Modelfree

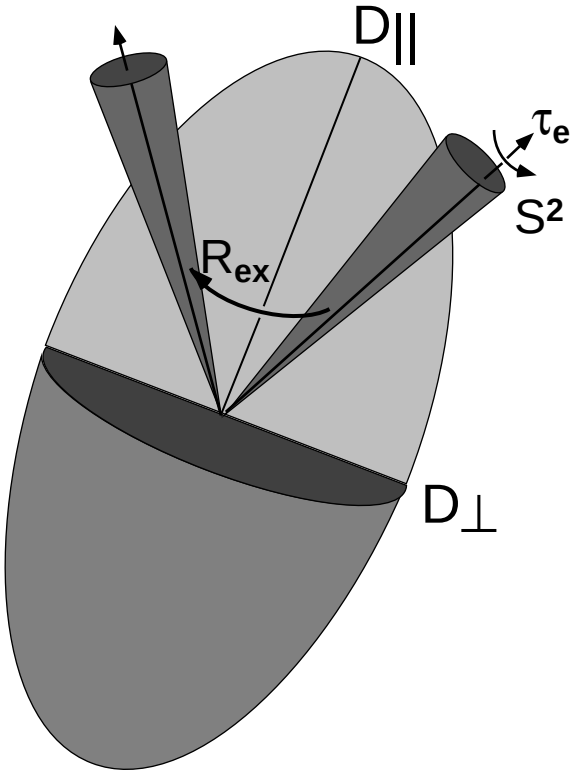


Table of Contents

License, Copyright, and Citations

Address any questions or comments to

Arthur G. Palmer
Department of Biochemistry and Molecular Biophysics
Columbia University
630 West 168th Street
New York, NY 10032

email: agp6@columbia.edu

The current version of the program should be referred to as Modelfree 4.15. The primary literature references for the Modelfree program are

Mandel, A. M.,Akke, M. & Palmer, A. G. (1995) *J. Mol. Bio 246*, 144-163.

Palmer, A. G.,Rance, M. & Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371-4380.

If you publish a paper that utilizes the Modelfree program, please email the reference information (authors, title, journal, volume, inclusive pages, year) to agp6@columbia.edu.

Introduction

Modelfree (version 4.0) is a program to fit the extended model free spectral density function to NMR spin relaxation data. The program can analyze the spin-lattice relaxation rate constant (R_1), the spin-spin relaxation rate constant (R_2), and the heteronuclear steady-state $\{^1H\}$-X nuclear Overhauser effect (*NOE*) for any combination of ^{13}C and ^{15}N spins at up to five static magnetic fields. The algorithm assumes dipolar and chemical shift anisotropy (CSA) relaxation mechanisms for R_1, R_2 and the *NOE* and includes an additive term to account for chemical exchange broadening of R_2. The exchange broadening parameter is scaled quadratically with respect to the static magnetic field if data for more than one field is available. Descriptions of the main principles utilized in the program for model selection, optimization and error analysis using Monte Carlo simulations have been published previously (Palmer et al., 1991; Mandel et al., 1995).

Modelfree incorporates three models for rotational diffusion. Brent's method is used to optimize a single global τ_m for an isotropic overall diffusion model. Either Powell's method (as implemented by Brent) or a simulated annealing protocol (based on the downhill simplex method) is used to optimize a global rotational diffusion tensor for an axially symmetric diffusion model. A local rotational correlation time, τ_{mi}, for each spin can be optimized by non-linear least squares regression.

Other internal model free parameters (order parameters, internal correlation times and chemical exchange terms) are optimized by restrained non-linear least squares. Any parameter can be fixed at its input value rather than optimized and simple bounds can be placed on any parameter.

Monte Carlo simulations are used to estimate uncertainties in model-free parameters and to perform statistical model selection based on F-testing. A good introductory discussion of the use of Monte Carlo simulations in error analysis is given by Press et al. (Press et al., 1986).

Required Software

Executable versions of **Modelfree** are freely available from http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer.

Modelfree utilizes software routines that have been copyrighted by Numerical Recipes Software, Inc. Consequently, source code is available only to users holding valid Numerical Recipes Software licenses. To obtain an academic workstation license, send your name, address, email address, workstation hostname, workstation internet address, workstation brand and model number, and a check for $50.00 (U.S.) to

Numerical Recipes Software
P.O. Box 243
Cambridge, MA 02238

Be certain to state that you want the FORTRAN version of the software. Numerical Recipes will send you a license and license number and instructions for obtaining all of their software subroutines by anonymous ftp (all the routines needed for **Modelfree** are provided with the **Modelfree** distribution). Up-to-date licensing information can be found at http://www.nr.com.

After you have a Numerical Recipes license, source code for Modelfree is obtained by emailing the Numerical Recipes license number to agp6@columbia.edu.

A FORTRAN 77 compiler is required to compile the **Modelfree** source code. In addition, you will need the BLAS and LAPACK libraries installed on your computer system. These are normally available from the vendor of your workstation. If not, you can obtain them from http://www.netlib.org.

STAR_BASE and XMGR are useful for analyzing the output from **Modelfree**. The STAR_BASE software is available from

ftp.cs.uwa.edu.au/pub/star	(precompiled SGI binary)
ftp.crystal.uwa.edu/pub/star	(source code)

The STAR_BASE distributions might be missing the man pages. An ascii text version of the documentation is provided with the **Modelfree** distribution. XMGR is available as part of the ACE/GR package from

ftp.teleport.com/pub/users/pturner

http://plasma-gate.weizmann.ac.il/Xmgr

Version History

Modelfree version 2.0 Initial public release

Modelfree version 3.0 Release date: 15Jun94

This is a nearly complete rewrite of the Modelfree package.

Modelfree version 4.0 Release date: 1Apr98

The software has been extensively re-written to model overall rotational diffusion using an axially symmetric diffusion tensor. At the same time, the program was modified to improve flexibility in data fitting, statistical analysis, and output format.

1. A separate grid search program (mfgrid in previous versions) does not exist. The grid search functionality has been incorporated into modelfree itself.

2. An axially symmetric rotational diffusion tensor for the molecule of interest is determined by using a simulated annealing or a conjugate gradient algorithm, as desired.

3. Modelfree now allows two different motional models to be specified for each spin. The F-statistic is calculated from the fitted results, within a single run of the program.

4. The output file format has been modified so that the file conforms to the STAR specifications. Particular output information can be extracted from the main output file by using the STAR_BASE program, NAWK, PERL or other text processing utilities

Modelfree version 4.01 Release date: 15Jul98

1. Fixed errors in calculation of F-statistic simulated distribution.

2. Added output of SSE for model 2 in simulation output file (-e flag)

3. Trimmed simulated F-distribution to include only simulations in which an improved SSE is obtained for model 2. The number of simulations used in determining the F-distribution is output as _F-simulations in output file.

4. Fixed output file so an overflow is not obtained for _t-value if _Uncertainty is zero (that is, if zero uncertainty is input for a relaxation datum).

Modelfree version 4.10 Release date: 4Feb00

1. Added uncertainty to csa value for error analysis

2. Added the relaxation interference rate constant ratio eta_xy/eta_z as an optional relaxation parameter

3. Fixed errors in X2 and F simulated distributions.

Modelfree version 4.15 Release date: 10Apr02

1. Fixed errors for IRIX due to type checking

2. Fixed flush-to-zero flag for SGI R12000 chip

Program Invocation

Modelfree requires that a number of input files be provided and it writes a series of output files. The file names are invoked on the command line as illustrated below:

modelfree4 -i *mfin* -d *mfdata* -p *mfpar* -m *mfmodel* -s *pdb* -o *mfout* -e *extension*

The output files must not exist or an error message will be generated.

Input Files

The names of the input files are specified on the command line when starting **Modelfree**. The files are referred to in this manual by their default names:

mfin: The main control file for the program.

mfmodel: The file specifying the internal motional models to be used for each spin being analyzed.

mfdata: The file containing the spin relaxation data.

mfpar: The file containing molecular parameters.

pdb: A standard Protein Data Bank (pdb) file containing atomic coordinates for heavy atoms and hydrogens participating in any dipolar interactions for which relaxation data is to be analyzed.

In normal applications, only the *mfin* and *mfmodel* files are modified to generate different analyses. The *mfdata*, *mfpar* and *pdb* files are re-usable in different analyses. All files are free format files. Fields can be separated by any number of spaces or tab characters. Blank lines and lines with a '*#*' character in the first column are ignored.

MFIN file format

The mfin file consists of a series of command lines setting global options for the **Modelfree** program. Each line begins with a keyword and is followed by one or more options:

optimization	*chisq*					
seed	*seed*					
search	*search_option*					
diffusion	*diffusion_model*	*diffusion_search*				
algorithm	*algorithm*	*options*				
simulations	*sim_type*	*#sim*	*trim*			
selection	*F_option*					
sim_algorithm	*algorithm*	*options*				
fields	*#fields*	*field*$_1$ ... *field*$_M$				
tm	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
Dratio	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
Theta	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
Phi	*value*	*flag*	*bound*	*lower*	*upper*	*steps*

Legal values for each parameter are given in Table 1 and are describe more fully below.

Table 1. MFIN variables

Keyword	Parameter	Values
optimization	chisq	tval frac
seed	seed	integer constant
search	search_option	none grid
diffusion	diffusion_model	local isotropic axial
	diffusion_search	none grid
algorithm	algorithm	fix theta brent powell anneal nonlin
	options	*<vide infra>*
simulations	sim_type	none pred expr
	#sim	integer constant
	trim	real constant (0-1.0)
selection	F_option	none ftest
sim_algorithm	algorithm	same as for algorithm key
	options	same as for algorithm key
fields	#fields	integer constant
	$field_1 ... field_M$	^1H fields (MHz)
tm, Dratio, Theta, Phi	value	real constant
	flag	0 or 1
	bound	-1, 0, 1, 2
	lower	real constant
	upper	real constant
	steps	integer

chisq sets the type of weighting function applied to the relaxation parameters. If *chisq* = *tval*, then the weights are the reciprocal of the variances in the relaxation data. Thus, the total χ^2 variable is given by

$$\chi^2 = \sum_{i=1}^{N} SSE(i) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left\{\frac{\left(R_{1ij} - \hat{R}_{1ij}\right)^2}{\sigma_{R_1ij}^{\ 2}} + \frac{\left(R_{2ij} - \hat{R}_{2ij}\right)^2}{\sigma_{R_2ij}^{\ 2}} + \frac{\left(NOE_{ij} - \hat{NOE}_{ij}\right)^2}{\sigma_{NOEij}^{\ 2}}\right\} \qquad [1]$$

in which R_{1ij}, R_{2ij}, and NOE_{ij} are the relaxation parameters for the *i*th spin and *j*th static magnetic field; $\hat{R}_{1ij}$, $\hat{R}_{2ij}$, and $\hat{NOE}_{ij}$ are the corresponding fitted values; and σ_{R_1ij}, σ_{R_2ij}, and σ_{NOEij} are the experimental uncertainties in the relaxation parameters. The total number of spins to be analyzed is *N* and the total number of static magnetic fields for which data is available in *M*. *SSE*(*i*) is the sum-squared-error residual for the *i*th spin. If *chisq* = *frac*, the weights are the reciprocal of the squares of the relaxation parameters. Thus, the total χ^2 variable is given by

$$\chi^2 = \sum_{i=1}^{N} SSE(i) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left\{\frac{\left(R_{1ij} - \hat{R}_{1ij}\right)^2}{R_{1ij}^{\ 2}} + \frac{\left(R_{2ij} - \hat{R}_{2ij}\right)^2}{R_{2ij}^{\ 2}} + \frac{\left(NOE_{ij} - \hat{NOE}_{ij}\right)^2}{NOE_{ij}^{\ 2}}\right\} \qquad [2]$$

In most cases, the experimental uncertainties in the relaxation parameters should be measured and the model-free analysis should use *chisq* = *tval*. The summations include only relaxation data for which the corresponding flag variable is set equal to 1 (*vide infra*).

seed is a random integer to serve as a seed for a random number generator. If the program has been compiled with the USE_GETSEED option enabled in the Makefile, then a value of *seed* = 0 will result in *seed* being set from the computer system clock; otherwise, *seed* = 0 will generate an error message. *seed* can be input as a positive or negative integer; however, the program will reset *seed* = –abs(*seed*) to ensure that *seed* is a negative integer in order to properly initialize the random number generator.

search_option determines whether an initial grid search of internal motional parameters is performed prior to any optimization of the overall diffusion model or least squares fitting of internal motional parameters. If *search_option* = *none*, no grid search is performed; if *search_option* = *grid*, then a grid search of internal parameters is performed using the lower and upper bounds for each motional parameters specified in the *mfmodel* file. Under normal circumstances, setting *search_option* = *grid* costs little computational time and is recommended.

diffusion_model defines the model to be used for overall rotational diffusion. If *diffusion_model* = *local*, then an independent local rotational correlation time, τ_{mi}, is used for each spin. If *diffusion_model* = *isotropic*, then a single global rotational correlation time, τ_m, is used for all spins. If *diffusion_model* = *axial*, then a global axially symmetric diffusion tensor is used for all spins.

diffusion_search controls whether an initial grid search should be performed for the initial values of τ_m or the diffusion tensor for isotropic or axially symmetric diffusion models, respectively. A grid search is performed if *diffusion_search* = *grid*. This field is not read if *diffusion_model* = *local*, in which case, grid searching of the local rotational correlation times is controlled by *search_option*.

algorithm controls the type of optimization to be performed on the global diffusion model. If *diffusion_model = local*, then *algorithm = nonlin* and non-linear least squares is used to optimize an independent local rotational correlation time, τ_{mi}, for each spin simultaneously with internal motional parameters. If *algorithm = fix*, the diffusion model is not optimized, but internal parameters are optimized for a diffusion model given by the input values for *tm* if *diffusion_model = isotropic* and for *tm*, *Dratio*, *Theta*, and *Phi* if *diffusion_model = axial*. If *diffusion_model = isotropic* and *algorithm = brent*, then *tm* will be optimized using Brent's univariate method. If *diffusion_model = axial* and *algorithm = powell*, then the axially symmetric diffusion model will be optimized using Brent's implementation of Powell's method for multidimensional minimization. If *diffusion_model = axial* and *algorithm = anneal*, then the axially symmetric diffusion model will be optimized using a simulated annealing method for multidimensional minimization. If *diffusion_model = axial* and *algorithm = theta*, then the diffusion model will be fixed and non-linear least squares will be used to optimize the orientation of each spin in the principal axis frame of the diffusion tensor by non-linear least squares along with other internal motional parameters.

sim_algorithm controls the type of optimization to be performed on the global diffusion model during Monte Carlo simulations. This is normally set either equal to the same value as *algorithm* or to *fix*. Different options can be specified for *algorithm* and *sim_algorithm* (*vide infra*).

sim_type determines the type of Monte Carlo simulations to be performed in order to generate fitting statistics. If *sim_type = none*, then no simulations are performed and the remainder of the line is ignored. If *sim_type = pred*, then simulated data sets are obtained by adding a random noise term to the best-fit relaxation parameters, $\hat{R}_{1ij}$, $\hat{R}_{2ij}$, and $\hat{NOE}_{ij}$. If *sim_type = expr*, then the simulated data sets are obtained by adding a random noise term to the experimental relaxation parameters, R_{1ij}, R_{2ij}, and NOE_{ij}. In either event, the noise terms for the three relaxation parameters are obtained by drawing random numbers from Gaussian distributions with mean 0 and standard deviations given by the experimental uncertainties, σ_{R_1ij}, σ_{R_2ij}, and σ_{NOEij}, respectively. In most cases, *sim_type = pred* is appropriate.

#sim is the number of simulated date sets to analyze. The maximum value allowed is 1000. Typically, values in the range 300 to 500 are satisfactory. This field is ignored if *sim_type = none*.

trim controls whether the upper and lower tails of the distribution of simulated results should be excluded when calculating final statistics. If trim is greater than 0, then the largest *#sim* × *trim* and smallest *#sim* × *trim* values are excluded. While *trim* = 0 normally should be used, a value of *trim* in the range 0.05 to 0.10 sometimes is helpful in excluding simulated data for which the fitting did not converge. The simulated data is written to output files if the -e flag is set (*vide infra*), consequently, the simulated data can be inspected and trimmed after the analysis if desired.

F_option controls whether F-statistics comparing two models should be generated. If *F_option = ftest*, then the *mfmodel* file should contain two motional models for each spin. If the distribution of the F-statistic is to be simulated, then *sim_type* should not be set to *none* and *#sim* should be greater than zero.

#fields is the number of static magnetic fields for which data is available (between 1 and 5). Complete data sets are not required at each field; which data exists at each field is set in the *mfdata* file (*vide infra*). The field strengths (1 to *#fields*) are given by the ^{1}H Larmor frequencies in MHz.

The diffusion model is defined by a set of six values for each parameter in the diffusion model. The possible parameters are *tm*, *Dratio*, *Theta*, and *Phi*. *tm* is the isotropic rotational correlation time or $1/(6D_{iso})$, in which $D_{iso} = \langle \mathbf{D} \rangle /3$ and $\langle \mathbf{D} \rangle$ is the is the trace of the diffusion

tensor. Dratio is the ratio $D_{\parallel}/D_{\perp}$ for the diffusion tensor. *Theta* and *Phi* are the polar angles for the symmetry axis of the diffusion tensor in the coordinate frame of the PDB file. None of the lines are read if *diffusion_model = local*. Only *tm* is read if *diffusion_model = isotropic*. Only *tm* and *Dratio* are read if *diffusion_model = theta*. For each parameter in the diffusion model, the following fields must be set:

value provides an initial estimate of the parameter.

flag = 0 fixes the parameter at its input value, while *flag* = 1 enables optimization of the parameter (assuming *algorithm* is not *fix*). If *algorithm = brent*, then *flag* must equal 1 for the parameter tm. If *algorithm = powell*, then *flag* must equal 1 for at least two of the four parameters for the axially symmetric diffusion tensor. If *algorithm = anneal*, then *flag* must equal 1 for at least one of the four parameters for the axially symmetric diffusion tensor.

bound determines whether the parameter should be restricted to a given range during optimization (0 = no bound, -1 = lower bound, 1 = upper bound, 2 = lower and upper bounds). At present, the optimization algorithms do not implement this option.

lower is the value of the lower bound on the parameter for either grid searches or optimizations (bounds on optimizations are not implemented as described for *bound*).

upper is the value of the upper bound on the parameter for either grid searches or optimizations (bounds on optimizations are not implemented as described for *bound*).

steps is the number of grid search steps to perform between *lower* and *upper* if *diffusion_search = grid*.

MFMODEL file format

The *mfmodel* file contains one record for each spin to be analyzed in the current run (which might be fewer spins than exist in the *mfpar* and *mfdata* files). Each record consists of an initial identifier followed by up to 12 lines defining the internal motional models to be applied to the spin:

spin	*title*						
M1	tloc	*tloc*	*flag*	*bound*	*lower*	*upper*	*steps*
M1	Theta	*theta*	*flag*	*bound*	*lower*	*upper*	*steps*
M1	Sf2	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M1	Ss2	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M1	te	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M1	Rex	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M2	tloc	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M2	Theta	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M2	Sf2	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M2	Ss2	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M2	te	*value*	*flag*	*bound*	*lower*	*upper*	*steps*
M2	Rex	*value*	*flag*	*bound*	*lower*	*upper*	*steps*

title is a 10-character identifier for the spin. The string should not contain any embedded blank characters.

Each line defining an internal motional parameter has eight fields given as follows:

field 1: model number
field 2: parameter name
field 3: value
field 4: flag
field 5: bound
field 6: lower
field 7: upper
field 8: steps

The model number has values M1 or M2. M1 defines the primary model and is required. M2 is necessary only if a secondary model is specified for model selection, in which case *selection = ftest* should be set in the *mfin* file. If *selection = none*, then any lines beginning with M2 are ignored.

The parameter name field is self-explanatory. The line defining the *tloc* parameter is only necessary if *diffusion_model = local* and is ignored otherwise. The line defining Theta is necessary only if *algorithm = theta* and is ignored otherwise.

value provides an initial estimate of the parameter.

flag = 0 fixes the parameter at its input value, while *flag* = 1 enables optimization of the parameter.

bound determines whether the parameter should be restricted to a given range during optimization (0 = no bound, -1 = lower bound, 1 = upper bound, 2 = lower and upper bounds).

lower is the value of the lower bound on the parameter for either grid searches or optimizations.

upper is the value of the upper bound on the parameter for either grid searches or optimizations.

steps is the number of grid search steps to perform between *lower* and *upper*.

MFPAR file format

The *mfpar* file contains two lines for each spin for which data exists. Which of the spins are analyzed in any single run is set by the entries in the *mfmodel* file. Each line consists of a keyword followed by one or more parameters. The required lines are:

spin	*title*				
constants	*residue*	*nucleus*	*gamma*	*rxh*	*csa*
vector	*atom1*	*atom2*			

title must be identical to the character string used in the mfdata and mfmodel files.

residue is the sequence number of the residue containing the spin. The residue number must match the residue numbering in the PDB file, if a PDB file is being utilized.

nucleus is a three character identifier for the nuclear spin. Typical entries are 15N and 13C.

gamma is the gyromagnetic ratio of the spin in units of $T^{-1}s^{-1}/10^7$. *gamma* is a signed quantity; thus, the appropriate entry for a ^{15}N spin is –2.71.

rxh is the bond length for the dipole-dipole interaction. Usually, this is the bond length of the X-H bond. *rxh* is given in units of Angstroms.

csa is the chemical shift anisotropy of the spin, measured in parts per million.

atom1 and *atom2* are the atom symbols in the PDB file that define the principle direction of the relaxation interactions. Normally, these two parameters designate the atom types for the X and H spins. This line is read only if *diffusion_model = axial*.

MFDATA file format

The *mfdata* file contains $3M+1$ lines for each spin for which relaxation data exists, in which M is the number of static magnetic fields utilized (given by *#fields*). Which of the spins are analyzed in any single run is set by the entries in the *mfmodel* file. Each line consists of a keyword followed by one or more parameters. The required lines are:

spin	*title*			
R1	$field_1$	R_1	σ_{R_1}	*flag*
R2	$field_1$	R_2	σ_{R_2}	*flag*
NOE	$field_1$	*NOE*	σ_{NOE}	*flag*
		•		
		•		
		•		
R1	$field_M$	R_1	σ_{R_1}	*flag*
R2	$field_M$	R_2	σ_{R_2}	*flag*
NOE	$field_M$	*NOE*	σ_{NOE}	*flag*

title must be identical to the character string used in the mfpar and mfmodel files.

$field_1$ ... $field_M$ is the ^1H Larmor frequency in MHz. The order of listing of the static magnetic fields must match the order of entry in the *mfin* file.

R_1, R_2, and *NOE* are the experimental values of the relaxation rates. R_1 and R_2 have units of s^{-1}; *NOE* is dimensionless and is given by $1 + \eta$, in which η is the nuclear Overhauser effect enhancement.

σ_{R_1}, σ_{R_2}, and σ_{NOE} are the experimental uncertainties in the relaxation rates. σ_{R_1} and σ_{R_2} have units of s^{-1}; σ_{NOE} is dimensionless.

flag determines whether a given datum will be included (*flag* = 1) or excluded (*flag* = 0) from the analysis. This provides a method of excluding missing or unreliable data from the analysis.

Optimization protocols

Certain of the optimization algorithms have user-settable options. These are described below. The algorithms *fix*, *nonlin*, and *theta* do not have any settable options.

algorithm	*brent*	*grid_search*			
algorithm	*powell*	*grid_search*	*restarts*		
algorithm	*anneal*	*grid_search*	*#temps*	*#steps*	*scale_factor*
sim_algorithm	*brent*	*grid_search*			
sim_algorithm	*powell*	*grid_search*	*restarts*		
sim_algorithm	*anneal*	*grid_search*	*#temps*	*#steps*	*scale_factor*

grid_search determines whether an initial grid search for internal parameters should be performed at each step of the optimization of the global diffusion model. Setting *grid_search* = *grid* results in better optimizations because some local minima are avoided, but increases the computation time. *grid_search* = *none* means that no grid searching will be performed (other than possibly an initial grid search set by *search_option*).

restarts can be equal to 1 or 2. If *restarts* = 2, then after the initial convergence of the routine, the algorithm will be re-initialized and restarted a second time. In some cases, this is valuable in checking for false minima. *restarts* should normally be set to 1 if performing Powell minimization for *sim_algorithm*.

#temps is the number of temperature steps to perform in simulated annealing. *#temps* = 100 is a reasonable value for initial optimization.

#steps is the number of cycles of simplex minimization to perform at each temperature. *#steps* = 50 is a reasonable value for initial optimization.

scale_factor is the factor by which the temperature should be reduced at each step of the protocol. A value of *scale_factor* = 0.9 is reasonable.

Output Files

Modelfree produces a number of output files. The files are referred to in this manual by their default names:

mfout: The main output file for the program.

pdb.rotate: A standard Protein Data Bank (pdb) file containing atomic coordinates rotated to the principal axis frame of the diffusion tensor. This file is only produced if *diffusion_model* = *axial*. The name is derived by appending the extension ".rotate" to the input *pdb* file name.

title.extension: If simulations are performed and the -e flag is set on the input line, an output file will be produced for each spin. The file name is derived by catenating *title* and *extension*.

MFOUT file format

mfout is a STAR (Self-defining Text Archive and Retrieval) (Hall, 1991; Hall & Spadaccini, 1994) compliant file containing both the input and output data. All input files can be regenerated if desired from the *mfout* file. In contrast to other implementations of the STAR format (such as the BioMagResBank file formats), *mfout* uses the nested loop data structure, but does not use the save frame data structure. The format of the output file is

```
# Modelfree STAR Format Output File

data_header
    _modelfree_version
    _date

    _Input_file
    _Model_file
    _Data_file
    _Parameter_file
    _PDB_file
    _PDB_rotate_file
    _Simulation_file

    _optimization
    _seed
    _search
    _diffusion
    _algorithm
    loop_
        _algorithm_option
    _simulations
    _iterations
    _trim_level
    _selection
    _sim_algorithm
    loop_
        _sim_algorithm_option
    _total_spins
    _number_of_fields
    loop_
        _1H_fields

data_title
loop_
        _Title  _Residue

data_chi_square
    _Total_X2
    loop_
        _Percentile  _simulated_X2

data_diffusion_tensor
loop_
    _Diffusion_name  _Units  _Fit_value  _Fit_error  _Flag  _Sim_value  _Sim_error  _Sim_abs  _Geary-Z

data_diffusion_correlation_matrix
loop_
    _Diffusion_name_1  _Diffusion_name_2  _Covariance
```

```
data_spin_parameters
loop_
    _Residue  _Model  _Nucleus  _Gamma  _Rxh  _CSA  _Atom_1  _Atom_2

data_relaxation
loop_
    _relaxation_rate_name  _relaxation_rate_unit  _field
    loop_
        _Residue  _Value  _Uncertainty  _Flag  _Fit_value  _t-value
    stop_

data_model_1
loop_
    _Model_free_name  _Model_free_unit
    loop_
        _Residue  _Fit_value  _Fit_error  _Flag  _Sim_value  _Sim_error  _Sim_abs  _Geary-Z

data_sse
loop_
        _Residue  _SSE
        loop_
            _Percentile  _simulated_SSE
        stop_

data_correlation_matrix
loop_
    _Residue
    loop_
       _Model_free_name_1  _Model_free_name_2  _Covariance
    stop_

data_model_2
loop_
    _Model_free_name  _Model_free_unit
    loop_
        _Residue  _Fit_value  _Fit_error  _Flag
    stop_

data_F_dist
loop_
    _Residue  _F-stat  _F-simulation
    loop_
        _Percentile  _simulated_F_dist
    stop_
```

In general, each data block (defined by the data_*name* keyword) contains related data. Thus, data_model_1 contains the model free results for the primary model fit to the relaxation data contained in the data block data_relaxation. Each data name (defined by the _*name* keyword) that is not part of a loop structure is followed by a data value. Thus, the data name _Input_file would be followed in the mfout file by the name of the MFIN file:

_Input_file *mfin*

The loop structure is opened by the keyword “loop_” and terminated by the keyword “stop_”, except that the outermost loop is not terminated by a “stop_” keyword. The loop_ command is followed by a list of data names and a list of data packets. Each data packet contains a list of data

values in a one-to-one mapping to the list of data names. Thus, the following loop structure defines the ^{1}H fields at which data was collected:

```
loop_
   _1H_fields
   500.13
   600.13
   750.13
```

The mfout file is an ASCII file and can be processed by using a text editor, AWK, PERL, or other text processing utilities. The mfout file also can be processed using the Star_Base software package (Spadaccini & Hall, 1994). Star_Base defines a query language that is used to generate requests. Output from Star_Base can be further processed by other text processing utilities. A number of simple Unix shell scripts that utilize the Star_Base program are provided with the **Modelfree** package.

Certain of the output entries are self-explanatory. Others are described below.

_Fit_value is the value of a parameter obtained by optimization of the input value if *flag* = 1, or the input value if *flag* = 0

_Fit_error is the error obtained from the covariance matrix for the least squares optimization of a parameter. This parameter is zero if flag = 0. This parameter is obtained only for internal motional parameters.

_Sim_value is the mean value of a parameter obtained from Monte Carlo simulations.

_Sim_error is the standard deviation of the parameters obtained from the Monte Carlo simulations

_Sim_abs is the mean absolute deviation of a parameter obtained from the Monte Carlo simulations.

_Geary-Z is the Geary statistic is given by $(1.2533\ U-1)(\#sim)^{1/2}/0.2661$, in which U is the ratio of the mean absolute deviation and standard deviation (Devore, 1982). The Geary Z value is a measure of the normality of the distribution of the simulated results. Positive values indicate that the distribution is narrower than a normal distribution; negative values indicate that the distribution is wider (heavier tails) than the normal distribution. Values outside of the range -2 to 2 are statistically significant; however, for 300-500 simulations, values in the range -10 to 10 are still quite satisfactory (giving absolute and standard deviations that differ by <20%).

_Covariance is the Pearson's r correlation coefficient between two parameters calculated from the Monte Carlo simulations.

_t-value is given by $\left(R - \hat{R}\right)/\sigma_R$ if *chisq* = *tval*, or $\left(R - \hat{R}\right)/R$ if *chisq* = *frac*, in which R is a relaxation parameter (R_1, R_2, *NOE*), $\hat{R}$ is the fitted value of R, and σ_R is the experimental uncertainty in R. If *chisq* = *tval*, but σ_R =0, then the value of *_t-value* is not meaningful.

_F-stat is the F-statistic comparing the improvement in fit afforded by model 2 (M2) compared with model 1 (M1) as input in the *mfmodel* file.

_F-simulation is the number of simulated points (less than or equal to *#sim*) used in generating the *F*-distribution.

Simulated distributions for the total χ^2, *SSE*(*i*) for individual spins, and the F-statistic for individual spins are reported as a two column table with entries

Percentile *Simulated_value*

in which the *Simulated_value* is greater than *Percentile* × 100% of the *#sim* values obtained from the Monte Carlo simulations.

Simulation file TITLE.EXTENSION format

The simulation output file is produced for each spin and has the name *title.extension*. The file contains an initial comment line followed by *#sim* lines of simulated data. The comment line contains the column headings. Each line of data has the format:

iteration *SSE* *parameter_1* **...** *parameter_n* [*SSE2*]

in which *iteration* is an index from 1 to *#sim*, *SSE* is the sum-of-squared-residuals for the spin designated by *title* for the given *iteration*, and *parameter_1* through *parameter_n* are the values of the optimized local parameters for the first model in the MFMODEL file. The *SSE2* field is the value of *SSE* for the second model in the MFMODEL file. This field is output only if *selection* = *ftest*. A similar file named *diffusion.extension* is produced that contains the simulation results for fitting the diffusion tensor. These files can be used to recalculate statistics from the simulation (changing the degree of trimming, for example) or for examining the covariance of the optimized parameters by constructing scatter plots.

Theoretical Background

Relaxation of protonated heteronuclei is dominated by the dipolar interaction with the directly attached ^{1}H spin and by the chemical shift anisotropy mechanism. Relaxation parameters are given by (Abragam, 1961):

$$R_1 = (d^2/4)\,[J(\omega_H-\omega_X) + 3J(\omega_X) + 6J(\omega_H+\omega_X)] + c^2J(\omega_X) \qquad [3]$$

$$R_2 = (d^2/8)\,[4J(0) + J(\omega_H-\omega_X) + 3J(\omega_X) + 6J(w_H) + 6J(\omega_H+\omega_X)]$$

$$+ (c^2/6)\,[4J(0) + 3J(\omega_X)] + R_{ex} \qquad [4]$$

$$NOE = 1 + (d^2/4R_1)\,(\gamma_X/\gamma_H)\,[6J(\omega_H+\omega_X)-J(\omega_H-\omega_X)] \qquad [5]$$

in which $d = \mu_0 h\gamma_X\gamma_H\left\langle r_{XH}^{-3}\right\rangle/(8\pi^2)$, $c = \omega_X\Delta\sigma/\sqrt{3}$, μ_0 is the permeability of free space; h is Planck's constant; γ_H and γ_X are the gyromagnetic ratios of ^{1}H and the X spin (X=^{13}C or ^{15}N), respectively; r_{XH} is the X-H bond length; ω_H and ω_X are the Larmor frequencies of ^{1}H and X spins, respectively; and $\Delta\sigma$ = is the chemical shift anisotropy of the X spin (assuming an axially

symmetric chemical shift tensor). The symmetry axis of the chemical shift tensor is assumed to be collinear with the X-H bond vector.

The model-free formalism, as described by Lipari and Szabo (Lipari & Szabo, 1982; Lipari & Szabo, 1982), and extended by Clore and co-workers (Clore et al., 1990), determines the amplitudes and time scales of the intramolecular motions by modeling the spectral density function, $J(\omega)$, as

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{\left(S_f^2 - S^2\right)\tau}{1+(\omega\tau)^2}\right] = \frac{2}{5}S_f^2\left[\frac{S_s^2\tau_m}{1+(\omega\tau_m)^2} + \frac{\left(1 - S_s^2\right)\tau}{1+(\omega\tau)^2}\right] \quad [6]$$

in which $\tau = \tau_s\tau_m / (\tau_s + \tau_m)$, τ_m is the isotropic rotational correlation time of the molecule, τ_s is the effective correlation time for internal motions, $S^2 = S_f^2 S_s^2$ is the square of the generalized order parameter characterizing the amplitude of the internal motions, and S_f^2 and S_s^2 are the squares of the order parameters for the internal motions on the fast and slow time scales, respectively. Generalized order parameters represent motions that are described by dynamics on the ns-ps time scale, with values ranging from zero for isotropic internal motions to unity for completely restricted motion in a molecular reference frame.

For an axially symmetric diffusion tensor (Woessner, 1962; Halle & Wennerström, 1981; Barbato et al., 1992),

$$J(\omega) = \frac{2}{5}S_f^2\sum_{j=1}^{3}A_j\left[\frac{S_s^2\tau_j}{1+(\omega\tau_j)^2} + \frac{\left(1 - S_s^2\right)\tau_j'}{1+(\omega\tau_j')^2}\right] \quad [7]$$

in which $\tau_j' = \tau_j\tau_s / (\tau_j + \tau_s)$, $\tau_1^{-1} = 6\,D_\perp$, $\tau_2^{-1} = 5\,D_\perp + D_\parallel$, $\tau_3^{-1} = 2\,D_\perp + 4D_\parallel$, $A_1 = (3\cos^2\theta - 1)^2/4$, $A_2 = 3\sin^2\theta\cos^2\theta$, $A_3 = (3/4)\sin^4\theta$, and θ is the angle between the X-H bond vector and the unique axis of the principal frame of the diffusion tensor. The functional form of Eq. [7] can be derived analytically for only a limited number of motional models (Schurr et al., 1994) and should be regarded as a heuristic approximation for other cases. In particular, no theoretical derivations of a spectral density function for internal motions on two time scales and axially symmetric overall diffusion have been reported in the literature.

The order parameter, S^2, is given by (Lipari & Szabo, 1982; Lipari & Szabo, 1982; Henry & Szabo, 1985; Brüschweiler & Wright, 1994):

$$S^2 = \sum_{m=-2}^{2}\left\langle Y_2^{m*}(\Omega)\right\rangle\left\langle Y_2^{m}(\Omega)\right\rangle \quad [8]$$

in which $Y_2^m(\Omega)$ are modified spherical harmonic functions (Brink & Satchler, 1993), $\Omega = (\theta, \phi)$ defines the orientation of the X-H vector in a molecular reference frame (principal axis system of the diffusion tensor), and angular brackets indicate ensemble averaging.

A phenomenological exchange term, R_{ex}, is included in equation [4] to account for chemical exchange processes that contribute to the decay of transverse magnetization during the CPMG pulse train or during the spin-locking period in the experiments used to measure R_2 (Bloom et al., 1965; Wennerström, 1972). The following expression approximates the effect of two site chemical exchange in a CPMG experiment (Luz & Meiboom, 1963):

$$R_{ex} = \frac{p_1 p_2 (\Delta\omega)^2}{k_{ex}} \left[1 - \frac{2}{k_{ex}\tau_{cp}} \tanh\left(\frac{k_{ex}\tau_{cp}}{2} \right) \right] \qquad [9]$$

in which τ_{cp} is the delay between 180° pulses in the CPMG sequence, $\Delta\omega$ is the difference in chemical shift of the nucleus in the two conformational states, $k_{ex} = k_{-1} / p_1 = k_1 / p_2$, $1 \geq p_1 \geq 0.5$, and $p_2 = (1-p_1)$ are the populations of the two conformational states, k_1 is the forward exchange rate constant and k_{-1} is the reverse exchange rate constant. This equation is in close agreement with exact formulations (Allerhand & Gutowsky, 1965); numerical calculations comparing Eq. [9] with the exact equation (Jen, 1978; Davis et al., 1994), indicate that for exchange parameters anticipated for ^{15}N nuclei in proteins ($\Delta\omega < 4$ ppm; $k_{ex}\tau_{cp} > 10^{-2}$), Eq. [9] is accurate to within 5%. The following expression gives the effect of two site chemical exchange in a $R_{1\rho}$ experiment (Deverell et al., 1970):

$$R_{ex} = p_1 p_2 (\Delta\omega)^2 \left[\frac{k_{ex}}{k_{ex}^2 + \omega_e^2} \right] \qquad [10]$$

in which $\omega_e = (\Omega^2 + \omega_1{}^2)^{1/2}$ is the effective field in the rotating frame, Ω is the chemical shift offset, and ω_1 is the spin-locking field strength. The **Modelfree** program reports the value of R_{ex} obtained for the first static magnetic field in the list of fields used. The value of R_{ex} is scaled quadratically for other fields within the program.

Using Modelfree

The analysis of relaxation data using **Modelfree** has at least three steps: initial estimation of the rotational correlation time or diffusion tensor, model selection, and final optimization. These steps may need to be iterated to obtain a convergent, self-consistent description of the data.

Estimation of rotational correlation time

If no three-dimensional structure of the molecule under investigation is available, or if the molecule is known to have a low degree of rotational anisotropy, then the overall rotational correlation time, τ_m, can be estimated from a trimmed mean value of R_2/R_1 by solving the equation (Kay et al., 1989):

$$\frac{R_2}{R_1} = \frac{4J(0) + J(\omega_X - \omega_H) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_X + \omega_H) + (c^2 / 3d^2)\{4J(0) + 3J(\omega_X)\}}{2J(\omega_X - \omega_H) + 6J(\omega_X) + 12J(\omega_X + \omega_H) + 2(c^2 / 3d^2)J(\omega_X)} \quad [11]$$

in which

$$J(\omega) = \frac{2}{5} S^2 \left[\frac{\tau_m}{1 + (\omega\tau_m)^2} \right] \quad [12]$$

is obtained from Eq. [6] assuming that internal motions are limited (large S^2) and fast (τ_e < 10 ps). Eq. [11] is independent of S^2 and depends only on τ_m. The program **tmest** solves this equation and is available at

http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer

Estimation of rotational diffusion tensor

If a three-dimensional structure of the molecule under investigation is available (determined either by x-ray crystallography or NMR spectroscopy), then the rotational diffusion tensor can be estimated from the R_2/R_1 ratio for a subset of spins in the molecule for which chemical exchange motions are absent and internal motion is restricted. This procedure has been described (Brüschweiler et al., 1995; Tjandra et al., 1995; Lee et al., 1996) and software for performing the analysis is available at

http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer

Model selection

The statistical approach to selection of model-free parameters has been outlined previously (Mandel et al., 1995). If the 3 relaxation parameters R_1, R_2 and *NOE* have been measured at a single static magnetic field strength, then no more than 3 model-free parameters (in addition to the overall rotational correlation time or diffusion tensor) can be fit to these data. Five possible sets of model-free parameters can be fit to 3 experimental data points:

Model 1: S2s

Model 2: S2s and te

Model 3: S2s and Rex

Model 4: S2s, te and Rex

Model 5 Sf2, Ss2, and te

For models 1-4, $S_f^2 = 1.0$ and $S^2 = S_s^2$. The **Modelfree** program always associates S_s^2 with τ_e; do not try to optimize S_f^2 and τ_e as a pair. Additional models may be testable with additional data acquired at multiple static magnetic field strengths.

Model selection is performed using *algorithm = fix* and *sim_algorithm = fix* in order to render the fitting of the relaxation data for individual spins independent of the data for other spins being analyzed. Thus, good initial values of the overall rotational correlation time or rotational diffusion tensor are important for model selection.

The fit of any single model to the relaxation data for a given spin is tested using the residual $SSE(i)$ value for that spin. If simulations were performed, then the distribution of simulated $SSE(i)$ values is also determined by the program. The $SSE(i)$ values are distributed approximately as a χ^2-statistic with $d = n - m$ degrees of freedom in which n is the number of relaxation parameters and m is the number of fitted model free parameters. The critical value for the statistical test can be obtained from the simulated distribution or the theoretical distribution of χ^2. The theoretical distribution of the χ^2-statistic is calculated assuming that the underlying data have a normal distribution, which is not *a priori* true for relaxation data; however, as noted in Mandel et al. (1995), the simulated distributions obtained from the Monte Carlo procedure agree well with theoretical results. At a confidence level α, the residual $SSE(i)$ is compared to the $(1-\alpha)$ 100% level of the distribution. If the $SSE(i)$ is less than the critical value, then the model adequately describes the data.

In most applications of **Modelfree**, the *F*-statistic is used to assess whether the improvement in fit obtained by using a more complicated model really is significant, or merely arises because of the random statistical reduction in $SSE(i)$ that follows upon incorporation of additional parameters. Both models to be compared are specified in the *mfmodel* file. The two models must be nested; that is, the parameters of the first model (M1 in the *mfmodel* file) must be a subset of the parameters of the second model (M2 in the *mfmodel* file). The *F*-statistic is calculated by the **Modelfree** program, and the distribution of the F-statistic is calculated if simulations are performed. The *F*-statistic is defined as:

$$F = [d_2 / (d_1 - d_2)] [SSE_1(i) - SSE_2(i)] / SSE_2(i) \qquad [13]$$

where $SSE_1(i)$ and $SSE_2(i)$ are the SSEs for models M1 and M2 with d_1 and d_2 degrees of freedom ($d_1 > d_2$). The *F*-statistic calculated from the SSEs of the fits to the experimental data should be compared with the $(1-\alpha)100\%$ critical value of the *F*-statistic obtained from the simulated data or the theoretical distribution of $F_{(d1-d2),d2}$. The theoretical distribution of the *F*-statistic is calculated assuming that the underlying data have a normal distribution, which is not *a priori* true for relaxation data. In addition, if the parameters that are added to the more complicated model are bounded, then the simulated distribution will be distorted relative to the theoretical distribution. For example, if model 1 (S2s) and model 3 (S2s, Rex) are compared and Rex is unrestricted, then the simulated distribution will closely approximate an *F*-distribution; however, if Rex is restricted to be > 0, then the distribution will be altered. Also, in these cases, simulations in which the parameter boundaries are reached are not included in the simulated *F*-distribution. Therefore, the value of *F-simulation* may be less than *#sim*. An F-statistic greater than the critical value implies that model M2 provides a better description of the data than model M1.

The rest of the selection scheme should be obvious from the flowchart and the discussion under Materials & Methods of Mandel et al. (1995).

The *F*-distribution simulated by the above procedure approximates the type I error: the probability of accepting the more complicated model when the simpler model is in fact correct.

Modelfree also can be used to simulate the *F*-distribution for type II errors: the probability that the simpler model is accepted when the more complicated model is correct. The distribution of type II errors reflects the power of the statistical test. See any standard statistical text for additional discussions of type II errors and power. To simulate type II errors, the more complicated model is input as M1 in the MODELIN file and the simpler model is input as M2. The *F*-statistic is calculated as

$$F = [d_1 / (d_2-d_1)] [SSE_2(i) - SSE_1(i)] / SSE_1(i) \quad [14]$$

Optimization

After model selection is completed, the overall rotational diffusion model and the internal motional parameters for each spin are optimized simultaneously.

In most cases, Powell minimization should be adequate for optimizing the rotational diffusion tensor. This algorithm requires that at least two parameters are optimized; therefore, the simulated annealing protocol must be used if only one of the four diffusion tensor parameters is to be optimized.

Fitting the diffusion tensor is slow (hopefully, future versions will be faster); therefore, all model selection (with fixed diffusion tensor values) should be completed before performing runs with large numbers of simulations in which the diffusion tensor is being optimized along with internal motional parameters.

References

Abragam, A. (1961). Principles of Nuclear Magnetism. Oxford, Clarendon Press.

Allerhand, A. & Gutowsky, H. S. (1965) *J. Chem. Phys. 42*, 1587-1599.

Barbato, G.,Ikura, M.,Kay, L. E.,Pastor, R. W. & Bax, A. (1992) *Biochemistry 31*, 5269-5278.

Bloom, M.,Reeves, L. W. & Wells, E. J. (1965) *J. Chem. Phys. 42*, 1615-1624.

Brink, D. M. & Satchler, G. R. (1993). Angular Momentum. Oxford, Clarendon Press.

Brüschweiler, R.,Liao, X. & P.E., W. (1995) *Science 268*, 886-889.

Brüschweiler, R. & Wright, P. E. (1994) *J. Am. Chem. Soc. 116*, 8426-8427.

Clore, G. M.,Szabo, A.,Bax, A.,Kay, L. E.,Driscoll, P. C. & Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989-4991.

Davis, D. G.,Perlman, M. E. & London, R. E. (1994) *J. Magn. Reson., Ser B 104*, 266-275.

Deverell, C.,Morgan, R. E. & Strange, J. H. (1970) *Mol. Phys. 18*, 553-559.

Devore, J. (1982). Probability and Statistics for Engineering and the Sciences. Monterey, Brooks/Cole Publishing Company.

Hall, S. R. (1991) *J. Chem. Inf. Comput. Sci. 31*, 326-333.

Hall, S. R. & Spadaccini, N. (1994) *J. Chem. Inf. Comput. Sci. 34*, 505-508.

Halle, B. & Wennerström, H. (1981) *J. Chem. Phys. 75*, 1928-1943.

Henry, E. R. & Szabo, A. (1985) *J. Chem. Phys. 82*, 4753-4761.

Jen, J. (1978) *J. Magn. Reson. 30*, 111-128.

Kay, L. E.,Torchia, D. A. & Bax, A. (1989) *Biochemistry 28*, 8972-8979.

Lee, L. K.,Rance, M.,Chazin, W. J. & Palmer, A. G. (1996) *Submitted*

Lipari, G. & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546-4559.

Lipari, G. & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559-4570.

Luz, Z. & Meiboom, S. (1963) *J. Chem. Phys. 39*, 366-370.

Mandel, A. M.,Akke, M. & Palmer, A. G. (1995) *J. Mol. Bio 246*, 144-163.

Palmer, A. G.,Rance, M. & Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371-4380.

Press, W. H.,Flannery, B. P.,Teukolsky, S. A. & Vetterling, W. T. (1986). Numerical Recipes. The Art of Scientific Computing. Cambridge, Cambridge University Press.

Schurr, J. M.,Babcock, H. P. & Fujimoto, B. S. (1994) *J. Magn. Reson., Ser. B 105*, 211-224.

Spadaccini, N. & Hall, S. R. (1994) *J. Chem. Inf. Comput. Sci. 34*, 509-516.

Tjandra, N.,Feller, S. E.,Pastor, R. W. & Bax, A. (1995) *J. Am. Chem. Soc. 117*, 12562-12566.

Wennerström, H. (1972) *Mol. Phys. 24*, 69-80.

Woessner, D. E. (1962) *J. Chem. Phys. 37*, 647-654.

Appendices

Appendices 1 through 6 give examples of the input and output files for a dataset consisting of 100 residues (the data is simulated with added noise; the actual values used to generate the simulated data are given as comments in the relevant files). The files are provided in the ./testing directory of the **Modelfree** distribution. Appendix 7 provides some examples of the use of the STAR_BASE program.

Appendix 1. Sample MFIN Input file

```
optimization tval

seed 0

search grid

diffusion axial none

algorithm powell grid 1

#algorithm anneal grid 100 50 0.9

simulations pred 200 0.00

selection none

sim_algorithm powell grid 1

fields 1   500.130
tm       8.5  1  2  8.0  10.0  10   #       9.200
Dratio   1.1  1  2  1.1   1.5   5   #       1.250
Theta   20.0  1  2  0.0  60.0  10   #      30.000
Phi     20.0  1  2 80.0 120.0  10   #     100.000
```

Appendix 2. Sample MFPAR Parameter File

```
spin    1
constants     1 N15     -2.710     1.020     -160.00
vector N HN

spin    2
constants     2 N15     -2.710     1.020     -160.00
vector N HN

spin    3
constants     3 N15     -2.710     1.020     -160.00
vector N HN
            •
            •
            •
spin   99
constants    99 N15     -2.710     1.020     -160.00
vector N HN

spin  100
constants   100 N15     -2.710     1.020     -160.00
vector N HN
```

Appendix 3. Sample MFDATA data file

```
spin    1
R1     500.130     1.724     0.030 1
R2     500.130    12.255     0.300 1
NOE    500.130     0.738     0.040 1

spin    2
R1     500.130     1.406     0.030 1
R2     500.130    10.142     0.300 1
NOE    500.130     0.791     0.040 1

spin    3
R1     500.130     1.566     0.030 1
R2     500.130    11.024     0.300 1
NOE    500.130     0.792     0.040 1
            •
            •
            •
spin   99
R1     500.130     1.556     0.030 1
R2     500.130    10.725     0.300 1
NOE    500.130     0.351     0.040 1

spin  100
R1     500.130     1.252     0.030 1
R2     500.130     8.461     0.300 1
NOE    500.130     0.762     0.040 1
```

Appendix 4. Sample MFMODEL model file

```
spin    1
M1 tloc  8.0   0   2        0.000      18.400 20  #       9.200
M1 Theta 0.0   0   2        0.000      90.000 20  #      55.304
M1 S2f   1.0   0   2        0.000       1.000 20  #       1.000
M1 S2s   1.0   1   2        0.000       1.000 20  #       0.929
M1 te    0.0   1   2        0.000     400.000 20  #      64.037
M1 Rex   0.0   0   2        0.000       0.000 20  #       0.000

spin    2
M1 tloc  8.0   0   2        0.000      18.400 20  #       9.200
M1 Theta 0.0   0   2        0.000      90.000 20  #      42.646
M1 S2f   1.0   0   2        0.000       1.000 20  #       1.000
M1 S2s   1.0   1   2        0.000       1.000 20  #       0.772
M1 te    0.0   0   2        0.000       0.000 20  #       0.000
M1 Rex   0.0   0   2        0.000       0.000 20  #       0.000

spin    3
M1 tloc  8.0   0   2        0.000      18.400 20  #       9.200
M1 Theta 0.0   0   2        0.000      90.000 20  #      47.657
M1 S2f   1.0   0   2        0.000       1.000 20  #       1.000
M1 S2s   1.0   1   2        0.000       1.000 20  #       0.839
M1 te    0.0   0   2        0.000       0.000 20  #       0.000
M1 Rex   0.0   0   2        0.000       0.000 20  #       0.000
           •
           •
           •
spin   99
M1 tloc  8.0   0   2        0.000      18.400 20  #       9.200
M1 Theta 0.0   0   2        0.000      90.000 20  #     162.450
M1 S2f   1.0   0   2        0.000       1.000 20  #       1.000
M1 S2s   1.0   1   2        0.000       1.000 20  #       0.794
M1 te    0.0   1   2        0.000     400.000 20  #     145.748
M1 Rex   0.0   0   2        0.000       0.000 20  #       0.000

spin  100
M1 tloc  8.0   0   2        0.000      18.400 20  #       9.200
M1 Theta 0.0   0   2        0.000      90.000 20  #      50.172
M1 S2f   1.0   0   2        0.000       1.000 20  #       1.000
M1 S2s   1.0   1   2        0.000       1.000 20  #       0.677
M1 te    0.0   0   2        0.000       0.000 20  #       0.000
M1 Rex   0.0   0   2        0.000       0.000 20  #       0.000
```

Appendix 5. Sample MFOUT output file

```
# Modelfree STAR Format Output File

data_header
     _modelfree_version 4.00
     _date 06-Mar-98

     _Input_file          mfinput
     _Model_file          mfmodel
     _Data_file           mfdata
     _Parameter_file      mfparam
     _PDB_file            mfpdb
     _PDB_rotate_file     mfpdb.rotate
     _Simulation_file     test

     _optimization                tval
     _seed                      -62290
     _search                      grid
     _diffusion                  axial
     _algorithm                 powell
     loop_
          _algorithm_option
               grid
               1
     _simulations                 pred
     _iterations                   200
     _trim_level                 0.000
     _selection                   none
     _sim_algorithm             powell
     loop_
          _sim_algorithm_option
               grid
               1
     _total_spins                  100
     _number_of_fields               1
     loop_
          _1H_fields
          500.130

data_title
loop_
          _Title  _Residue
     1                 1
     2                 2
     3                 3
           •
           •
           •
     99               99
     100             100

data_chi_square
           _Total_X2       129.5058
     loop_
               _Percentile  _simulated_X2
                    0.0500         1.4327
```

```
                    0.1000        11.7193
                    0.1500        96.1450
                    0.2000        99.5251
                    0.2500       103.0972
                    0.3000       106.0962
                    0.3500       108.3757
                    0.4000       111.7140
                    0.4500       113.5875
                    0.5000       116.2783
                    0.5500       119.5867
                    0.6000       122.0340
                    0.6500       124.3859
                    0.7000       126.0021
                    0.7500       128.7768
                    0.8000       131.6494
                    0.8500       134.4023
                    0.9000       137.4994
                    0.9500       146.0203
                    1.0000       168.9172

data_diffusion_tensor
loop_
     _Diffusion_name  _Units  _Fit_value  _Fit_error  _Flag  _Sim_value
            _Sim_error  _Sim_abs  _Geary-Z
     tm      (ns)        9.179  0.000  1   9.176  0.030  0.024   0.194
     Dratio  ()          1.249  0.000  1   1.254  0.020  0.016   0.189
     Theta   (degrees)  29.917  0.000  1  30.231  2.426  1.904  -0.874
     Phi     (degrees)  96.284  0.000  1  96.606  5.251  4.183  -0.092

data_diffusion_correlation_matrix
loop_
        _Diffusion_name_1   _Diffusion_name_2         _Covariance
          tm             Dratio            -0.3750
          tm             Theta              0.4467
          tm             Phi               -0.0181
          Dratio         Theta              0.0609
          Dratio         Phi                0.0652
          Theta          Phi               -0.0757

data_spin_parameters
loop_
  _Residue  _Model  _Nucleus  _Gamma  _Rxh  _CSA  _Atom_1  _Atom_2
  1        0000110   N15   -2.7100   1.0200  -160.0000   N   HN
  2        0000100   N15   -2.7100   1.0200  -160.0000   N   HN
  3        0000100   N15   -2.7100   1.0200  -160.0000   N   HN
                              •
                              •
                              •
  99       0000110   N15   -2.7100   1.0200  -160.0000   N   HN
  100      0000100   N15   -2.7100   1.0200  -160.0000   N   HN

data_relaxation
loop_
     _relaxation_rate_name _relaxation_rate_unit _field
     loop_
            _Residue  _Value  _Uncertainty  _Flag  _Fit_value  _t-value

     R1          (1/s)        500.130
```

```
                  1         1.724        0.030    1        1.759      -0.115E+01
                  2         1.406        0.030    1        1.417      -0.382E+00
                  3         1.566        0.030    1        1.576      -0.348E+00
                                    •
                                    •
                                    •
                 99         1.556        0.030    1        1.559      -0.884E-01
                100         1.252        0.030    1        1.248       0.118E+00
     stop_

     R2          (1/s)       500.130
                  1        12.255        0.300    1       11.736       0.173E+01
                  2        10.142        0.300    1        9.979       0.542E+00
                  3        11.024        0.300    1       10.873       0.504E+00
                                    •
                                    •
                                    •
                 99        10.725        0.300    1       10.684       0.137E+00
                100         8.461        0.300    1        8.513      -0.173E+00
     stop_

     NOE         ()          500.130
                  1         0.738        0.040    1        0.759      -0.528E+00
                  2         0.791        0.040    1        0.803      -0.289E+00
                  3         0.792        0.040    1        0.802      -0.251E+00
                                    •
                                    •
                                    •
                 99         0.351        0.040    1        0.353      -0.473E-01
                100         0.762        0.040    1        0.802      -0.994E+00
     stop_

data_model_1
loop_
_Model_free_name    _Model_free_unit
  loop_
  _Residue _Fit_value _Fit_error _Flag _Sim_value _Sim_error _Sim_abs _Geary-Z

Theta     (degrees)
  1    54.817  0.000  0   55.097  2.397  1.838  -119.009
  2    40.913  0.000  0   41.192  2.494  1.939   -78.225
  3    47.070  0.000  0   46.968  2.536  2.034    16.544
            •
            •
            •
  99  163.902  0.000  0  163.448  2.409  1.824  -155.522
  100  50.570  0.000  0   50.784  2.498  2.012    29.023
  stop_

S2        ()
  1     0.928  0.015  1    0.927  0.014  0.011    -1.268
  2     0.772  0.013  1    0.771  0.013  0.011    -1.051
  3     0.850  0.013  1    0.849  0.014  0.011    -1.781
            •
            •
            •
  99    0.785  0.013  1    0.785  0.013  0.010     0.003
  100   0.669  0.013  1    0.668  0.014  0.011     1.302
```

```
  stop_

S2f       ()
  1     1.000  0.000  0    0.000  0.000  0.000     0.000
  2     1.000  0.000  0    0.000  0.000  0.000     0.000
  3     1.000  0.000  0    0.000  0.000  0.000     0.000
            •
            •
            •
  99    1.000  0.000  0    0.000  0.000  0.000     0.000
  100   1.000  0.000  0    0.000  0.000  0.000     0.000
  stop_

S2s       ()
  1     0.928  0.015  1    0.927  0.014  0.011    -1.268
  2     0.772  0.013  1    0.771  0.013  0.011    -1.051
  3     0.850  0.013  1    0.849  0.014  0.011    -1.781
            •
            •
            •
  99    0.785  0.013  1    0.785  0.013  0.010     0.003
  100   0.669  0.013  1    0.668  0.014  0.011     1.302
  stop_

te        (ps)
  1    42.713  37.761  1  43.292  34.076  27.696    0.990
  2     0.000   0.000  0   0.000   0.000   0.000    0.000
  3     0.000   0.000  0   0.000   0.000   0.000    0.000
            •
            •
            •
  99  159.888  21.600  1 161.940  22.288  17.495   -0.861
  100   0.000   0.000  0   0.000   0.000   0.000    0.000
  stop_

Rex       (1/s)
  1     0.000   0.000  0   0.000   0.000    0.000   0.000
  2     0.000   0.000  0   0.000   0.000    0.000   0.000
  3     0.000   0.000  0   0.000   0.000    0.000   0.000
            •
            •
            •
  99    0.000   0.000  0   0.000   0.000    0.000   0.000
  100   0.000   0.000  0   0.000   0.000    0.000   0.000
  stop_

data_sse
loop_
            _Residue           _SSE
     loop_
              _Percentile _simulated_SSE
              1         4.6041
                   0.0500         0.0028
                   0.1000         0.0138
                   0.1500         0.0291
                   0.2000         0.0505
                   0.2500         0.0764
                   0.3000         0.1307
```

```
              0.3500          0.1781
              0.4000          0.2588
              0.4500          0.3770
              0.5000          0.4927
              0.5500          0.5674
              0.6000          0.7506
              0.6500          0.9060
              0.7000          1.1291
              0.7500          1.4703
              0.8000          1.7706
              0.8500          2.2444
              0.9000          3.0337
              0.9500          3.8777
              1.0000         12.8112
stop_
         2          0.5230
              0.0500          0.0510
              0.1000          0.1131
              0.1500          0.2595
              0.2000          0.3430
              0.2500          0.4226
              0.3000          0.5374
              0.3500          0.6903
              0.4000          0.8316
              0.4500          0.9621
              0.5000          1.1821
              0.5500          1.4003
              0.6000          1.5922
              0.6500          1.7261
              0.7000          2.0797
              0.7500          2.3304
              0.8000          2.8025
              0.8500          3.3143
              0.9000          4.3741
              0.9500          5.6480
              1.0000         11.7535
stop_
         3          0.4378
              0.0500          0.0909
              0.1000          0.2141
              0.1500          0.2965
              0.2000          0.3773
              0.2500          0.4501
              0.3000          0.5318
              0.3500          0.6470
              0.4000          0.7435
              0.4500          1.0161
              0.5000          1.2137
              0.5500          1.4042
              0.6000          1.7121
              0.6500          1.9222
              0.7000          2.1720
              0.7500          2.4348
              0.8000          2.9344
              0.8500          3.5032
              0.9000          4.2873
              0.9500          5.3749
              1.0000         14.0286
```

```
     stop_
           •
           •
           •
            99         0.0289
                  0.0500          0.0008
                  0.1000          0.0102
                  0.1500          0.0329
                  0.2000          0.0522
                  0.2500          0.0810
                  0.3000          0.1458
                  0.3500          0.2212
                  0.4000          0.2984
                  0.4500          0.3520
                  0.5000          0.4340
                  0.5500          0.5505
                  0.6000          0.6805
                  0.6500          0.8878
                  0.7000          1.0687
                  0.7500          1.3906
                  0.8000          1.6506
                  0.8500          1.8955
                  0.9000          2.3423
                  0.9500          3.0790
                  1.0000          5.6043
     stop_
           100         1.0318
                  0.0500          0.0998
                  0.1000          0.1856
                  0.1500          0.2827
                  0.2000          0.3481
                  0.2500          0.4612
                  0.3000          0.6590
                  0.3500          0.7907
                  0.4000          1.0503
                  0.4500          1.2715
                  0.5000          1.4094
                  0.5500          1.7375
                  0.6000          1.9535
                  0.6500          2.3303
                  0.7000          2.5579
                  0.7500          2.7618
                  0.8000          3.1572
                  0.8500          3.4517
                  0.9000          4.0447
                  0.9500          5.7424
                  1.0000         10.9812
     stop_

data_correlation_matrix
loop_
     _Residue
     loop_
           _Model_free_name_1   _Model_free_name_2          _Covariance

        1
              Theta           Theta                1.0000
              Theta           S2s                  0.0000
```

```
          Theta          te                0.0000
          S2s            S2s               1.0000
          S2s            te               -0.1635
          te             te                1.0000
stop_

    2
          Theta          Theta             1.0000
          Theta          S2s               0.0000
          S2s            S2s               1.0000
stop_

    3
          Theta          Theta             1.0000
          Theta          S2s               0.0000
          S2s            S2s               1.0000
stop_
      •
      •
      •

   99
          Theta          Theta             1.0000
          Theta          S2s               0.0000
          Theta          te                0.0000
          S2s            S2s               1.0000
          S2s            te                0.4299
          te             te                1.0000
stop_

  100
          Theta          Theta             1.0000
          Theta          S2s               0.0000
          S2s            S2s               1.0000
stop_
```

Appendix 6. Sample TITLE.EXTENSION simulation output file

The simulation output file ./testing/15.test contains the following information:

```
# index           SSE             S2s              te
    1          0.1351           0.732         198.473
    2          2.8248           0.697         178.577
    3          0.5068           0.691         152.949
            •
            •
            •
  199          4.0432           0.706         202.441
  200          5.7484           0.748         178.375
```

This file can be examined to assess in detail the outcome of the Monte Carlo simulations. For example, the file can be processed with NAWK, PERL or other text processing utilities to generate input files for XMGR that produce plots of the distribution of a single parameter or scatter plots showing the correlation between pairs of paramters. For example, the following script, provided as ./star/scatter.xmgr extracts pairs of columns from the files and produces output suitable for plotting scatter plots in XMGR:

```
#! /bin/sh

#usage: scatter.xmgr parameter_1 parameter_2 simulation_file

nawk 'BEGIN{
              print "@ S0 TYPE xy"
              print "@ S0  LINESTYLE 0"
              print "@ S0  SYMBOL 2"
              print "@ S0  SYMBOL SIZE 0.5"
              print "@ S0  SYMBOL FILL 1"
              print "@ XAXIS LABEL \""par1"\""
              print "@ YAXIS LABEL \""par2"\""
          }

          $1 == "#" {for (i=1;i<=NF;i++) {
                       if ($i == par1) field1 = i-1
                       if ($i == par2) field2 = i-1
                     }
                    }
          $1 !~ /#/ {print $field1,"\t",$field2}' par1=$1 par2=$2 $3
```

Executing the command line

`scatter.xmgr S2s te 15.test > scatter.xmgr`

using the ./testing/15.test file generates a file containing the following output:

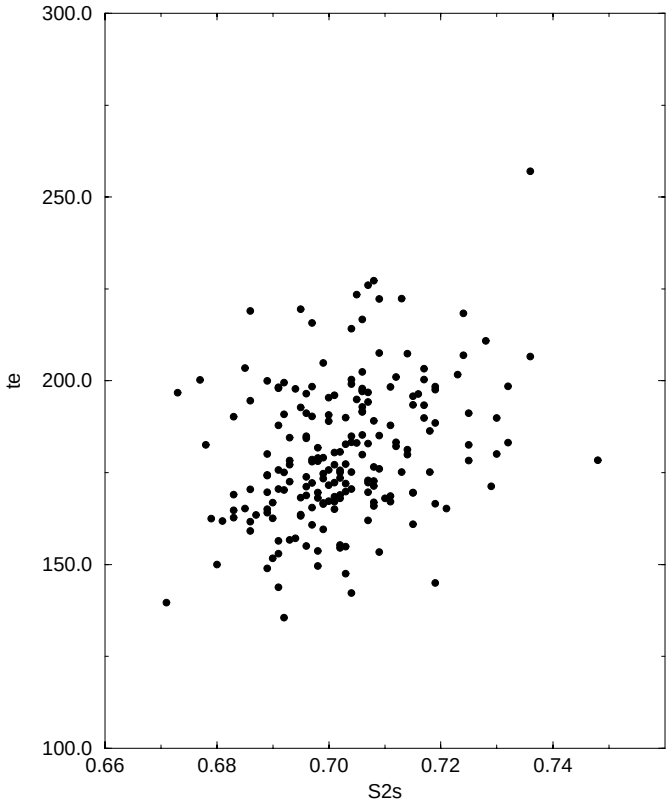
```
@ S0 TYPE xy
@ S0  LINESTYLE 0
@ S0  SYMBOL 2
@ S0  SYMBOL SIZE 0.5
@ S0  SYMBOL FILL 1
@ XAXIS LABEL "S2s"
@ YAXIS LABEL "te"
0.732        198.473
```

```
0.697        178.577
0.691        152.949
              •
              •
              •
0.706        202.441
0.748        178.375
```

The command

xmgr scatter.xmgr

generates the following graph:



Appendix 7. Sample STAR_BASE scripts

The following shell script, provided as ./star/get.mfpar in the **Modelfree** distribution, performs a simple extraction of the data for any of the model-free parameters from an output file:

```
#! /bin/sh

#usage: get.mfpar parameter mfout

sbdir=/usr/local/modelfree/STAR

$sbdir/sb -v none -r "
       if_ data_model_1
       scope_data_block_
            if_ _Model_free_name ~= $1
            scope_loop_packet_
              _Residue _Fit_value _Fit_error _Flag _Sim_value _Sim_error \
                   _Sim_abs _Geary-Z
            endscope_
            endif_
       endscope_
       endif_
" $2
```

Executing the command line

`get.mfpar S2 mfout`

using the sample output file in ./testing directory generates the following output:

```
data_model_1
loop_
    loop_
        _Residue
        _Fit_value
        _Fit_error
        _Flag
        _Sim_value
        _Sim_error
        _Sim_abs
        _Geary-Z
    stop_


        1 0.928 0.015 1 0.927 0.014 0.011 -1.268
        2 0.772 0.013 1 0.771 0.013 0.011 -1.051
        3 0.850 0.013 1 0.849 0.014 0.011 -1.781
              •
              •
              •
         99 0.785 0.013 1 0.785 0.013 0.010 0.003
        100 0.669 0.013 1 0.668 0.014 0.011 1.302
        stop_
```

The following shell script, provided as ./star/get.mfpar.xmgr in the **Modelfree** distribution, performs a simple extraction of the data for any of the model-free parameters from an output file and writes a file suitable for plotting with the XMGR program. This script uses NAWK to further process the output produced by STAR_BASE.

```
#! /bin/sh

#usage: get.mfpar.xmgr parameter mfout

sbdir=/user/local/modelfree/STAR

$sbdir/sb -v none -r "
       if_ data_model_1
       scope_data_block_
            if_ _Model_free_name ~= $1
            scope_loop_packet_
              _Residue _Fit_value _Fit_error _Flag _Sim_value _Sim_error \
                   _Sim_abs _Geary-Z
            endscope_
            endif_
       endscope_
       endif_
" $2 | nawk 'BEGIN{
              print "@ S0 TYPE xydy"
              print "@ S0  LINESTYLE 0"
              print "@ S0  SYMBOL 2"
              print "@ S0  SYMBOL SIZE 0.5"
              print "@ S0  SYMBOL FILL 1"
              print "@ XAXIS LABEL \"residue\""
              print "@ YAXIS LABEL \""mfpar"\""
          }

          $1 ~ /^[0-9]/ {print $1,"\t",$2,"\t",$6}' mfpar=$1 -
```

Executing the command line

`get.mfpar.xmgr S2 mfout > S2.xmgr`

using the sample output file in ./testing directory generates a file containing the following output:

```
@ S0 TYPE xydy
@ S0  LINESTYLE 0
@ S0  SYMBOL 2
@ S0  SYMBOL SIZE 0.5
@ S0  SYMBOL FILL 1
@ XAXIS LABEL "residue"
@ YAXIS LABEL "S2"
1        0.928   0.014
2        0.772   0.013
3        0.850   0.014
         •
         •
         •
99       0.785   0.013
100      0.669   0.014
```

The command

xmgr -type xydy S2.xmgr

generates the following graph: