FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data

Roger Cole & J. Patrick Loria*
Department of Chemistry, Yale University, P.O. Box 208107, New Haven, CT 06520, U.S.A.

Abstract

Herein we describe the program FAST-Modelfree for the fully automated, high throughput analysis of NMR spin-relaxation data. This program interfaces with the program Modelfree 4.1 and provides an intuitive graphical user interface for configuration as well as complete standalone operation during the model selection and rotational diffusion parameter optimization processes. FAST-Modelfree is also capable of iteratively assigning models to each spin and optimizing the parameters that describe the diffusion tensor. Tests with the protein Ribonuclease A indicate that using this iterative approach even poor initial estimates of the diffusion tensor parameters will converge to the optimal value within a few iterations. In addition to improving the quality of the final fit, this represents a substantial timesaving compared to manual data analysis and minimizes the chance of human error. It is anticipated that this program will be particularly useful for the analysis and comparison of data collected under different conditions such as multiple temperatures and the presence and absence of ligands. Further, this program is intended to establish a more uniform protocol for NMR spin-relaxation data analysis, facilitating the comparison of results both between and within research laboratories. Results obtained with FAST-Modelfree are compared with previous literature results for the proteins Ribonuclease H, *E. coli* glutaredoxin-1 and the Ca^{2+}-binding protein S100B. These proteins represent data sets collected at both single and multiple static magnetic fields and which required analysis with both isotropic and axially symmetric rotational diffusion tensors. In all cases results obtained with FAST-Modelfree compared favorably with the original literature results.

Introduction

Nuclear magnetic spin-relaxation rates provide sensitive probes of the dynamic behavior of proteins and nucleic acids in solution. In particular, laboratory frame spin-relaxation experiments can be analyzed within the model-free formalism (Clore et al., 1990; Lipari and Szabo, 1982a, b) to provide details on macromolecular dynamics with atomic resolution. Information on the spatial restriction of the bond vector under investigation (S^2, the generalized order parameter), the timescale for internal motion (τ_e), the timescale for macromolecular tumbling (τ_m) and the contributions to the transverse relaxation rate due to

μs-ms chemical exchange phenomena (R_{ex}) can be obtained. The utility of these experiments for the study of the physical properties of macromolecules is evident from the interesting applications that have appeared. In addition to identifying flexible and rigid regions in proteins, spin-relaxation data has been used to measure free energy changes upon cooperative Ca^{2+} binding to the protein calbindin (Akke et al., 1993), entropy changes in a protein folding transition for an SH3 domain (Yang and Kay, 1996), and entropic contributions involved in the disorder-to-order transition for the transcription factor GCN4 (Bracken et al., 1999). Changes in backbone S^2 values have been determined upon small molecule binding (Stivers et al., 1996; Zidek et al., 1999) and upon site-specific mutation (Cameron and Benkovic, 1997) and subsequently implicated in enzyme function. The utility of NMR

*To whom correspondence should be addressed. E-mail: patrick.loria@yale.edu

for dynamics analysis relies upon the precise and re-producible extraction of parameters for internal and overall motion from the experimental spin-relaxation data.

Obtaining values for the amplitude and timescale of bond vector motion requires fitting of measured spin-relaxation data to one of several model forms of the spectral density equations followed by sta-tistical analysis for model comparison. The most common computer program for performing such data fitting and statistical analysis is Modelfree (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html). In fact, a survey of the scien-tific literature for 2001 in which S^2 values are reported reveals that 50% of all analysis was performed with Modelfree (http://isi6.isiknowledge.com/portal.cgi). This number far surpasses any other single program currently in use for fitting spin-relaxation data.

Unfortunately, there are also several drawbacks to using Modelfree. First, during the analysis the user is required to intervene at several instances during the model selection process. While protocols for this have been published (Mandel et al., 1996), modifications to these procedures are often necessary, such as when data is acquired at multiple static magnetic fields. In addition to the required files containing raw spin-relaxation data, four properly formatted files need to be manually prepared for input to Modelfree. Finally, after fitting the data for each spin to every model a series of statistical selection procedures are employed to determine the best fit motional model (Mandel et al., 1995). This step is frequently done manually and can be tedious. Manual intervention at each stage of the fitting process substantially increases the likelihood of introducing errors in the final outcome. Furthermore, the need for manual input is quite time-consuming. These issues become magnified if multiple relaxation data sets are being analyzed. Recently there has been increased interest in comparing dynamics at multi-ple temperatures (Mandel et al., 1996; Stone, 2001), static magnetic fields (Engelke and Ruterjans, 1997; Kroenke et al., 1999; Tjandra et al., 1996), or in the presence and absence of ligands (Alexandrescu et al., 1998; Fushman et al., 1994; Stivers et al., 1996). It is clear that an improved method of analysis is highly desirable and necessary to reliably and accu-rately complete such studies in a reasonable amount of time (Dosset et al., 2000). Undoubtedly many indi-vidual research laboratories have developed scripts for automating various aspects of modelfree data fitting,

however a single robust and automated program for performing this analysis has not been reported.

Here we describe the program FAST-Modelfree (Facile Analysis and Statistical Testing for Model-free). FAST-Modelfree interfaces with the program Modelfree 4.1 and is designed to completely automate the task of input file preparation, model selection, ro-tational diffusion parameter optimization and process iteration until a final global diffusion tensor is reached which is self consistent with the set of assigned mod-els. FAST-Modelfree is written in Perl and has a graph-ical Tk based interface that allows for easy program setup and execution. Although FAST-Modelfree is completely automated, almost every parameter avail-able in Modelfree may be set to the user's preference. Further, the direct output of all files from Modelfree is automatically stored and can be inspected by the interested user.

This automated method represents a substantial advance in the analysis of laboratory frame NMR spin-relaxation data. This approach minimizes the possibility of errors and provides a consistent protocol for model assignment and reproducibility of results from within and between research laboratories. Fur-ther, by eliminating the need for human intervention FAST-Modelfree greatly diminishes the time needed for complete analysis. This approach is especially use-ful if several sets of relaxation data are to be analyzed. Finally, given a set of R1, R2, and NOE data a be-ginning user can start a Modelfree run in well under 5 minutes. One goal of this program is to make the model selection process as transparent as possible.

Materials and methods

All calculations were performed on Intel P4 based computers running the Linux operating system. FAST-Modelfree is written in the Perl program-ming language and was executed with Perl ver-sion 5.6.1 with the optional Tk module. FAST-Modelfree is available for download from the au-thor's web site (http://xbeams.chem.yale.edu/~loria/) and is distributed under the terms of the GNU public license. Perl and the optional Tk module are both available for download from the CPAN site. (http://www.cpan.org). Modelfree 4.1 was ob-tained from website of Professor Arthur G. Palmer (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html). FAST-Modelfree was tested using previously published relaxation data from three

proteins. The data analyzed were from ribonuclease H (Mandel et al., 1996), a 2.3 mM sample of oxidized glutaredoxin-1 (Kelley et al., 1997), and the apo form of the Ca^{2+} binding protein S100B (Inman et al., 2001)

The enzyme ribonuclease H (RNase H) was analyzed using R1, R2 and NOE relaxation data collected at 500 MHz (Mandel et al., 1996). An isotropic diffusion tensor was employed and therefore analysis did not require the use of a PDB structure file. Similarly, the oxidized form of *E. coli* glutaredoxin-1 was analyzed using R1, R2 and NOE relaxation data collected at 500 MHz (Kelley et al., 1997) with an isotropic diffusion tensor, using data downloaded from the Indiana Dynamics Database (http://pooh.chem.indiana.edu/IDD.html). Finally, the protein apo-S100b was analyzed with two sets of R1, R2 and NOE relaxation data collected at 600 MHz and 400 MHz (Inman et al., 2001). An axially symmetric diffusion tensor was employed and the atomic coordinates were obtained from the PDB file 1B4C (Drohat et al., 1999). The PDB file was prepared for use by first adding hydrogen atoms using the program MolMol (Koradi et al., 1996) followed by translation to the center of mass with the program 'pdbinertia' (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html). The program 'r2r1_diffusion', which uses the method described by Tjandra et al. (1995) was then used to estimate an axially symmetric diffusion tensor and orient the coordinate system of the PDB file to the principal axis of the diffusion tensor (Tjandra et al., 1995). This final PDB file was then used as input to FAST-Modelfree. The length of the N-H bond and value for ^{15}N-CSA were identical to those used in original publications. Residues that did not fit to models 2 or 3 were assigned to model 1 if the sum squared error for that spin, Γ_i, was less than 20; the cutoff value of Γ_i can be set to the user's preference. When parameters describing the rotational diffusion tensor (τ_m, $D_{||}$, $D_{\perp}$, etc.) were optimized, data for residues that had been initially fit with S^2 values less than 0.7 were excluded from the optimization.

Theory

While the theory of the model-free formalism (Clore et al., 1990; Lipari and Szabo, 1982a, b) and model selection protocols (Mandel et al., 1995) have been previously elaborated in great detail, we present a brief overview for completeness and to facilitate following the FAST-Modelfree procedure described below. The relaxation of the amide ^{15}N nucleus is determined by the dipolar interaction with its attached proton and the chemical shift anisotropy (CSA) mechanism of ^{15}N (Abragam, 1961):

$$R_1 = \left(d^2/4\right)\left[J\left(\omega_H - \omega_N\right) + 3J\left(\omega_N\right) + 6J\left(\omega_H + \omega_N\right)\right] + c^2 J\left(\omega_N\right) \tag{1}$$

$$R_2 = \left(d^2/8\right)\left[4J(0) + J\left(\omega_H - \omega_N\right) + 3J\left(\omega_N\right) + 6J\left(\omega_H\right) + 6J\left(\omega_H + \omega_N\right)\right] + \left(c^2/6\right)\left[4J(0) + 3J\left(\omega_N\right)\right] + R_{ex} \tag{2}$$

$$NOE = 1 + \left(d^2/4R_1\right)\left(\gamma_N/\gamma_H\right)\left[6J\left(\omega_H + \omega_N\right) - J\left(\omega_H - \omega_N\right)\right], \tag{3}$$

in which $d = \left(\mu_0 h \gamma_H \gamma_N / 8\pi^2\right)\left\langle r_{NH}^{-3}\right\rangle$, $c = \omega_N \Delta\sigma/\sqrt{3}$, μ_0 is the permeability of free space, h is Planck's constant, γ_H and γ_N, ω_N and ω_H are the magnetogyric ratios and Larmor frequencies of ^{1}H and ^{15}N, respectively, r_{NH} is the N-H bond length and typically values are 1.02 or 1.04 Å, $\Delta\sigma$ is the ^{15}N CSA value and R_{ex} describes the loss of transverse magnetization due to chemical or conformational exchange. In the model-free analysis the spectral density function corresponding to an isotropic diffusion tensor is modeled as (Clore et al., 1990; Lipari and Szabo, 1982a, b):

$$J(\omega) = \frac{2}{5}\left[\frac{S^2 \tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S_f^2)\tau_f'}{1 + \left(\omega\tau_f'\right)^2} + \frac{(S_f^2 - S^2)\tau_s'}{1 + \left(\omega\tau_s'\right)^2}\right], \tag{4}$$

in which τ_m is the overall rotational correlation time of the macromolecule, τ_f and τ_s are the effective correlation times for fast ($\tau_f < 100 - 200$ ps) and slow ($\tau_f < \tau_s < \tau_m$) internal motion, $\tau_f' = \tau_f \tau_m / \left(\tau_f + \tau_m\right)$, $\tau_s' = \tau_s \tau_m / \left(\tau_s + \tau_m\right)$, $S^2 = S_f^2 S_s^2$ are the squares of the generalized order parameters where S_f^2 and S_s^2 are the order parameters for fast and slow internal motion, respectively. In the case of an axially symmetric diffusion tensor Equation 4 becomes

$$J(\omega) = \frac{2}{5}\sum_{j=1}^{3} A_j\left[\frac{S^2 \tau_j}{1 + (\omega\tau_j)^2} + \frac{(1 - S_f^2)^2 \tau_f'}{1 + (\omega\tau_f')^2} + \frac{(S_f^2 - S^2)\tau_s'}{1 + (\omega\tau_s')^2}\right]. \tag{5}$$

The parameters are defined similarly to the case of isotropic diffusion except that $\tau_1^{-1} = 6D_\perp$, $\tau_2^{-1} = D_{||} + 5D_\perp$, $\tau_3^{-1} = 4D_{||} + 2D_\perp$, $\tau_s' = \tau_j\tau_e/(\tau_j + \tau_e)$, where τ_e is either τ_f or τ_s, $A_1 = (3\cos^2\theta - 1)^2/4$, $A_2 = 3\sin^2\theta\cos^2\theta$, $A_3 = (3/4)\sin^4\theta$, and θ is the angle between the N-H bond vector and the unique axis of the diffusion tensor.

Extraction of the motional parameters requires fitting the experimental NMR relaxation data to one of five dynamical models derived from Equation 4 or Equation 5. Each model contains parameters specific to the dynamics of a particular spin, such as S^2 and R_{ex}, as well as parameters describing the rotational diffusion tensor which are the same for every spin, such as τ_m, $D_{||}$ and $D_\perp$. It is important to note that, using Modelfree, one cannot fit parameters specific to a particular spin without knowing, or at least estimating, rotational diffusion tensor parameters that are identical for every spin. Similarly, one cannot fit diffusion tensor parameters without knowing which spin specific dynamics parameters to include for each residue. A list of models and the corresponding spin specific parameters is given below.

Model 1 S^2
Model 2 $S^2, \tau_e = \tau_f$
Model 3 S^2, R_{ex}
Model 4 $S^2, \tau_e = \tau_f, R_{ex}$
Model 5 $S_f^2, S^2, \tau_e = \tau_s$

To begin use of FAST-Modelfree, a set of three delimited files containing R1 $\pm$ error, R2 $\pm$ error, NOE $\pm$ error for each amino acid residue is needed. There are no constraints on the format of these files. The names of the data files are entered in the FAST-Modelfree input window (Figure 1). If the program SPARKY (Goddard and Kneller) is used for NMR data visualization, the program 'Sparky2rate' is provided (http://xbeams.chem.yale.edu/~loria/) which automatically extracts peak heights and relaxation times from the output generated by the Sparky command 'rh'. 'Sparky2rate' uses the program 'Curvefit' (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html) to fit the data to a single exponential decay and perform error analysis.

Beginning an analysis with FAST-Modelfree consists of two steps. First, the user executes the program 'setupFMF' to open the graphical user interface window. Parameters regarding statistical cutoffs, model assignment and diffusion tensor parameters may be chosen quickly and easily through a simple menu (Figure 1). When all parameters have been set as desired, the configuration is saved and the main program is executed.

The utility of FAST-Modelfree is that it automates the model selection strategy described in Figure 9 of Mandel et al. (1995). During the process of model selection, the rotational diffusion tensor parameters are fixed at their estimated values, relaxation data for all residues are fit to model 1 and the sum squared error for each spin, Γ_i, is calculated. Those residues in which Γ_i is less than the α-critical value of the simulated distribution of Γ_i are assigned to model 1. This comparison is performed automatically and the value for the α-critical value cutoff, adjustable by the user, is entered in the FAST-Modelfree parameter window. Residues for which model 1 is an insufficient descriptor of the experimental relaxation data are then fit to models containing two motional parameters (models 2 and 3). Model selection at this stage is based on comparison of Γ_i and F-statistics. The value for the F-test cutoff is user adjustable; the statistical comparison and model selection is performed automatically. If relaxation data are available at only a single magnetic field, this approach cannot be used to determine if the three parameters models (models 4 and 5) are appropriate because the number of degrees-of-freedom is zero. For this reason, if a particular spin does not fit models 1, 2 and 3 but has a Γ_i for model 1 almost within the statistical cutoff it is assigned to model 1 without testing against models 4 and 5. However, if Γ_i for model 1 is large, the particular spin may be assigned to model 4 or 5 if Γ_i for the three-parameter model is zero. If relaxation data are available at more than one static magnetic field, FAST-Modelfree will instead use the statistical selection protocol to determine if model 4 or 5 is appropriate. In these cases the synthetic data sets are generated with models 3 and 2 when performing the relevant F-test for models 4 and 5, respectively. FAST-Modelfree is also capable of recognizing if sufficient data for this protocol is present for some residues but not others, and will use the appropriate protocol for each.

Once all residues have been assigned to a model, the motional parameters for each spin are optimized simultaneously with the diffusion tensor parameters. Since models were initially selected using only an estimate of the global diffusion tensor parameters, there is no guarantee that the models chosen indeed represent the best fit solution. To overcome this problem, the newly calculated diffusion tensor parameters are

Figure 1. Graphical user interface for FAST-Modelfree. All information needed to run FAST-Modelfree is entered in this window, including data file names, statistical factors for model selection, motional parameter estimates, and rotational diffusion tensor estimates.

then used in a new round of model selection. If the true optimum diffusion tensor parameters had been initially used, then the correct model would have been selected for each spin and the optimization of diffusion tensor parameters would return the same globally optimum diffusion tensor used during the model selection process. As a result, by comparing the newly calculated diffusion tensor parameters with those used in the previous round of model selection one can gauge how close the current parameter set is to the global optimum. Further, by iterating the process of model selection and diffusion tensor parameter optimization one can eventually converge to a truly optimum set of parameters for both the diffusion tensor, which affects every spin, as well as the local dynamics parameters, which are specific to each spin. This iterative process is performed automatically by FAST-Modelfree and the relevant tolerances for convergence may be specified by the user through the graphical interface by changing the value in the convergence limit input field for the rotational diffusion parameters.

The automation built into the FAST-Modelfree program does not preclude user intervention and additionally records the results of all iterations in user-readable text files for inspection. At the end of each iteration, a log file is updated with the model selections and tensor parameters from the current iteration. An output file detailing the dynamics parameters is created as well, and in the case of an axially symmetric diffusion tensor a PDB file is created with the coordinates rotated to the principal axis of the diffusion tensor.

Results

The robustness of the Modelfree program should not be sacrificed for the convenience and speed of using FAST-Modelfree. To judge this we compared published results obtained using Modelfree with those obtained with FAST-Modelfree. We selected relaxation data obtained at a single as well as multiple static magnetic fields. In addition, we have chosen proteins for which an isotropic or axially symmetric diffusion
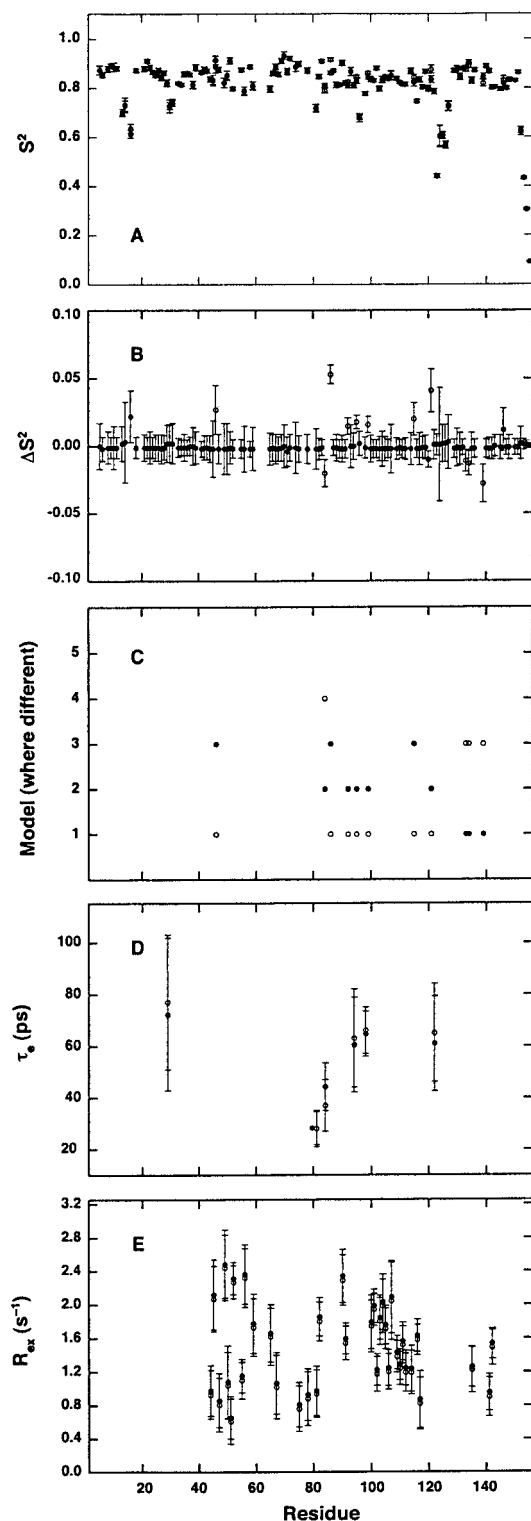
tensor was appropriate. The results of this comparison are depicted in Figures 2, 3 and 4.

In the case of the protein RNase H a rotational correlation time, $\tau_m = 9.26 \pm 0.01$ ns is obtained that compares favorably with the value of 9.28 ± 0.02 ns reported previously. Of the 125 residues analyzed, FAST-Modelfree assigned 15 residues to different models than previously reported. This value is within the tolerance reported for unique spin-relaxation data sets acquired on the same protein (Mandel et al. 1996). Residues Asn 15, Gly 20 and Glu 61 could not be fit to any model, whereas in previous reports they fit models 5, 5 and 2 respectively. In total only 5 residues yielded S^2 values which were not within error of the previously reported results. For each of these five residues a different model was found than previously reported, and the relative difference in S^2 was in all 5 cases less than 6%.

The results for oxidized glutaredoxin-1 agree similarly well with previously published values. The calculated value of τ_m equals 7.88 ± 0.02 ns and is in close agreement with the literature value of 7.95 ± 0.03 ns (Kelley et al., 1997). Although the values are not within the reported errors, they differ by less than 1% from the average. Data from 72 residues were analyzed, and of these 14 residues were assigned to different models than previously reported. Two residues, Asp 84 and Ala 85, which were reported to fit no model were both found to fit model 5. Similarly, no model was found to fit the data for residue Asp 30 although this spin had previously been assigned to model 2. Only 4 residues were found whose calculated S^2 values were not within error of the published results. Each of these 4 residues was assigned to a different model than those reported in the literature.

The protein S100B has an axially symmetric rotational diffusion tensor and a correlation time of 7.80

Figure 2. Comparison of FAST-Modelfree derived motional parameters for RNase H with previously published results. (A) Comparison of S^2 values calculated with FAST-Modelfree with literature values. (B) Difference in S^2 values between FAST-Modelfree results and literature values ($S^2_{FMF} - S^2_{published}$). In panel B open circles denote residues where the model assigned by FAST-Modelfree is different from the literature model. Filled circles indicate residues where the literature model matches the FAST-Modelfree assigned model. (C) Comparison of models assigned by FAST-Modelfree with literature values for those cases where a model is assigned. (D) Comparison of τ_e values calculated using FAST-Modelfree to literature values. (E) Comparison of R_{ex} values calculated using FAST-Modelfree to literature values. In panels A, C, D and E filled circles represent values calculated with FAST-Modelfree, open circles indicate published values.
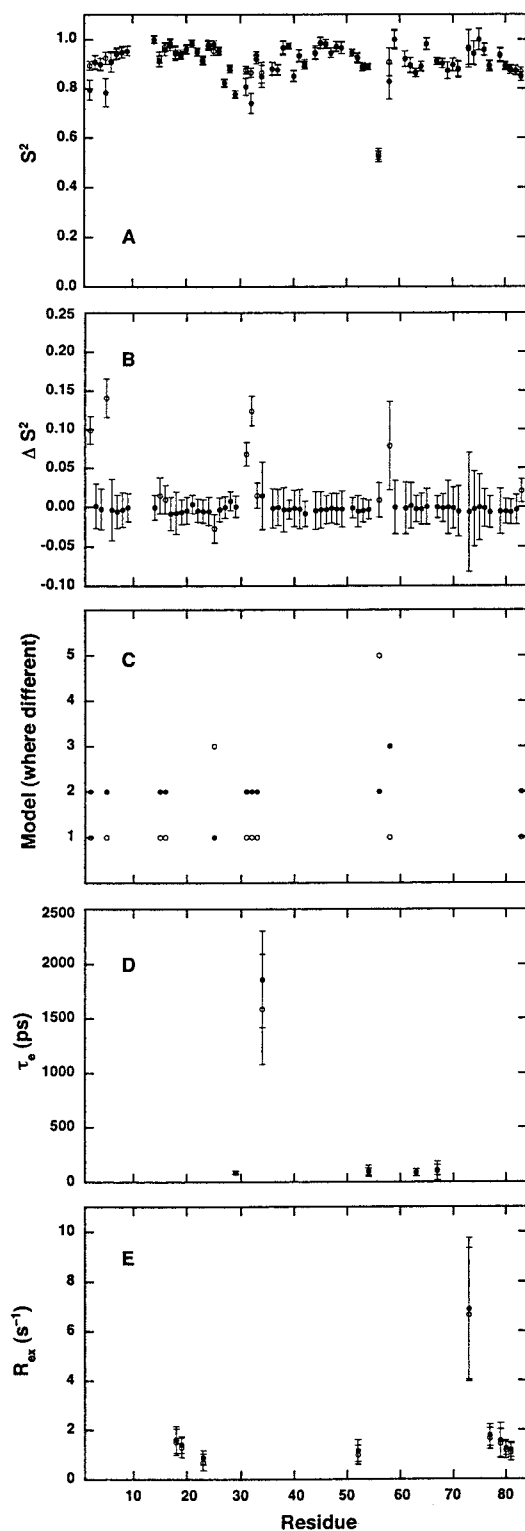
ns (Inman et al., 2001). This value differs slightly from that obtained with FAST-Modelfree in which $\tau_m = 7.91 \pm 0.01$ ns. This represents a difference of 1.4% from the average. $D_{||}/D_{\perp}$ was calculated to be 1.17 ± 0.01, which compares favorably with the published result of 1.15 ± 0.02. Of the 80 residues analyzed, 14 were assigned to different models. Five residues were found whose calculated S^2 values were not within error of previous results. In each of these cases different models were assigned, and the calculated S^2 values were no more than 5% different than those in the literature. In the original publication (Inman et al., 2001), a group of residues could not be fit to models 1–5 and were subsequently tested against more complex models as well as the local diffusion approximation. This method is currently not implemented in FAST-Modelfree. As a result, residues in this group were tested against models 1–5 using FAST-Modelfree to verify that they indeed do not fit any of these models, but no additional analysis of these residues was performed.

FAST-Modelfree utilizes an iterative procedure of fitting the overall and internal dynamics parameters to obtain self-consistent results. The accuracy and robustness of this procedure was tested with data for the protein Ribonuclease A (RNase A). The importance of the initial estimates of the rotational diffusion tensor parameters on the final results obtained by the iterative process of model selection and diffusion tensor parameter optimization as employed by FAST-Modelfree was tested. The 'modelfree' analysis of this data has been previously reported by Cole and Loria (2002). The spin relaxation data were best fit with an axially symmetric rotational diffusion tensor with $D_{||}/D_{\perp} = 0.84 \pm 0.01$, $\Phi = 150.3 \pm 12.3°$, $\Theta = 14.9 \pm 3.0°$ and a rotational correlation time of 6.49

Figure 3. Comparison of FAST-Modelfree derived motional parameters for 2.3 mM oxidized glutaredoxin-1 with previously reported literature results. (A) Comparison of S^2 values calculated with FAST-Modelfree with literature values. (B) Difference in S^2 values between FAST-Modelfree results and literature values ($S^2_{FMF} - S^2_{published}$). In panel B open circles denote residues where the model assigned by FAST-Modelfree is different from the literature model. Filled circles indicate residues where the literature model matches the FAST-Modelfree assigned model. (C) Comparison of models assigned by FAST-Modelfree with literature values for those cases where a model is assigned. (D) Comparison of τ_e values calculated using FAST-Modelfree to published values. (E) Comparison of R_{ex} values calculated using FAST-Modelfree to literature values. In panels A, C, D and E filled circles represent values calculated with FAST-Modelfree, open circles indicate literature values.
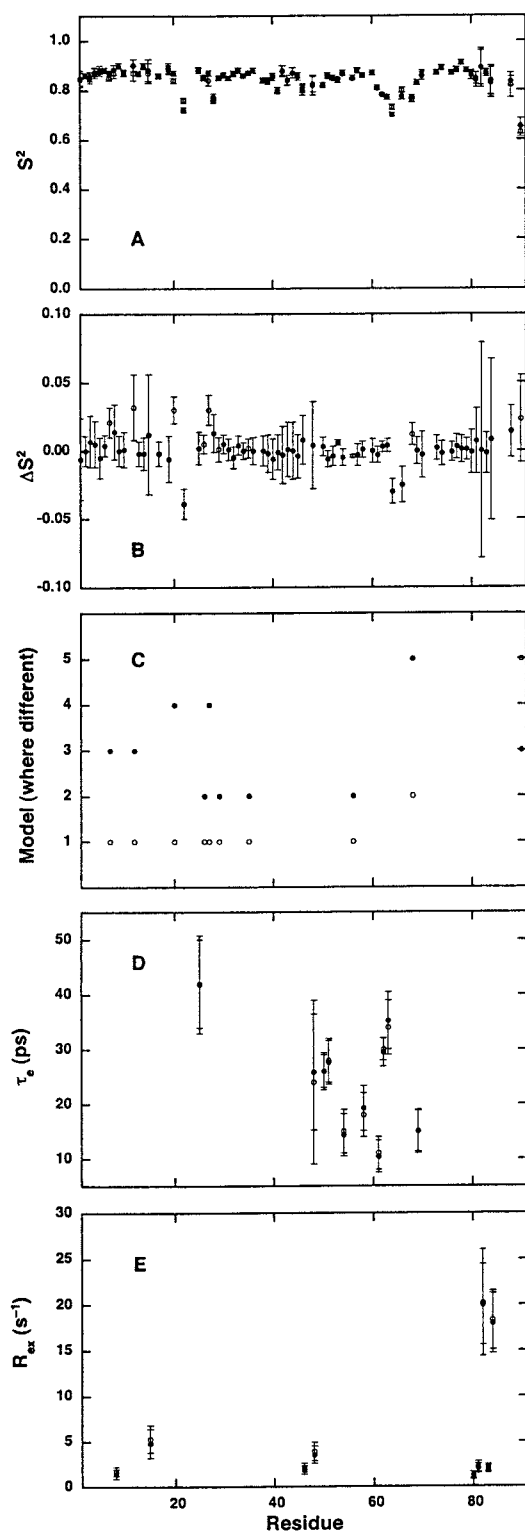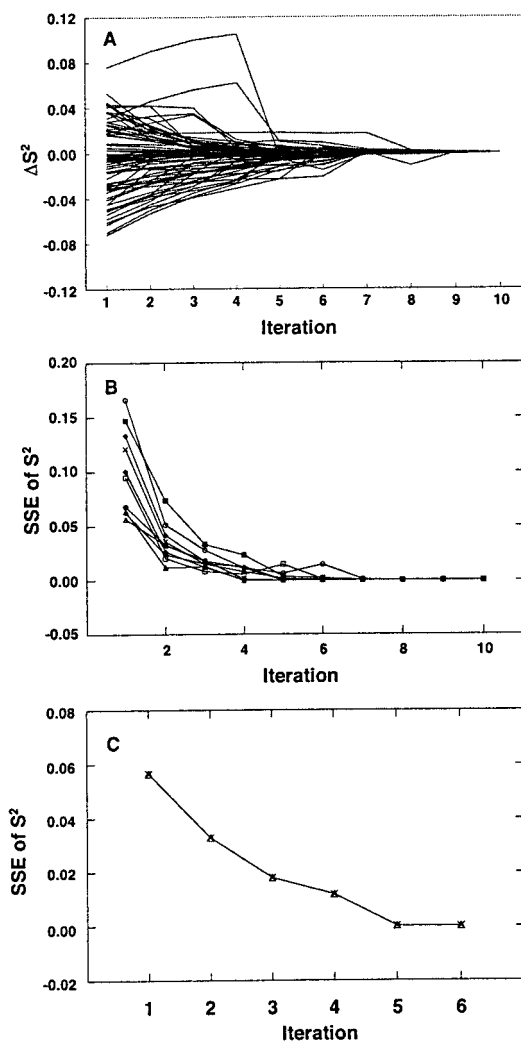
± 0.01 ns. To assess the ability of FAST-Modelfree to obtain these values even when the program is supplied with incorrect starting parameter estimates, ten different FAST-Modelfree runs were performed on the same set of relaxation data using randomly generated initial diffusion tensor parameters. For these runs, rotational correlation times were randomly distributed between 6.0 ns and 7.0 ns, $D_{||}/D_{\perp}$ was randomly distributed between 0.5 and 1.0, Θ was randomly distributed between 0° and 45° and Φ was randomly distributed between 0° and 360°. While convergence generally took several additional iterations than when the initial estimates of the rotational diffusion tensor parameters were close to the final value, in all cases identical results were achieved within 9 iterations (ca. 8–10 h). This convergence is illustrated in Figure 5. In addition for one of the starting conditions, the iterative procedure was performed manually to assure that programming errors in FAST-Modelfree are not present. Identical results are obtained by the manual and automated results (Figure 5C).

Discussion

As can be seen from Figure 5, the iterative method employed by this program effectively removes any dependence of the derived motional parameters on the initial estimates of the rotational diffusion tensor parameters. Since both the models selected and the motional parameters derived from those models are highly dependent on the diffusion tensor parameters, it is expected that this method will enhance the reproducibility and reliability of data analysis.

FAST-Modelfree obtains nearly identical motional parameters with previously reported values. For the protein RNase H, the global correlation time was

Figure 4. Comparison of FAST-Modelfree derived motional parameters for S100B with previously reported literature results. (A) Comparison of S^2 values calculated with FAST-Modelfree with literature values. (B) Difference in S^2 values between FAST-Modelfree results and literature values ($S^2_{FMF} - S^2_{published}$). In panel B open circles denote residues where the model assigned by FAST-Modelfree is different from the literature model. Filled circles indicate residues where the literature model matches the FAST-Modelfree assigned model. (C) Comparison of models assigned by FAST-Modelfree with literature values for those cases where a model is assigned. (D) Comparison of τ_e values calculated using FAST-Modelfree to literature values. (E) Comparison of R_{ex} values calculated using FAST-Modelfree to literature values. In panels A, C, D and E filled circles represent values calculated with FAST-Modelfree, open circles indicate literature values.

found to be within experimental error of the previous data. Although the analysis of S100B and oxidized glutaredoxin-1 data furnished correlation times that were not within error of previous results, the difference was, in both cases, extremely small (< 1.5%). In general, between 80% and 90% of the residues in each protein were assigned to the same models previously reported. This is illustrated in panel C of Figures 2–4 and this discrepancy can be traced to several factors. First, the statistical method for selecting models is based on Monte Carlo algorithms. In cases where data for a given residue fits two models almost equally well, the exact random number generator and random number seed can influence the model ultimately selected. Further, the method of assigning models allows for flexibility in assigning a spin to model 1 before test-

ing to see if models 4 or 5 are appropriate. The exact method in which this is done can vary depending on the preference of the experimenter, which can slightly alter the final results. Third, residues that exhibit a high degree of flexibility are typically excluded from the diffusion tensor optimization process. The FAST-Modelfree program accomplishes this by excluding residues with S^2 values less than a user specified cutoff value. Different methods exist for exclusion of flexible residues in the diffusion tensor optimization process and variations in this selection process can account for some of the differences (Kroenke et al., 1998; Mandel et al., 1995; Pawley et al., 2001; Seewald et al., 2000). Finally, the exact versions of Modelfree used in these comparisons differ. FAST-Modelfree calculations are done using Modelfree 4.1 while older versions were used in two (Kelley et al., 1997; Mandel et al., 1995) of the comparative studies. Slight differences in random number generator and Monte Carlo simulations exist between versions of Modelfree and could contribute to the differences in motional parameters observed here.

Given the computational effort in running Model-free, it is difficult to manually test all of the potential

sources that result in the different motional parameters between FAST-Modelfree and the reported literature values. However, to support the reasoning given above and to eliminate the additional possibility of a programming error, multiple runs of FAST-Modelfree were performed using different random number seeds for the proteins S100B, glutaredoxin, and RNase H. For RNase H, 10 different random number seeds were used in the data analysis. Of the fifteen residues with different reported motional models, one changed from model 3 (FAST-Modelfree data) to model 1 (literature value) (Mandel et al., 1995). The model change occurred for one of the 10 different runs. For this residue, when model 1 is chosen, the S^2 value is now within error with the reported value.

For the protein glutaredoxin, fifteen random numbers were used as seeds for the FAST-modelfree runs. Three of the fourteen residues assigned different models by FAST-Modelfree switched in some of the analysis to reported literature values. For example, residues 25, 31, and 33 changed to the reported model in 1, 9, and 5 cases. In instances where motional models assigned by FAST-Modelfree were identical to literature values, the orders parameters were also within error of the reported values.

For S100B, three different random numbers were used to seed the modelfree runs. Of the fourteen differences between the FAST-Modelfree results and the reported values, one (residue 56) undergoes a model change that is in agreement with the literature value (Inman et al., 2001). Again, when the motional model is identical for the two analyses, the calculated S^2 values are in agreement. These results underscore the sensitivity of model selection on the random number seed and additionally indicate the correlation in the calculated S^2 value with the chosen model (Mandel, et al., 1996). Not all of the residues are observed to change motional models and achieve agreement with the literature values. It is likely that a larger range of random number seeds would have resulted in further changes but the computational expense of many Modelfree calculations precluded further trials. In addition to the random number seed, values of Γ_i used in model selection will also influence the final results.

It is important to recognize that while some of the residues may be assigned to different models than previously reported, very few of the calculated S^2 values changed appreciably. In almost all of these cases the calculated S^2 and S_f^2 values were within error of the previous results, with the others generally within 6% of the literature values. As can be seen from panels A and B of Figures 2, 3 and 4 this is a relatively minor difference. In principle, this is not surprising since the cases where different models were assigned by FAST-Modelfree relative to the original reported values should represent residues were the particular motional parameters in question have small effects on the predicted values of R1, R2 and NOE. In those cases the small effects discussed above can influence the final model assignment of a particular spin.

As can be seen from panels D and E of Figures 2–4 the motional parameters τ_e, and R_{ex} were in every case found to compare very well with the literature values. Within the three proteins studied there were a total of 75 residues that required one or more of these motional parameters to adequately fit the data. In every case the values calculated by FAST-Modelfree were within error of the literature values. While model selection determines which, if any, of these motional parameters are used, these results indicate that the value of these two parameters is not strongly dependent on the exact protocol used for model selection and tensor optimization.

Conclusions

These results indicate that FAST-Modelfree provides a reliable method for analysis of laboratory frame spin-relaxation experiments. Through automation of each step, the chances for error are minimized and the amount of human intervention in the process of 'model free' analysis is greatly reduced. Further, it is hoped that by establishing a more consistent protocol it will become easier to compare results obtained by different laboratories.

Acknowledgements

References

Akke, M., Brüschweiler, R. and Palmer, A.G. (1993) *J. Am. Chem. Soc.*, **115**, 9832.

Alexandrescu, A.T., Rathgeb-Szabo, K., Rumpel, K., Jahnke, W., Schulthess, T. and Kammerer, R.A. (1998) *Protein Sci.*, **7**, 389.

Bracken, C., Carr, P.A., Cavanagh, J. and Palmer, A.G. (1999) *J. Mol. Biol.*, **285**, 2133.

Cameron, C.E. and Benkovic, S.J. (1997) *Biochemistry*, **36**, 15792.

Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990) *J. Am. Chem. Soc.*, **112**, 4989.

Cole, R. and Loria, J.P. (2002) *Biochemistry*, **41**, 6072.

Dosset, P., Hus, J.-C., Blackledge, M. and Marion, D. (2000) *J. Biomol. NMR*, **16**, 23.

Drohat, A.C., Tjandra, N., Baldisseri, D.M. and Weber, D.J. (1999) *Protein Sci.*, **8**, 800.

Engelke, J. and Ruterjans, H. (1997) *J. Biomol. NMR*, **9**, 63.

Fushman, D., Weisemann, R., Thüring, H. and Rüterjans, H. (1994) *J. Biomol. NMR*, **4**, 61.

Goddard, T. and Kneller, D.G. SPARKY 3, University of California, San Francisco.

Inman, K.G., Baldisseri, D.M., Miller, K.E. and Weber, D.J. (2001) *Biochemistry*, **40**, 3439.

Kelley, J.J.R., Caputo, T.M., Eaton, S.F., Laue, T.M. and Bushweller, J.H. (1997) *Biochemistry*, **36**, 5029.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51.

Kroenke, C.D., Loria, J.P., Lee, L.K., Rance, M. and Palmer, A.G. (1998) *J. Am. Chem. Soc.*, **120**, 7905.

Kroenke, C.D., Rance, M. and Palmer, A.G. (1999) *J. Am. Chem. Soc.*, **121**, 10119.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.*, **104**, 4546.

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.*, **104**, 4559.

Mandel, A.M., Akke, M. and Palmer, A.G. (1995) *J. Mol. Biol.*, **246**, 144.

Mandel, A.M., Akke, M. and Palmer, A.G. (1996) *Biochemistry*, **35**, 16009.

Pawley, N.H., Wang, C., Koide, S. and Nicholson, L.K. (2001) *J. Biomol. NMR*, **20**, 149.

Seewald, M.J., Pichumani, K., Stowell, C., Tibbals, B.V., Regan, L. and Stone, M.J. (2000) *Protein Sci.*, **9**, 1177.

Stivers, J.T., Abeygunawardana, C. and Mildvan, A.S. (1996) *Biochemistry*, **35**, 16036.

Stone, M.J. (2001) *Acc. Chem. Res.*, **34**, 379.

Tjandra, N., Feller, S.E., Pastor, R.W. and Bax, A. (1995) *J. Am. Chem. Soc.* **117**, 12562.

Tjandra, N., Wingfield, P., Stahl, S. and Bax, A. (1996) *J. Biomol. NMR*, **8**, 273.

Yang, D. and Kay, L.E. (1996) *J. Mol. Biol.*, **263**, 369.

Zidek, L., Novotny, M.V. and Stone, M.J. (1999) *Nat. Struct. Biol.*, **6**, 1118.