truncated signals may appear deceptively precise while being inaccurate. That is, the error-bars associated with such a rate would not account for the systematic underestimation induced by truncation, and the probability that the true rate would be contained within the error-bar-interval would be reduced. However, while inaccuracies introduced by truncation are insensitive to SNR (Supplementary Fig. 1), the error associated with a fitted rate increases with decreasing SNR and it becomes necessary to consider inaccuracy *relative* to error when assessing the validity of a fitted relaxation rate and its error-bars. Thus, we introduce a parameter δ/σ that reports on both inaccuracy (δ) and precision as measured by the standard deviation (σ) of fitted rates in a Monte Carlo analysis. This parameter provides insights on the confidence level associated with a fitted rate's error-bar-interval. To understand this parameter we consider the following. For a given distribution of fitted rates, inaccuracy moves the mean of the distribution away from the "true" rate. A value of $\delta/\sigma = 0.5$ implies that the mean of the distribution deviates from the true rate by half of the distribution's standard deviation. Thus, the probability of finding the *true* rate within the fitted rate's error bars is reduced. For example, for a Gaussian distribution, error bars of $\pm$ σ about the fitted rate describe a 68 % confidence interval. An inaccuracy $\delta = 0.5\sigma$ ($\delta/\sigma = 0.5$) reduces the confidence level of the same error-bar-interval to 62 %. Likewise, a value of $\delta/\sigma = 1$ reduces the confidence level of the error-bar-interval from 68 to 48 %. Figure 6 displays the ratio of the absolute inaccuracy (δ) to the standard deviation (σ) of fitted rates over a 1,000 round Monte Carlo simulation. This ratio is plotted as a function of SNR for rates of 10 s^{-1} (Fig. 6a) and 30 s^{-1} (Fig. 6b), where SNR is defined as the amplitude of the first point of the decaying signal divided by the standard deviation of the noise in the time domain. The horizontal dashed line highlights a value of $\delta/\sigma = 0.5$, below which we consider rates to be valid with high confidence (greater than 62 % if the rates were drawn from Gaussian distribution). That is, for a value of δ/σ below 0.5 one would expect the error bars to encompass the true rate with high probability. The colored lines correspond to region widths of 3, 4, 5 and 6 FWHM and therefore reflect different degrees of truncation during extraction. Displaying δ/σ as a function of SNR

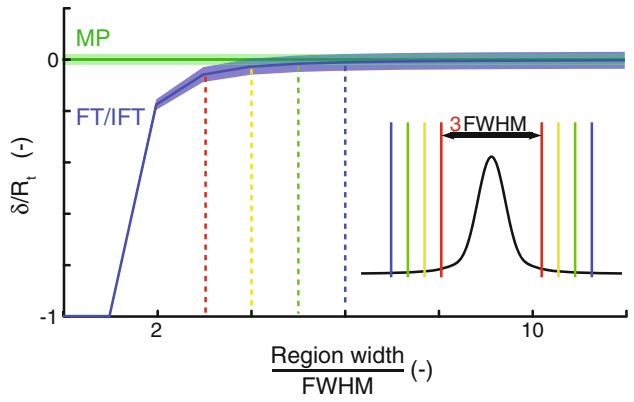

Fig. 5 Normalized inaccuracy as a function of the relative region width when fitting R_{obs} with the FT/IFT method. Normalized inaccuracy, δ/R_t, was calculated as the absolute inaccuracy $\delta = R_m - R_t$ divided by the true rate used to generate the signal, $R_t = 30\ s^{-1}$. R_m is the mean of the extracted rates, R_{obs}, over a Monte Carlo simulation. Normalized error was calculated as the SD of the Monte Carlo simulation, σ, divided by the true rate, R_t. *Bold*, *centerlines* denote the normalized inaccuracy while the shaded regions represent the extent of the normalized error (±σ). The inaccuracy and precision of the FT/IFT method (*blue*) is compared with that of the MP method (*green*), which is displayed identically at all window widths for visualization. Fits of R_{obs} were calculated over a 1,000 round Monte Carlo performed on a decaying exponential with signal-to-noise ratio 20, decay rate $30\ s^{-1}$ and defined by 128 complex points at intervals of 500 μs. During each round of the Monte Carlo, pseudo-random noise was added to the pure signal and the resulting FID was fit using the MP method. The FID was then anodized with a cosine-squared window function, zero-filled to 4,096 points, Fourier transformed, and fit using the FT/IFT method at various window widths. Relative region widths were chosen as integer multiples of the full width at half max (FWHM) as measured on the noiseless peak. The inlay displays the corresponding peak line-shape. Solid bars in red, yellow, green and blue indicate region widths of 3, 4, 5 and 6 times the FWHM respectively. Dashed lines in matching colors indicate the respective inaccuracy at each level of truncation

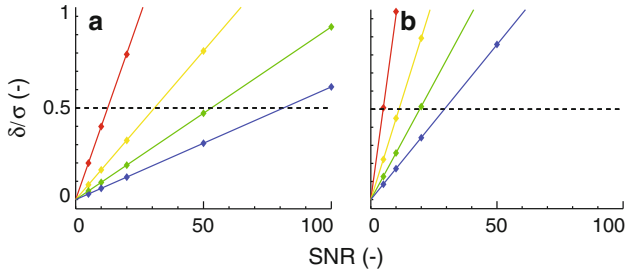

Fig. 6 The combined variation of inaccuracy and precision as a function of SNR at various region widths using the FT/IFT method. Data is taken from simulated signals with decay rates of **a** $10\ s^{-1}$ and **b** $30\ s^{-1}$. In each plot, the fixed value of inaccuracy (δ), which is independent of SNR, is divided by the standard deviation (σ) of 1,000 fits of the relaxation rate at different values of SNR. The parameters for each round of the Monte Carlo simulation are as described in Fig. 5. The *colored lines* indicate region widths of three (*red*), four (*yellow*), five (*green*) and six (*blue*) times the FWHM. The *dashed line* indicates an inaccuracy that is half the SD of the fitted rate at the given SNR. If these rates were drawn from a Gaussian distribution, this value would correspond to a reduction in confidence of the error-bar-interval (± σ) from 68 to 62 %. For all rates, increasing the extracted region width increases the confidence level by improving accuracy [lower δ, e.g. *red* (three FWHM) vs. *blue* (six FWHM) lines]. At low SNR, the inherently large σ dominates and even substantial truncation (three FWHM, *red lines*) has little effect on the level of confidence of the error-bar-interval. Smaller rates have lower slopes and therefore remain within the 0.5 σ region at higher values of SNR when compared to larger rates [(**a**) vs. (**b**)]. The validity of rates obtained with FT/IFT is most questionable for signals with large relaxation rates yet high SNR, where the δ/σ ratio would be large, reflecting that the error-bar-interval would be unlikely to contain the true value. In practice, such a situation is unlikely to occur because fast relaxation leads to reduced SNR

provides a linear dependence in which the slope describes how the confidence interval is affected as the SNR varies. Small slopes mean that the region width provides valid rates for a large range of SNR. Figure 6b shows that, for a rate of $10\ s^{-1}$, an extracted window width of three FWHM provides valid rates for SNRs below 12, four FWHM is an acceptable width for SNRs up to 30, five FWHM for SNRs up to 55 and six FWHM for SNRs up to 85. For a rate of $30\ s^{-1}$ the acceptable signal-to-noise ratios decrease to 5, 11, 20 and 30 for extracted window widths of 3, 4, 5 and 6 times the FWHM, respectively. This comparison reflects the general trend that larger rates are more prone to inaccuracies caused by truncation. However, this analysis compares signals broadened by larger relaxation rates to narrow signals associated with smaller rates at equal signal-to-noise ratios. In practice, signals with higher relaxation rates have reduced SNRs and de-facto satisfy the more stringent requirement for a lower SNR. Additionally, Fig. 6 reflects that lower SNR provides relief from signal truncation. The increased error associated with low SNR overwhelms the inaccuracies introduced by truncation, and in such a case, the large error-bar-interval still contains the true rate with high probability. In summary, the FT/IFT procedure may introduce inaccuracies for excessive truncation of signal line-shapes, and these inaccuracies may not be accounted for by the experimental error associated with the estimated rate when truncation is performed at high SNR. To restore the rate's validity in such a case, users should either increase the width of the extracted region (and consider including additional oscillators if necessary) or they should use the symmetrization feature if possible. Figure 6 as well as Supplementary Fig. 2, which displays similar plots for rates of 5 and $100\ s^{-1}$, provide an empirical means to evaluate the validity of the estimated rates using FT/IFT.

Precision of rates obtained by FT/IFT

We have already discussed how the noise influences the precision of rates obtained by FT/IFT. However, the

precision of these rates is also affected by a reduction of the signal intensity stemming from apodization. The ability to isolate (groups of) signals in the frequency domain requires apodization of the time domain data. Indeed, the t_1 FID must be free of truncation to prevent "sinc-wiggle" artifacts that would otherwise make the extraction procedure inapplicable. This constraint is also present for other methods that extract relaxation rates following Fourier transform in t_1, such as line-shape analysis (Guenneugues et al. 1999) or simply line-width measurements (Chen and Tjandra 2009). In these previous works the authors restricted their analyses to data acquired either with large values of κ (Chen and Tjandra 2009) or subject to rates that were intrinsically large (Guenneugues et al. 1999), leading in both cases to fully relaxed FIDs that required no apodization. Unfortunately, this solution is not applicable in general because most proteins display a large dynamic range of relaxation rates. To prevent sinc-wiggles, the user would have to sample relaxation times dictated by the smallest rates. However, such a constraint would greatly reduce the signal-to-noise ratio of signals subject to larger relaxation rates, for which the majority of the data acquired would be noise. In a typical, well-folded protein, slowly relaxing residues are a minority, located in loops and terminal regions of the polypeptidic chain, and the experiment would be sub-optimal for a majority of signals. Therefore, it is preferable to design the acquisition to maximize the sensitivity for all residues and to overcome truncation artifacts with apodization. Indeed, Lefèvre and co-workers have already implemented a procedure that includes apodization, allowing accordion data to be analyzed in the frequency domain even with a value of $\kappa = 0$ (i.e. a reference experiment). However, their solution still required that all signals in an F_1 slice be fit simultaneously and did not take advantage of the frequency domain separation achieved by the Fourier transform. To account for the apodization, it may appear efficient to divide the reconstructed FID by the apodization function prior to fitting, thus restoring the signal-to-noise ratio and providing access to the (simpler) target functions of the MP procedure (Eqs. 7 and 12). However, such a process frequently introduces singularities in the reconstructed FID due to division by numbers tending toward zero. In contrast, inclusion of the apodization function in the target function requires only a point by point multiplication and does not increase the number of fitted parameters. Unfortunately, we note that apodization results in a loss of signal intensity that translates into a decrease in the precision of fitted parameters. For example, a signal with a relaxation rate of 5 s^{-1} is accompanied by an error of 0.056 s^{-1} with our implementation of the MP method. The error increases to 0.11 s^{-1} when fitting with the FT/IFT method, corresponding to a two-fold loss in precision. Clearly the added imprecision due to apodization decreases with increasing relaxation rate, as less signal is lost (see Supplementary Fig. 1). Thus, rates of 30 s^{-1} have only a 1.5-fold loss of precision in otherwise identical simulations. Nevertheless, apodization is necessary for a broad application of our FT/IFT method.

Accuracy of the symmetrization feature

The symmetrization feature within the FT/IFT method provides two potential advantages. Given partially overlapped peaks, it allows users to make a comparison between fitting both peaks simultaneously and fitting each signal separately after symmetry reconstruction. Symmetrization may also help users identify overlapped residues that had otherwise gone unnoticed. In such a case, a seemingly isolated signal would give rise to two substantially different rates when reconstructing from each half of the peak. The identification of overlapped peaks is clearly valuable in the context of protein functional studies, as improper fitting can result in relaxation rate outliers and subsequent erroneous mechanistic interpretations. However, symmetrization can introduce additional inaccuracies, and SARA implements a special procedure to minimize them. Symmetrization produces erroneous results when the peak maximum does not coincide exactly with a point in the spectrum, but this problem can be mitigated with extensive zero-filling prior to Fourier transformation in t_1. The most severe inaccuracy is introduced when a peak's maximum is centered exactly between two points in the spectrum. The magnitude of this inaccuracy is then a function of the distance between the true center of the peak and the closest point in the spectrum. That is, inaccuracies introduced by symmetrization will depend on the digital resolution of the spectrum. To quantify this effect and to design a solution in SARA, we have performed simulations at various relaxation rates and signal-to-noise ratios for signals subject to this worst-case scenario. Figure 7 displays the accuracy of rates obtained by symmetrization as a function of the number of points encompassed by the FWHM. For each peak, symmetrization was performed by using either the right half, which contains one more point (Fig. 7b top), or the left half, which contains one fewer point (Fig. 7b bottom). Thus, the former strategy artificially broadens the signal and overestimates the rate (Fig. 7a, green) while the latter narrows the signal and underestimates the rate (Fig. 7a, blue). Comparison of simulations for rates of 1, 5, and 100 s^{-1} (Fig. 7a) reveals that larger relaxation rates are the least affected by variation in spectral resolution, whereas smaller rates require a large digital resolution to be accurate. This observation simply reflects that a sharp signal will be more sensitive to inaccuracies stemming from a symmetrization not

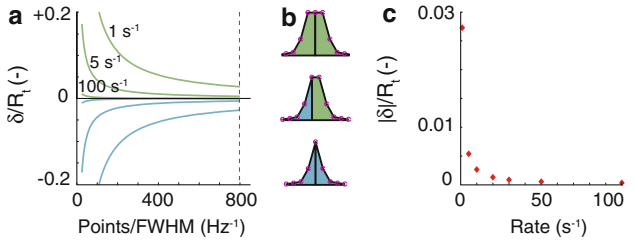

Fig. 7 Inaccuracy introduced by symmetrization. The worst-case scenario for symmetrization occurs when the true peak maximum lies exactly between two points in the spectrum [*purple dots* (**b**) center]. In this case symmetrization would reconstruct a peak that is too narrow [(**b**) bottom, *blue*] or too wide [(**b**) top, *green*]. In either case, the inaccuracy decreases when increasing the extent of zero-filling prior to Fourier transformation. Panel (**a**) displays the inaccuracy when fitting either the narrow (*blue*, bottom) or wide (top, *green*) symmetrized signal as a function of the number of points spanning the FWHM of the peak. Symmetrization simulations were performed using the same parameters as in Figs. 5 and 6 except that in each case the peak frequency was offset from zero by one half of the digital resolution. The simulated decay curves were anodized, zero-filled to multiple values based on the known FWHM and subsequently Fourier transformed, symmetrized, inverse Fourier transformed and fit. Curves are displayed for relaxation rates of 1, 5 and 100 s^{-1}. The *dashed line* in panel (**a**) indicates the amount of zero-filling used in panel (**c**). Panel (**c**) displays the magnitude of the inaccuracy as a function of the relaxation rate at a zero-filling level leading to 800 points per FWHM for each rate. Rates as low as 5 s^{-1} are accurate to within 0.6 % under such conditions. SARA includes an automated procedure using this level of zero-filling to minimize inaccuracies introduced by symmetrization

performed around its true maximum. Figure 7a reveals that fitting after symmetrization remains accurate to within 2.75 % for a rate as low as 1 s^{-1} if the signal's FWHM contains 800 points (vertical red dashed line). This criterion was hence chosen as a means to minimize inaccuracies for all rates. Figure 7c shows normalized accuracy (δ/R_t) as a function of relaxation rate under the 800-points-per-FWHM condition (dashed line in Fig. 7a). All rates larger than or equal to 5 s^{-1} remain accurate within 0.6 % following symmetrization. In a manner similar to that described for Fig. 6, the validity of rates was assessed by monitoring precision and accuracy as a function of SNR for various values of zero-filling (Supplementary Fig. 3). When symmetrizing signals with 800 points per FWHM, even a rate of 1 s^{-1} and SNR higher than 250 are associated with errors-bars providing a high confidence level (above 62 % when assuming a Gaussian distribution). This criterion was used to design a procedure to limit inaccuracies introduced during symmetrization in an automated manner. For any spectral width and apodization function provided by the user, SARA automatically calculates the amount of zero-filling needed for accurate symmetrization. More specifically, a signal decaying with a rate of 1 s^{-1} is simulated and apodized with the same function as the user-supplied spectrum. It is then Fourier transformed so that its FWHM can be measured, and SARA calculates the amount of zero-filling needed to reach 800 points within this FWHM (ZF_{sym}). Symmetrization is not performed until the user has defined the extraction window within the user-loaded spectrum and the resolution of the data is not modified until after this stage. After selecting one of the "Sym" options (Fig. 3), SARA extracts the corresponding t_1 FID from the interferogram, zero-fills it to ZF_{sym} points, and re-Fourier transforms the data. Symmetrization is then performed on this single, high-resolution spectral trace before inverse Fourier transformation and fitting. This procedure ensures that symmetrization does not introduce inaccuracies higher than 0.6 % for rates as low as 5 s^{-1}.

Concerted use of MP and FT/IFT as a means to monitor accuracy and improve precision

Although the features offered by symmetrization and the FT/IFT method in general do come with caveats, they provide a necessary alternative to the MP method that can be used to validate rates and identify setbacks. In order to capture the advantages of both procedures, fitting strategies may be developed and tested using the FT/IFT method and subsequently implemented in the MP method to maximize precision.

In the end, we have found that the MP and FT/IFT procedures are highly complementary. The MP procedure has the indisputable advantage of analyzing the raw data and therefore maximizes precision in the fitted parameters because no apodization is necessary. However, the FT/IFT procedure may identify signals that were fit poorly in the MP protocol. In particular, the FT/IFT method overcomes limitations due to the cumulative effects of numerous and overlapping signals in the indirect dimension. Thus, we recommend beginning the analysis with the FT/IFT procedure because it calls for the closest inspection by the user. This allows the user to clearly identify overlaps as well as "bad" signals (e.g. signals close to t_1 noise or axial peak artifacts). We then suggest using the user-interactive MP protocol to maximize precision. Users may minimize the risk of converging to a local minimum in the MP method by using the amplitudes, frequencies and rates obtained in the FT/IFT procedure as input to the optimization. Additionally, parameters known from independent measurements can be incorporated or held fixed during fitting. The local minima we encountered when applying the MP method to ubiquitin highlight the importance of user intervention when applying non-linear fitting methods and emphasize the need for an interactive and user-friendly software environment such as SARA. Because SARA provides a visual comparison between the data and its fitted function in both procedures, any discrepancy between them can be identified and possibly resolved. We believe such a

protocol allows researchers to analyze accordion data reliably and to clearly identify potential inaccuracies.

Conclusions

We have developed a graphical, user-friendly software package for processing data from accordion relaxation experiments. It harnesses the development speed and accessibility of the MATLAB environment to bring accordion data analysis to a wide audience. In order to provide users with the tools necessary to investigate a wide array of proteins, we have implemented two analysis methods in SARA that span the majority of approaches presented in the literature so far. The MP method offers the highest possible fitting precision, but it may be applied sub-optimally without close inspection by the user. As an alternative, we have developed the FT/IFT method which harnesses the ability of the Fourier transform to separate signals prior to analysis. We provide guidelines for its appropriate application and discuss the limitations of its use. We also include a symmetrization feature that may be used both to identify cases of strong overlap and resolve cases of slight overlap. The strength of SARA lies in its ability to evaluate and fit data from multiple perspectives, so we suggest a protocol relying on both the MP and FT/IFT analysis procedures that, together, ensure a reliable and precise estimation of relaxation rates using a robust and interactive software environment. The accordion method is an efficient way to measure relaxation rates in a fraction of the time needed using traditional experiments. We hope that the tools we have designed for analyzing relaxation rates will facilitate and promote routine application of the accordion method during biological studies and thereby encourage a more systematic investigation of protein dynamics by NMR. SARA is available for download at http://www.jhttonline.jhu.edu/TechnologyDetail.aspx?TechID=178F4955-3FB5-446A-9BAF-03D5ACF90264&JHURef=C12656 or upon request from the corresponding author.

Acknowledgments This work was supported by NIH, Grant R01-GM104257.

References

Akke M, Brüschweiler R, Palmer AG (1993) NMR order parameters and free energy: an analytical approach and its application to cooperative calcium (2+) binding by calbindin D9 k. J Am Chem Soc 115:9832–9833

Barkhuijsen H, DeBeer R, VanOrmondt D (1987) Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. J Magn Reson 557:553–557

Bodenhausen G (1982) Direct determination of rate constants of slow dynamic processes by two-dimensional "accordion" spectroscopy in nuclear magnetic resonance. J Am Chem Soc 3:1304–1309

Bodenhausen G, Ernst R (1981) The accordion experiment, a simple approach to three-dimensional NMR spectroscopy. J Magn Reson 45:367–373. doi:10.1016/0022-2364(81)90137-2

Carr PA, Fearing DA, Palmer AG (1998) 3D accordion spectroscopy for measuring 15 N and 13CO relaxation rates in poorly resolved NMR spectra. J Magn Reson 132:25–33. doi:10.1006/jmre1998.1374

Chen K, Tjandra N (2009) Direct measurements of protein backbone 15 N spin relaxation rates from peak line-width using a fully-relaxed Accordion 3D HNCO experiment. J Magn Reson 197:71–76. doi:10.1016/j.jmr.2008.12.001

Delaglio F, Grzesiek S, Vuister GW et al (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Eisenmesser EZ, Millet O, Labeikovsky W et al (2005) Intrinsic dynamics of an enzyme underlies catalysis. Nature 438:117–121. doi:10.1038/nature04105

Guenneugues M, Gilquin B, Wolff N et al (1999) Internal motion time scales of a small, highly stable and disulfide-rich protein: a 15 N, 13C NMR and molecular dynamics study. J Biomol NMR 14:47–66

Gunther U, Ludwig C, Ruterjans H (2000) NMRLAB-advanced NMR data processing in matlab. J Magn Reson 145:201–208. doi:10.1006/jmre 2000.2071

Henzler-Wildman KA, Thai V, Lei M et al (2007) Intrinsic motions along an enzymatic reaction trajectory. Nature 450:838–844. doi:10.1038/nature06410

Herman P, Lee JC (2012) The advantage of global fitting of data involving complex linked reactions. In: Fenton AW (ed) Allostery. Springer New York, New York, pp 399–421

Korzhnev DM, Kay LE (2008) Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. Acc Chem Res 41:442–451. doi:10.1021/ar700189y

Mandel AM, Palmer AG (1994) Measurement of relaxation-rate constants using constant-time accordion NMR spectroscopy. J Magn Reson, Ser A 110:62–72. doi:10.1006/jmra1994.1182

Rabier P, Kieffer B, Koehl P, Lefèvre J-F (2001) Fast measurement of heteronuclear relaxation: frequency-domain analysis of NMR accordion spectroscopy. Magn Reson Chem 39:447–456. doi:10.1002/mrc.870

Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021–1025. doi:10.1038/nature05858

Wand AJ (2013) The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. Curr Opin Struct Biol 23:75–81. doi:10.1016/j.sbi.2012.11.005