Journal of Biomolecular NMR, **25:** 25–39, 2003.
KLUWER/ESCOM

The use of model selection in the model-free analysis of protein dynamics

Edward J. d'Auvergne* & Paul R. Gooley*
Department of Biochemistry and Molecular Biology, University of Melbourne, Melbourne, Victoria 3010, Australia

Received 3 July 2002; Accepted 16 October 2002

Key words: AIC, model-free analysis, model selection, NMR relaxation, protein dynamics

Abstract

Model-free analysis of NMR relaxation data, which is widely used for the study of protein dynamics, consists of the separation of the global rotational diffusion from internal motions relative to the diffusion frame and the description of these internal motions by amplitude and timescale. Five model-free models exist, each of which describes a different type of motion. Model-free analysis requires the selection of the model which best describes the dynamics of the NH bond. It will be demonstrated that the model selection technique currently used has two significant flaws, under-fitting, and not selecting a model when one ought to be selected. Under-fitting breaks the principle of parsimony causing bias in the final model-free results, visible as an overestimation of S^2 and an underestimation of τ_e and R_{ex}. As a consequence the protein falsely appears to be more rigid than it actually is. Model selection has been extensively developed in other fields. The techniques known as Akaike's Information Criteria (AIC), small sample size corrected AIC (AICc), Bayesian Information Criteria (BIC), bootstrap methods, and cross-validation will be compared to the currently used technique. To analyse the variety of techniques, synthetic noisy data covering all model-free motions was created. The data consists of two types of three-dimensional grid, the R_{ex} grids covering single motions with chemical exchange $\{S^2, \tau_e, R_{ex}\}$, and the Double Motion grids covering two internal motions $\{S_f^2, S_s^2, \tau_s\}$. The conclusion of the comparison is that for accurate model-free results, AIC model selection is essential. As the method neither under, nor over-fits, AIC is the best tool for applying Occam's razor and has the additional benefits of simplifying and speeding up model-free analysis.

Abbreviations: AIC – Akaike's Information Criteria; AICc – small sample size corrected AIC; BIC – Bayesian Information Criteria; CSA – chemical shift anisotropy; NOE – nuclear Overhauser effect; pdf – probability distribution function.

Introduction

The study of NMR relaxation data is the richest source of experimental information on protein dynamics and can reveal details on an atomic level. By analysis of the backbone amide nitrogen relaxation, a global picture of the dynamics of a protein can be revealed. Standard analysis consists of the measurement of three relaxation values per magnetic field strength, the $\{^1H\}$-^{15}N steady-state NOE, and the ^{15}N R_1 and R_2 relaxation rates. The significance of these rates is revealed through model-free theory. Two distinct components influence the relaxation values - the global rotational diffusion of the protein and internal motions of the NH bond vector relative to the rotational diffusion frame. Model-free theory separates these two components as well as describes the internal motion by amplitude and timescale. The original model-free theory (Lipari and Szabo, 1982a, b) describes a single internal motion using the two parameters S^2 and τ_e, where S^2 is the square of the generalised order parameter reflecting the amplitude of motion, and τ_e is the effective correlation time reflecting the picosecond to nanosecond timescale of the motion. S^2 will be referred to as the order parameter and τ_e as the correlation time. The theory was extended to include internal motions on

*To whom correspondence should be addressed. E-mails: ejdauv @pgrad.unimelb.edu.au; prg@unimelb.edu.au

two timescales with the faster of these described by the parameters S_f^2 and τ_f, and the slower by S_s^2 and τ_s (Clore et al., 1990). An additional term used in model-free analysis is R_{ex} which is included to account for the relaxation due to chemical exchange and is an indicator of motions on micro to millisecond timescales. Five model-free models describing different types of motion are constructed through various combinations of the model-free parameters. These are model 1 $\{S^2\}$, model 2 $\{S^2, \tau_e\}$, model 3 $\{S^2, R_{ex}\}$, model 4 $\{S^2, \tau_e, R_{ex}\}$, and model 5 $\{S_f^2, S_s^2, \tau_s\}$. Each type of relaxation value is described by different combinations of spectral density values at five frequencies (Abragham, 1961), and in turn, the model-free equations describe these spectral densities in terms of the model-free parameters (Lipari and Szabo, 1982a, b; Clore et al., 1990). Model-free analysis consists of fitting the relaxation data by χ^2 minimisation, using the model-free and relaxation equations. The fitting is repeated five times, once for each of the five model-free models.

The problem addressed in this paper is which of the five model-free models should be selected to describe the internal motion of the NH bond. The current model-free model selection technique (Mandel et al., 1995), which is almost universally used in model-free analysis, is based on hypothesis testing using chi-squared and F-tests. An older technique (Farrow et al., 1994), a set of rules based on the fitted values of the model-free parameters and their errors, will also be examined. The study of model selection is an important field in statistics which has been extensively developed for use in a broad variety of disciplines (Linhart and Zucchini, 1986; Burnham and Anderson, 1998; Zucchini, 2000). Three major categories of model selection include hypothesis testing, Bayesian methods, and frequentist methods. Traditional hypothesis testing can be classed into the step-up, step-down, or step-wise methods, with the current model-free technique (Mandel et al., 1995) classified as a highly modified step-up procedure. The use of hypothesis testing for model selection has many shortcomings (Burnham and Anderson, 1998), the major concern being significance levels or α-levels for the tests. As the variance of a model increases with the number of parameters, due to more noise being reflected in the model, the α-levels should be adjusted for both the sample size and parameter number yet no rules exist for the selection of correct α-levels. Due to the arbitrary choice of constant α-levels the current model selection is inconsistent, as different α-levels lead to different selection results. In addition, different models will be selected depending whether step-up, step-down, or step-wise testing is used. Importantly, hypothesis testing cannot be used for non-nested models. For model-free analysis this is testing between model-free models 2 and 3, 3 and 5, and 4 and 5. The adequacy of the current model-free model selection has previously been called into question but no solution has been put into practice (Korzhnev et al., 1997; Jin et al., 1998; Andrec et al., 1999). Bayesian statistical methods have been applied to the analysis of NMR relaxation data as a means of determining both the global rotational diffusion parameters (Andrec et al., 2000) and the model-free parameters (Andrec et al., 1999). As of yet, Bayesian statistics have not been implemented as a tool for model-free model selection although the methodology has been proposed (Jin et al., 1998).

Central to many of the frequentist techniques is the concept of parsimony, which states that the simplest model which fits well should be used to describe the data (Burnham and Anderson, 1998). The principle of parsimony is a manifestation of Occam's razor and is the balance between bias and variance. Bias is a distortion of the results due to oversimplification while variance is the incorporation of more experimental noise into the final model. Generally, as the number of parameters in the model increases, bias will decrease and variance will increase. Therefore, in the selection of a biased model, under-fitting results in an overestimation of precision due to the low variance. Over-fitting increases the variance in the final model-free results. The best balance is achieved by selection of the model with the lowest value of a quantity known as the expected discrepancy (Zucchini, 2000). Unfortunately this can never be calculated for real data but can be estimated using another value termed a criterion. All the advanced techniques studied in this paper consist of the calculation of five criteria, one for each model-free model, and the selection of the model with the lowest criterion. The frequentist methods of AIC or Akaike's Information Criteria (Akaike, 1973), AICc or small sample size corrected AIC (Hurvich and Tsai, 1989), bootstrap model selection (Lindhart and Zucchini, 1986), and cross-validation (Lindhart and Zucchini, 1986), as well as the Bayesian method BIC, Bayesian Information Criteria or Schwarz Criteria (Schwarz, 1978), will be applied to model-free analysis and a comparison of various model selection techniques presented.

Theory and methods

Model selection theory for NMR relaxation

For a single nucleus four different types of relaxation data sets exist, the true set R^{true}, the sample set R, the true back calculated set $R^{true}(\theta)$, and the back calculated set $R(\theta)$. A relaxation data set is defined as the collection of all the relaxation values for a single nucleus and θ is the vector whose elements are the model-free parameters. The true set is the true relaxation data underlying the measured data, but can never be observed due to noise. The sample set is the experimentally available or measured relaxation data set and is the true set plus noise. The true back calculated and back calculated sets are determined from the model-free parameters which are fitted using the true or sample sets respectively. The differences between the five model-free models are reflected in the two back calculated sets while the true and sample sets remain constant. For each of the four data sets there is a corresponding error set with the same dimension. By assuming gaussian errors the data and error sets together describe a set of normal probability distribution functions (pdfs) with one normal pdf for each data point. It is assumed that all four error sets are identical and therefore the one error set (σ) will be used in association with all four data sets.

The method of maximum likelihood is used to find the best fit model-free parameter values for each model-free model. By assuming gaussian errors a maximum likelihood estimate of the parameters is found by chi-squared minimisation, where the chi-squared statistic is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(R_i - R_i(\theta))^2}{\sigma_i^2}. \quad (1)$$

n is the dimension of the sets, R_i are the data points of the sample set R, $R_i(\theta)$ are the data points of the back calculated set $R(\theta)$, and σ_i are the values from the error set σ. After minimisation, the model-free parameter vector is denoted by the symbol $\hat{\theta}$. An alternative chi-squared statistic, labelled the true chi-squared statistic, is defined as

$$\chi^2_{true} = \sum_{i=1}^{n} \frac{(R_i^{true} - R_i(\theta))^2}{\sigma_i^2}. \quad (2)$$

R_i^{true} are the data points of the true set R^{true}.

Central to the frequentist model selection techniques is the concept of discrepancies. A discrepancy is represented by the symbol Δ, and can be any measure for lack of fit. The discrepancy chosen for this work is the common Kullback–Liebler discrepancy, Δ_{K-L} (Kullback and Leibler, 1951), which is associated with the concept of likelihood and is a statistical measure of the distance between the probability distribution functions of the true set and the back calculated set. The model with the lowest discrepancy is therefore defined as the model which best describes the true relaxation data set. Four variations of the Kullback–Liebler discrepancy are the realised discrepancy (Δ), the expected discrepancy ($E\Delta$), the empirical discrepancy (Δ_n), and the expected empirical discrepancy ($E\Delta_n$). The realised discrepancy, or simply the discrepancy for a given sample, is between the true set (R^{true}) and the back calculated set ($R(\theta)$). As different measurements will result in different sample sets, and as the true set is constant, the discrepancy varies due to the dependence of the back calculated set on the sample set. The discrepancy is therefore a random variable, and its average value is known as the expected discrepancy. The expected discrepancy is independent of the errors in the sample set and is fundamental to the derivation of the frequentist model selection techniques. The empirical discrepancy is defined as being between the sample set (R) and the back calculated set ($R(\theta)$). The expected empirical discrepancy is defined as the mean value of the discrepancies between back calculated sets, created by randomisation of the sample set, and the sample set and is a natural estimator of the expected discrepancy. Two additional classes of the Kullback–Liebler discrepancy are the discrepancy due to approximation, defined as being between R^{true} and $R^{true}(\theta)$, and the discrepancy due to estimation, defined as being between $R^{true}(\theta)$ and $R(\theta)$.

The empirical Kullback–Liebler discrepancy, Δ_n, is related to the likelihood formula where the likelihood for a given model-free model is represented by $L(\theta)$ and the log-likelihood by $l(\theta)$. θ is the k-dimensional model-free parameter vector, and n is the dimension of the data set. The formula for the empirical discrepancy is

$$\Delta_n = -\frac{1}{n} l(\hat{\theta}) \quad (3)$$

and the likelihood for a relaxation data sample set is

$$L(\theta) \equiv L(\theta|R) = \prod_{i=1}^{n} L(\theta|R_i). \quad (4)$$

$L(\theta|R)$ is the likelihood of the model-free parameter vector θ given the sample set R, and $L(\theta|R_i)$ is the

likelihood of the model-free parameter vector θ given the sample data point R_i. Fundamental to likelihood theory is the relationship (Edwards, 1972)

$$L(\theta|R_i) \propto P(R_i|\theta). \quad (5)$$

$P(R_i|\theta)$ is the probability of the sample data point R_i given the model-free parameter vector θ. The difference between likelihood and probability is that in probability, R_i is variable and θ is constant, while in likelihood, R_i is constant and θ is variable. Because the constant of proportionality is invariant between the model-free models, it can be ignored. By assuming gaussian errors for each relaxation data point in the set, the likelihood for a single data point is given by the equation

$$L(\theta|R_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(R_i - R(\theta))^2}{2\sigma_i^2}}. \quad (6)$$

By combining Equations 1, 4, 5, and 6, the likelihood is therefore

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(R_i - R(\theta))^2}{2\sigma_i^2}}, \quad (7)$$

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{i=1}^{n} \frac{1}{\sigma_i}\right) e^{-\sum_{i=1}^{n} \frac{(R_i - R(\theta))^2}{2\sigma_i^2}}, \quad (8)$$

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{i=1}^{n} \frac{1}{\sigma_i}\right) e^{-\frac{\chi^2}{2}}. \quad (9)$$

The log-likelihood is the natural logarithm of the likelihood and is therefore

$$l(\theta) = \ln(2\pi)^{-\frac{n}{2}} + \ln\left(\prod_{i=1}^{n} \frac{1}{\sigma_i}\right) - \frac{\chi^2}{2}, \quad (10)$$

$$l(\theta) = -\frac{n}{2}\ln 2\pi - \sum_{i=1}^{n} \ln \sigma_i - \frac{\chi^2}{2}. \quad (11)$$

Because the first two terms of the log-likelihood are constant under all models, the log-likelihood can be simplified to

$$l(\theta) = -\frac{\chi^2}{2}. \quad (12)$$

By combining Equations 3 and 12 the empirical discrepancy is therefore

$$\Delta_n = \frac{\chi^2}{2n}. \quad (13)$$

$L(\theta)^{true}$ and $l(\theta)^{true}$ are the likelihood and log-likelihood of the model-free parameter vector θ given the true set. If $L(\theta), l(\theta)$, and R_i are replaced with $L(\theta)^{true}, l(\theta)^{true}$, and R_i^{true} in the derivation of the log-likelihood the formula for the realised discrepancy is

$$\Delta = \frac{\chi^2_{true}}{2n}. \quad (14)$$

Importantly, the back calculated set is not replaced by the true back calculated set, otherwise the discrepancy due to approximation is derived rather than the realised discrepancy. Because the factor of $2n$ is constant for all model-free models, the realised discrepancy can be simplified to the value of the true chi-squared statistic. The simplest method for finding the value of the expected discrepancy is by Monte Carlo simulations. The realised discrepancy is calculated for x sample sets, created by randomisation of the true set using the error set and assuming gaussian errors, and the average of the x true chi-squared values is the expected discrepancy.

In practice, the expected discrepancy can never be calculated because the true set cannot be measured. Therefore an estimator of the expected discrepancy which uses only the sample set and back calculated set is required, and is known as a criterion. All of the frequentist model selection techniques used in this paper attempt to estimate the expected discrepancy through criteria, selecting the model with the lowest criterion value. The AIC and AICc criteria belong to a class of the frequentist model selection techniques known as the asymptotic methods. When the derivation of the expected discrepancy using only the sample and back calculated sets is extremely complex, as with NMR relaxation data, a simpler alternative is to derive the asymptotic value of the expected discrepancy as the sample size approaches infinity. AIC or Akaike's Information Criterion (Akaike, 1973) is the simplest of all model selection techniques and is defined by the formula

$$\frac{AIC}{2n} = \Delta_n + \frac{k}{n}. \quad (15)$$

By substitution with Equation 13, the formula for relaxation data is

$$AIC = \chi^2 + 2k. \quad (16)$$

k is the number of model-free parameters in the model. AICc is the small sample size corrected AIC (Hurvich and Tsai, 1989) and is defined by

$$\frac{AICc}{2n} = \Delta_n + \frac{k}{n} + \frac{k(k+1)}{n(n-k-1)} \quad (17)$$

and by substitution with Equation 13, the criterion is

$$AICc = \chi^2 + 2k + \frac{2k(k+1)}{n-k-1}. \quad (18)$$

n is the dimension of the relaxation data set. AICc is simply AIC with an additional term used to compensate for the size of the sample by causing under-fitting. Analogous to the asymptotic methods is the Bayesian Information Criteria or BIC, otherwise known as the Schwarz Criteria (Schwarz, 1978). BIC is similar in form to the asymptotic criteria but is derived independently from within a Bayesian framework. The defining formula is

$$\frac{BIC}{2n} = \Delta_n + \frac{k \ln n}{2n} \quad (19)$$

and by substitution with Equation 13, the criterion for relaxation data is

$$BIC = \chi^2 + k \ln n. \quad (20)$$

The bootstrap criterion is another method for estimating the expected discrepancy (Zucchini, 2000). If the sample set is considered as the true set, the expected empirical discrepancy ($E\Delta_n$) can be used as an estimator of the expected discrepancy ($E\Delta$). The sample set is randomised with gaussian noise using the error set to create x new sample sets. The empirical discrepancy is calculated for each randomised sample set as the true chi-squared statistic, defined as being between the original sample set which is equivalent to the true set, and the back calculated set. The mean value or expected empirical discrepancy over the x data sets is the bootstrap criterion.

The cross-validation model selection technique used in this paper is single-item-out cross-validation (Zucchini, 2000). The sample set is split into two different sets, the validation set consisting of a single data point R_i, and the calibration set consisting of the remaining $(n-1)$ data points. The model-free parameters are fitted using the calibration set and then used to compute the back calculated data point $R_i(\theta)$. The χ^2 value between the validation set data point R_i and the back calculated data point $R_i(\theta)$ is calculated. This is repeated for every value of i and the average χ^2 value is the cross-validation criterion. Because models 4 and 5 contain three parameters cross-validation can only be used on NMR relaxation data sets with $n > 3$, and therefore single field strength data cannot be used.

To minimise the under-fitting of the model selection technique currently used for model-free analysis (Mandel et al., 1995), high significance levels or α-levels for the hypothesis testing were used. For the chi-squared tests the level was set to 0.1, while for the F-tests the level was set to 0.2. The high chi-squared test is part of the flow diagram (Mandel et al., 1995) between the tests for models 2 and 3 and the tests for models 4 and 5. It checks the value of the chi-squared statistic for model 1 and if below an arbitrary cut-off value, model 1 is selected, otherwise the tests for models 4 and 5 are carried out. The chi-squared cut-off value was set to 20. The current technique was only designed for use on single field strength data ($n = 3$) but because this analysis uses both single and double field strength data, the method was extended for the larger data sets by the addition of extra chi-squared and F-tests (supplementary material). The original method was retained for the single field strength data. For the older model-free model selection technique the model selection rules were followed exactly as stated (Farrow et al., 1994).

Data analysis

By specifying the underlying motion of the NH bond vector, the performance of the different model selection techniques can be compared. To test the various methods two types of three-dimensional grid, labelled the R_{ex} grids and the Double Motion grids, were constructed to cover all model-free motions. The R_{ex} grids cover the motions represented by single model-free motions with chemical exchange and have the three dimensions S^2, τ_e, and R_{ex}. The Double Motion grids cover the model-free space where there are two internal motions and have the dimensions S_f^2, S_s^2, and τ_s. The values for the parameter dimensions were non-linear, being concentrated at regions where the distinction between the complex and the simpler parametrically restricted models are blurred by noise. This occurs for high S^2 values, and low τ_e and R_{ex} values. The values used for the order parameter dimensions are $S^2, S_s^2, S_f^2 = \{0.001, 0.388, 0.582, 0.698, 0.776, 0.831, 0.873, 0.905, 0.931, 0.952, 0.970\}$. The values for the correlation time dimensions in picoseconds are $\tau_e, \tau_s = \{0.1, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192\}$. The values for the R_{ex} dimension at 600 MHz are $R_{ex} = \{0, 0.149, 0.223, 0.332, 0.495, 0.739, 1.102, 1.644, 2.453, 3.660, 5.460, 8.145, 12.151, 18.127, 27.043\}$. The kinetic regime was set to fast exchange and therefore the R_{ex} values were scaled quadratically for the 500 MHz data. This quadratic field strength dependence may not be a valid assumption for real data (Millet et al., 2000) but is sufficient for the purposes of this analysis. Due to the

range of parameter values some of the grid points of the R_{ex} and Double Motion grids will never be observed in a real system. For the sake of completeness these were not excluded. An extremely popular misconception is that the range of correlation times for single internal motions is limited to motions which are in the extreme narrowing limit. While the extraction of correlation times in the nanosecond timescales is less accurate than picosecond correlation times (Lipari and Szabo, 1982a), the only limit on the parameter is the accuracy and quantity of the relaxation data (Jin et al., 1998). For the extended model-free theory the only limit on the correlation times is the separation of time scales by a magnitude determined by the local curvature of the model-free space which itself is dependent on the relaxation data, experimental errors, and model-free parameter values. After fitting, all model selection techniques additionally increase the magnitude of separation of time scales which can be extracted from the data by selecting simpler models. The total number of grid points for each R_{ex} grid and Double Motion grid are 2640 and 1936, respectively.

The ideal model to be selected for the R_{ex} grids is model 4, but when either one or both of the τ_e and R_{ex} dimensions are considered insignificant because of the noise, models 1, 2, and 3 are equally valid as these are parametric restrictions of the full model. The model-free parameter values for these simpler models are correct if the dropped parameter is truly insignificant. For the Double Motion grids the ideal model to be selected is model 5, but when the two motions cannot be distinguished, model 2 is a valid approximation. By equating the fast and slow correlation times in the full extended model-free correlation function (Clore et al., 1990) the model 2 parameters are given by $S^2 \equiv S_f^2 S_s^2$ and $\tau_e \equiv \tau_f \equiv \tau_s$ and the extracted model-free parameters can still be interpreted. Model 2 is also a parametric restriction of model 5 if either of the order parameters is equal to one. If S_f^2 is equal to one, $S^2 \equiv S_s^2$ and $\tau_e \equiv \tau_s$, and if S_s^2 is equal to one, $S^2 \equiv S_f^2$ and $\tau_e \equiv \tau_f$. In addition to this, if the values of the correlation times are insignificant, model 1 is also valid as it is a parametric restriction of model 5 where $S^2 \equiv S_f^2 S_s^2$ and $\tau_f = \tau_s = 0$.

Three perfect R_{ex} grids and three perfect Double Motion grids, consisting of noise-free relaxation values at either 500 MHz ($n = 3$), 600 MHz ($n = 3$), or both 500 and 600 MHz (double field strength, $n = 6$), were back calculated for each grid point. A schematic describing the creation of all grids is presented in the supplementary material. For the R_{ex} grids the model 4 equations were used for the back calculation, for the Double Motion grids the equations were those for model 5. The diffusion tensor used in the back calculation was isotropic with a correlation time of 10 ns. The NH bond length was fixed to 1.02 Å, the chemical shift anisotropy fixed to -160 ppm, and the angle between the principle axes of the chemical shift and dipolar tensors assumed to be zero.

The error sets were chosen to reflect experimental noise and consisted of the fixed values 0.04 for the 600 MHz NOE, 0.05 for the 500 MHz NOE, and 2% for all R_1 and R_2 values. When the NOE is zero, the corresponding error is at a minimum and equal to the error from the saturated spectrum divided by the peak intensity from the reference spectra. The exact NOE error depends on the peak intensities and errors from both saturated and reference spectra but, as this information is no longer present in the NOE, fixed errors were chosen as a balance between simplicity and accuracy. Each of the six perfect grids was randomised five times with gaussian noise creating thirty random grids represented by synthetic noisy relaxation data. Each true set representing a single point from the perfect grids and each sample set representing a single point from the random grids were fit separately to models 1 to 5. The program utilised for the fitting procedure was Modelfree 4.01 for Linux (Palmer et al., 1991; Mandel et al., 1995). The various model selection techniques were used to select the best model for each grid point.

For the Modelfree calculations an initial linear grid search over the model-free parameters was carried out before minimisation. Twenty increments were used for each parameter dimension with the search from 0 to 1 for the order parameters, 0 to 10,000 ps for the correlation times, and 0 to 20 s^{-1} for the R_{ex} values. Lower and upper bounds of 0 and 1 were placed on the order parameters while the correlation times and chemical exchange values were restricted to positive values. The chi-squared statistic (Formula 1) was used for model-free minimisation by setting the Modelfree program variable 'optimisation' to 'tval'. The diffusion tensor was fixed to 10 ns isotropic tumbling and the NH bond length and CSA values were set to the previously mentioned values. No trimming of the Monte Carlo results was used. For the perfect grids, which were used for the ideal model selection using the expected discrepancy, 500 Monte Carlo simulations were performed. The Modelfree program Monte Carlo variable 'sim_type' was set to 'expr' to randomise the original true set. For the random

grids, which were used for AIC, AICc, and BIC, and the current (Mandel et al., 1995) and older (Farrow et al., 1994) model-free model selection techniques, 200 Monte Carlo simulations were performed. So that the new sets would be created by randomisation of the back calculated set, 'sim_type' was set to 'pred'. These simulations were used to create the chi-squared and F-distributions necessary for the hypothesis tests. Separate model-free calculations were required for the bootstrap and cross-validation model selection techniques. For the bootstrap method the random grids were used in combination with 500 Monte Carlo simulations with 'sim_type' set to 'expr'. For the cross-validation method a model-free minimisation was carried out per single-item-out iteration and no Monte Carlo simulations were used. The results of model selection were visualised using the program OpenDX 4.1.3 from IBM (http://www.opendx.org).

Results

The results are best summarised as the percentages of the model-free models selected for each model selection technique. For the double field strength R_{ex} and Double Motion grids the results are shown in Tables 1 and 2, respectively, while for the single field strength 600 and 500 MHz R_{ex} and Double Motion grids the results are shown in Tables 3–6 in the supplementary material. For a single table the percentages are the average of the 5 random grids, except for the expected discrepancy which is obtained from the single perfect grid. These percentages are specific to the synthetic non-linear grids and can only be used for comparison of the performance of the various model selection techniques.

A problem inherent in the tables is that even though the percentages may appear reasonable a certain fraction may actually correspond to incorrect model selection where the parameter values of the fitted and selected model are far from the true values specified in the creation of the grids. By creating surfaces where the height corresponds to the difference between the final fitted and selected model-free parameter values and the original true values, anomalies which are present but invisible in the tables will result in a distortion of the surface. All problematic regions can be uncovered by visual inspection of the surfaces. If a parameter is not represented in the selected model, it is assumed to be equal to one for order parameters or zero for correlation times and relaxation due to exchange. Because the dimensionality of the most complex model used is three, a surface corresponding to the difference in a single model-free parameter will be four-dimensional. The fourth dimension, which is set to the S^2 and S_s^2 dimensions for the R_{ex} and Double Motion grids respectively, is split into a sequence of eleven three-dimensional surfaces. A selection of these plots is shown in Figures 1–3 while entire sequences are presented in the supplementary material. These surfaces represent the results of the first randomised grids. In these diagrams, a black sphere (red in the supplementary material) represents a grid point where no model is selected. For the R_{ex} grid surfaces, the S^2 difference is set to zero when no model is selected to enable the resolution of fine differences between selection techniques by preventing large distortions in the surface. For the Double Motion grid surfaces an additional sphere, which is white and smaller (cyan in the supplementary material), represents grid points where model 5 is selected and indicates where it is possible to differentiate between the two internal motions. From a dynamics perspective the final model-free parameter values are of more interest than the knowledge of which model was actually selected. The only information available from the model type is the number of internal motions yet the selection of a model describing a single motion does not necessarily mean that two internal motions are absent (Table 2, Figure 2). These surfaces demonstrate the accuracy and quality of the final dynamic picture uncovered using the various model selection techniques independent of which model was selected.

Discussion

From the results, two important flaws in the current model selection technique (Mandel et al., 1995) used in model-free analysis are evident. These are, not selecting a model when one should be selected, and under-fitting. For the current technique approximately 13% of the R_{ex} grids (Tables 1, 3, and 5) and approximately 9% of the Double Motion grids (Tables 2, 4, and 6) have no model selected. These percentages are independent of field strength or set size thereby validating the extension of the technique for multiple field strength data ($n > 3$) by the addition of chi-squared and F-tests for models 4 and 5 (supplementary material). A few of the grid points for which no model is selected are clustered in a region in the model-free space covered by the R_{ex} grids where τ_e values are in

Table 1. Percentages of model-free models selected for the five randomised R_{ex} grids (double field strength data) consisting of the three dimensions $\{S^2, \tau_e, R_{ex}\}$

Selection method	No model	Model 1	Model 2	Model 3	Model 4	Model 5
AIC	0	8.6	13.5	27.4	49.1	1.5
AICc	0	14.5	19.0	32.9	32.6	1.0
BIC	0	8.0	13.2	27.0	50.2	1.6
Bootstrap	0	5.1	9.0	24.0	57.4	4.6
Cross-validation	0	7.9	11.3	27.6	47.9	5.3
Old technique	18.9	4.8	9.4	22.7	43.9	0.3
Current technique	13.3	16.2	13.7	25.0	31.8	0.1
Realised discrepancy	0	4.6	10.4	26.3	56.3	2.5
Expected discrepancy	0	6.6	10.8	27.7	52.3	2.6

Table 2. Percentages of model-free models selected for the five randomised Double Motion grids (double field strength data) consisting of the three dimensions $\{S_f^2, S_s^2, \tau_s\}$

Selection method	No model	Model 1	Model 2	Model 3	Model 4	Model 5
AIC	0.4	30.9	33.7	2.3	2.7	30.0
AICc	0.4	42.1	36.0	0.4	0.2	21.0
BIC	0.4	29.6	33.8	2.6	3.1	30.5
Bootstrap	0	24.1	24.7	3.6	6.8	40.8
Cross-validation	0	29.9	22.2	3.0	6.7	38.1
Old technique	21.5	16.0	31.2	3.8	7.6	19.9
Current technique	8.7	43.2	27.1	0.4	0.7	19.9
Realised discrepancy	0	27.8	35.7	1.6	2.3	32.6
Expected discrepancy	0	33.7	27.6	0.0	0.7	37.9

the range of 128 to 512 picoseconds (supplementary material). The origin of this problem was traced to an apparent bug in the Modelfree program where in certain circumstances the Levenberg-Marquardt minimisation failed to optimise the parameter values obtained from the initial grid search. This problem is currently being investigated (A.G. Palmer, personal communication). Apart from this region the majority of affected points appear to be randomly distributed across the model-free space and are most likely the result of randomly large errors in the affected sample sets. In the older technique (Farrow et al., 1994) the problem of not selecting a model is more severe with approximately 20% of the double field strength grids (Tables 1 and 2) and approximately 15% of the single field strength grids with no model selected (Tables 3–6 in the supplementary material).

The second major problem of the current model selection technique (Mandel et al., 1995) is under-fitting. This is revealed by comparison with the expected discrepancy and AICc model selection. Theoretically the expected discrepancy is the perfect balance between bias and variance, meaning that neither under nor over-fitting occurs, and therefore the percentages of models selected are the ideal situation. In comparison, the small sample size corrected AIC or AICc uses the property of under-fitting to select the best model, where the additional term in the AICc formula induces under-fitting as a compensation for the size of the sample. From Tables 1 and 2 the percentages of models selected for the current technique are similar to AICc but are far from the balance of the expected discrepancy demonstrating the under-fitting. In model-free analysis, under-fitting is manifested as the selection of a model which is a parametric restriction of the best model. The model-free results incorporate less experimental noise but this artificially underestimates parameter uncertainty (Andrec et al., 1999). In addition there is an increase in bias which skews the final results. When the underlying motion is best described

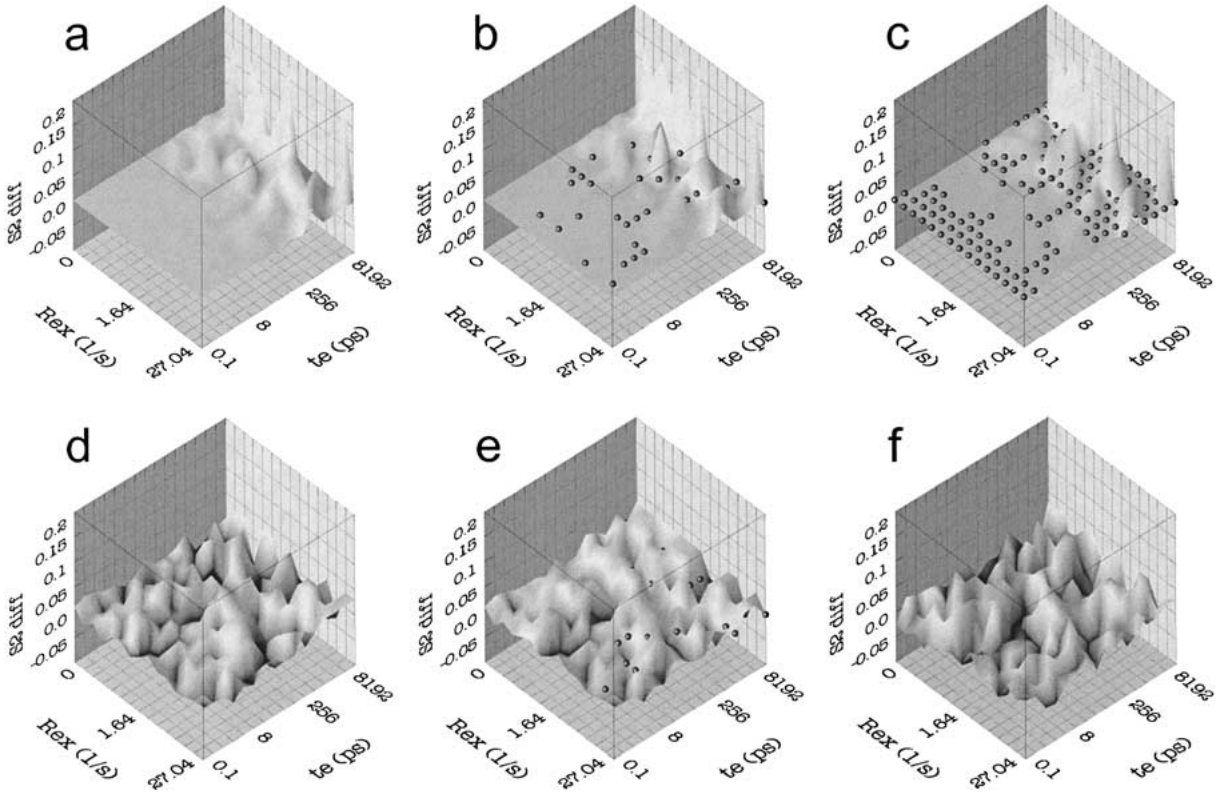

Figure 1. S^2 difference surfaces for the first randomised R_{ex} grids, where the difference is between the final fitted and selected parameter value and the true value specified in the creation of the grids. The S^2 dimension of the grid has been split into a series of graphs, with only a single member of the series depicted. Black spheres represent grid points where no model has been selected, in these cases the difference is set to zero to visualise the fine distinctions between the techniques. The model selection results for the double field strength grid when $S^2 = 0.001$ are: (a) AIC model selection (Akaike, 1973); (b) current model-free model selection (Mandel et al., 1995); (c) an older model-free model selection technique (Farrow et al., 1994). These results demonstrate a flaw of the current and the older model-free model selection techniques whereby a high proportion of the grid points have no model selected. The model selection results for the double field strength grids when $S^2 = 0.970$ are: (d) AIC model selection; (e) current model selection. These reveal the second flaw of the current model-free model selection, under-fitting, which results in the overestimation of S^2 and underestimation of τ_e and R_{ex}. The overestimation of S^2 is evident as the lighter elevated surface in (e); (f) AIC model selection when $S^2 = 0.970$ for the 600 MHz grid demonstrating the increase in noise of the model-free parameters compared to the double field strength data in (d).

by model 4 but either models 1, 2, or 3 are selected, the result is that S^2 is overestimated and τ_e and/or R_{ex} are underestimated. An example of this is the lighter elevated S^2 difference surface in Figure 1e when $S^2 = 0.970$. A numerical example of the potential S^2 overestimation is for the grid point $\{S^2 = 0.698, \tau_e = 64, R_{ex} = 12.151\}$ of the first randomised double field strength R_{ex} grid, where the fitted parameter values are model 1 $\{S^2 = 0.913\}$, model 2 $\{S^2 = 0.877, \tau_e = 88.74\}$, model 3 $\{S^2 = 0.770, R_{ex} = 10.885\}$, and model 4 $\{S^2 = 0.697, \tau_e = 67.91, R_{ex} = 11.952\}$. As the parameter number decreases the value of S^2 generally increases demonstrating how under-fitting causes overestimation of S^2. τ_e and R_{ex} are underestimated, their value being zero, because of their exclusion from the final model. When the underlying motion is best described by two internal motions (model 5) under-fitting results in the selection of models 1 or 2 (Tables 2, 4, and 6, Figure 2) hiding one of the two motions by merging the two order parameters into the single parameter $S^2 = S_f^2 S_s^2$. This occurs naturally when the two timescales are close to each other but is exaggerated by under-fitting. Over-fitting results in the selection of an overly complex model in which one of the model-free parameters is insignificant, its existence being blurred by noise. This occurs where order parameters are close to one and correlation times and chemical exchange are close to zero. The exclusion of the parameter would be beneficial to the final result by reducing the variance, thereby in-

creasing the stability and precision of the final model. More experimental noise is reflected in the selected model which randomly increases the error associated with the final model-free parameter estimates. Due to the randomness of this effect the consequences of over-fitting are far less detrimental to the final results than the consequences of under-fitting, even if single field strength data is all that is available. Because the current technique (Mandel et al., 1995) under-fits, it breaks the principle of parsimony distorting the results and causing the protein to appear much more rigid than it actually is.

When the expected discrepancy is used for model selection, the proportion of grid points with model 5 selected in the double field strength Double Motion grids is much higher (38%) than can be achieved using any of the other model selection techniques (Table 2). For the criteria based model selection techniques the best that can be realistically achieved is 30%. The bootstrap and cross-validation techniques do have proportions similar to the expected discrepancy but this is due to the over-fitting of the techniques and these grid points tend to be randomly placed in the region of the Double Motion grids where the correlation time is low and the two internal motions cannot be separated (Figure 3c). The fitted and selected model-free results for these points are far from the true parameter values. An example in the first randomised double field strength Double Motion grid using bootstrap model selection is where the true values of the model-free parameters are $\{S_f^2 = 0.931, S_s^2 = 0.582, \tau_s = 16\}$ and the fitted and selected values are $\{S_f^2 = 0.580, S_s^2 = 0.928, \tau_s = 298.7\}$. As the true correlation time is located below the τ_s cut-off, defined as the lowest correlation time where the two motions can be discriminated, the best result would be the selection of either model 1 or 2 but because model 5 is selected the corresponding percentage is artificially high. Often in these cases the order parameters are reversed and the correlation time overestimated, the result being the selection of an artificial motion on a time scale much slower than the true model where in reality no such motions exist. There are two explanations why the selection of model 5 by the criteria based methods can only reach 30%, while the expected discrepancy reaches 38%. Firstly, a proportion of the extra 8% is due to the expected discrepancy incorrectly selecting model 5 because of the bug in the Modelfree program, as described previously. There is a trend for these points to be clustered in a band with the same τ_s value (Figure 3a) located below the τ_s cut-off with its exact position dependent on the value of S_s^2. Secondly, the τ_s cut-off appears to move to higher correlation times for noisy data. This is reflected in the differences in percentages between the expected and realised discrepancies which are calculated using noise-free and noisy data respectively. Since no model selection technique using noisy relaxation data can replicate results of the expected discrepancy due to the τ_s cut-off change and the effects of the program bug, the model selection techniques were compared to the realised rather than expected discrepancy for the Double Motion grids.

Due to the skewing of results caused by under-fitting, AICc model selection is not recommended for model-free analysis. Another reason for not using AICc is because model selection actually fails for single field strength data (Tables 3–6). This is because if data sets with $n \leq 4$ are used together with AICc model selection the denominator of the additional term becomes zero for certain models causing the value of the criterion to be infinite. Due to the extremely large proportion of the grids with no model selected (Figure 1c), as well as model selection inconsistencies between the R_{ex} and Double Motion grids, the older method used for model-free analysis (Farrow et al., 1994) is not recommended either. By comparison of the R_{ex} grid percentages to the expected discrepancy (Tables 1, 3, and 5), the results are not too far from the ideal situation with neither under nor over-fitting occurring. The percentages are lower for all models due to the high proportion of the grids with no model selected. For the Double Motion grids, comparison of the older technique to the realised discrepancy (Tables 2, 4, and 6) reveals a different pattern of results, under-fitting occurring for models 2 and 5 and over-fitting occurring for models 1, 3, and 4. These inconsistencies arise because of fundamental flaws in the selection technique.

Both the bootstrap and cross-validation model selection perform relatively well. The percentages are similar to the expected discrepancy (all tables), the only difference being slight over-fitting. Cross-validation is not represented in Tables 3–6 in the supplementary material because the technique can only work when $n > 3$. The slight over-fitting of the bootstrap method is evident in the R_{ex} grids where the percentages for models 1 to 3 are lower than the expected discrepancy while models 4 and 5 are higher. In the Double Motion grids models 1 and 2 are lower than the realised discrepancy and models 3 to 5 are higher. The results of the cross-validation model se-

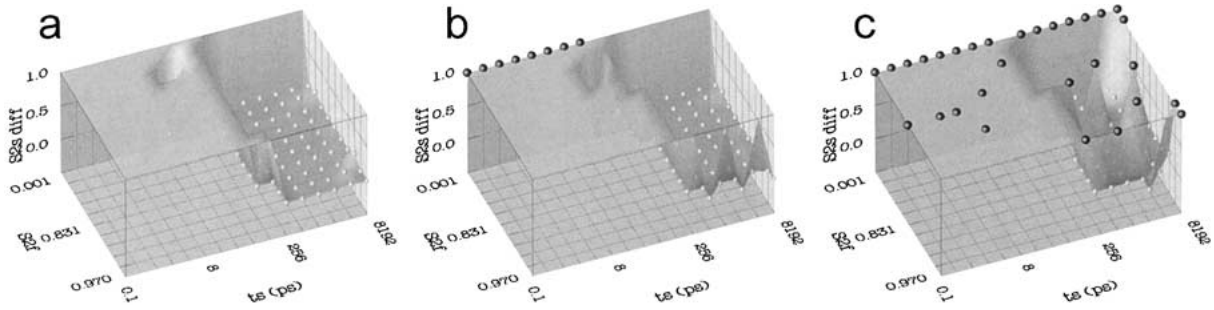

Figure 2. S_s^2 difference surfaces for the first randomised double field strength Double Motion grids. The S_s^2 dimension of the grid has been split into a series of graphs and only results for $S_s^2 = 0.001$ are shown. Black spheres represent grid points where no model has been selected, white spheres represent grid points where model 5 has been selected. When no model is selected the S_s^2 value is assumed to be one. The plots represent: (a) Model selection using the expected discrepancy (Zucchini, 2000); (b) AIC model selection (Akaike, 1973); (c) current model-free model selection (Mandel et al., 1995). These surfaces demonstrate where the two internal motions can be separated. The expected discrepancy neither under nor over-fits, therefore this is the ideal result. Because the expected discrepancy cannot be calculated for real NMR data this figure demonstrates the improvement in results using AIC compared to the current technique.

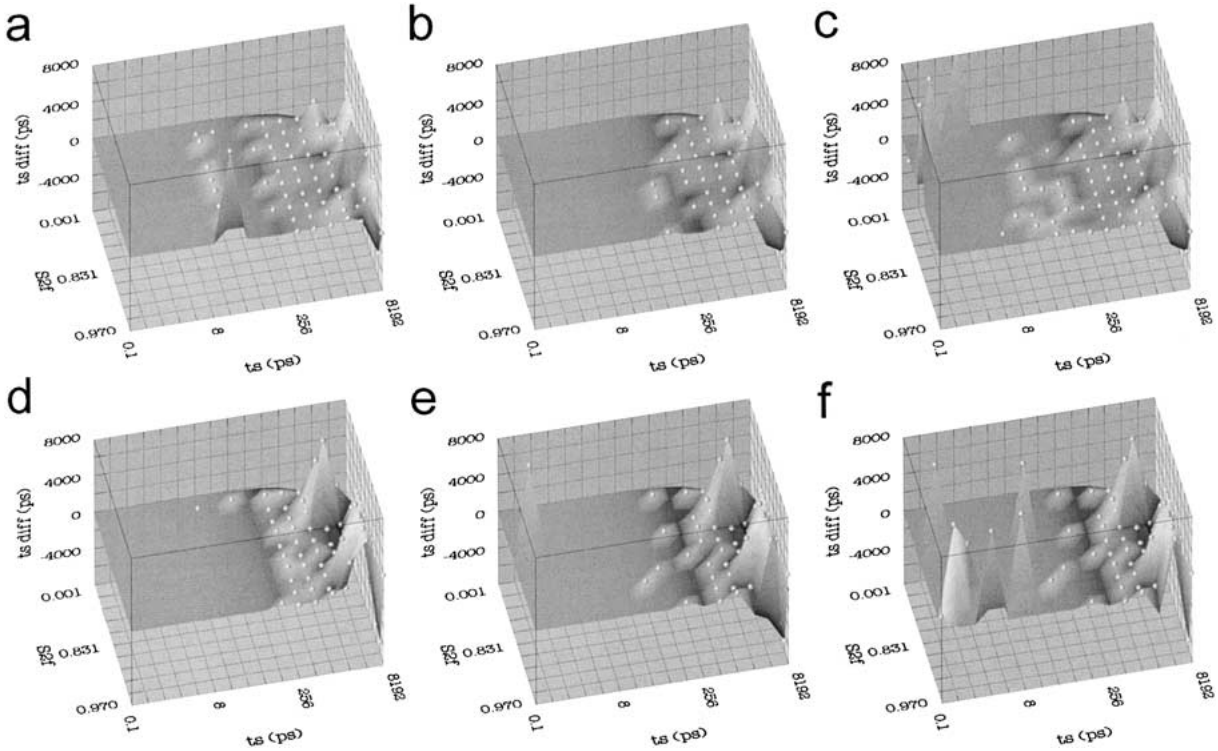

Figure 3. τ_s difference surfaces for the first randomised Double Motion grids, where the S_s^2 dimension of the grid has been split into a series of graphs. White spheres represent grid points where model 5 has been selected. The model selection results for the double field strength grid when $S_s^2 = 0.582$ are: (a) Model selection using the expected discrepancy (Zucchini, 2000); (b) AIC model selection (Akaike, 1973); (c) bootstrap model selection (Zucchini, 2000). These three surfaces demonstrate why both the expected discrepancy and bootstrap model selection have higher percentages for model 5 in the Double Motion grids compared to the other techniques. The additional percentages are artificial, being placed below the τ_s cut-off which is located at approximately 128 ps in this case, and these models have incorrect parameter estimates. The model selection results for the 600 MHz grid when $S_s^2 = 0.873$ are: (d) The expected discrepancy; (e) AIC model selection; (f) BIC model selection (Schwarz, 1978). The BIC model 5 percentages are higher than AIC for the 600 MHz Double Motion grids. These surfaces show this is artificial rather than truly reflecting the percentages of the expected discrepancy.

lection are different. For the R_{ex} grids the percentages for models 1–3 are almost the same as the expected discrepancy while model 4 is lower and model 5 is higher. For the Double Motion grids, model 1 is similar to the realised discrepancy while model 2 is lower and models 3–5 are higher. The bootstrap technique is known to over-fit (Zucchini, 2000) but the reason for the distinctive cross-validation results is unclear, especially since these results are not what would be expected from normal over-fitting. The result is probably due to either the Modelfree bug as discussed above or due to artefacts of the technique. Because over-fitting results in the selection of artificial motions in the Double Motion grids, as discussed above, bootstrap and cross-validation model selection techniques are not recommended.

The best performing techniques are AIC and BIC model selection with percentages closest to the expected discrepancy in Tables 1, 3, and 5 and closest to the realised discrepancy in Tables 2, 4, and 6. Because of the similarity in percentages, AIC and BIC therefore neither under nor over-fit. By avoiding under-fitting, the final results do not overestimate precision and are free of bias which would otherwise respectively cause error estimates to be too small and order parameters to be overestimated and correlation times and chemical exchange to be underestimated. By avoiding over-fitting, the final results do not underestimate precision, which would cause excessive parameter uncertainty and model instability, and are free of artificial motions. The validity of these percentages is confirmed by visual inspection of the full difference surfaces. For double field strength data the two techniques perform almost perfectly with the average difference value centred at zero (Figures 1–3 and supplementary material). After accounting for the experimental noise, visible as random perturbations throughout the plots, the AIC and BIC surfaces are free of distortions due to model selection. No grid points have difference values visible above the noise with the exception of two regions which are due to minimisation artefacts rather than problems associated with the selection technique. These regions, which are explained below, are therefore independent of model selection and visible throughout all surfaces. AIC and BIC model selection perform almost identically for double field strength data but, when the results for the single field strength grids are studied, AIC actually performs better than BIC. For the R_{ex} grids the AIC percentages are similar to the expected discrepancy while the BIC percentages are closer to the realised discrepancy. When both single field strength Double Motion grids are studied the performance of the two techniques varies. BIC appears to select model 5 more often but this is abnormal as these grid points are randomly placed below the τ_s cut-off resulting in approximately 4% of all grid points displaying artificially slow motions (Tables 4 and 6, and Figure 3f). AIC also suffers from this problem although much less frequently with around 0.2% of grid points affected. BIC also has the disadvantage of selecting quite a high percentage of models 3 and 4. Therefore, because of the almost perfect percentages and surfaces for all data studied in this analysis, AIC is the recommended technique to use in combination with any type of relaxation data set.

Two regions in the difference surfaces for all model selection techniques stand out from the noise and are caused by minimisation problems rather model selection artefacts. The first is in the correlation time surfaces for both the R_{ex} and Double Motion grids when the true internal correlation time approaches the global rotational correlation time (Figure 3 and supplementary material). Model-free fitting fails in finding the correct parameter values and in most cases the correlation time dimension of the model-free space never converges with the final value either stuck at the lower or upper bounds. Because the value of the correlation time cannot be resolved, the estimates of the S^2 and R_{ex} parameters are distorted yet are still close to their true values (Figures 1 and 2). The problem increases in severity as S^2 values approach one and is currently unavoidable. The second region is in the R_{ex} grid where correlation times are around 128 to 512 ps and is due to the bug in the Modelfree program discussed previously.

Both the size and the highest field strength of the data set make a difference to the final model-free results. One of the major advantages in the collection of six data points is the reduction of the noise levels of the final model-free parameters (Figures 1d, 1f), thereby increasing the accuracy of the final results. By comparing AIC model selection for the double field strength grids with the single field strength grids (all tables), the proportion of models 4 and 5 for both the R_{ex} and Double Motion grids is higher for both the double field strength and 600 MHz grids, than for the 500 MHz grids. Therefore from these results, the extraction of more complicated motions is dependant on the strength of the highest magnetic field that data was collected at rather than the size of the relaxation data set. The reason for this may be due to the reduction of

experimental noise as field strength increases, or may be due to a property of the relaxation and model-free equations. Further study is required to validate and explain this result.

In reality the underlying dynamics of the NH bond vector is far more complex than could ever be modelled by model-free or any other theory which uses limited relaxation data. This complexity is evident when the information content of model-free analyses is compared with MD simulations. Therefore trying to find the truth, the correct model, which is defined as the complete description of the dynamics, is experimentally impossible. The selected model must be considered as an approximation to the truth and in using the frequentist methods this is the model which best describes the true data set (R^{true}). This does not mean that the information extracted is not physically meaningful. Without prior knowledge of the true underlying dynamics, it is not possible to determine the degree of approximation. However the quality of the model can be measured using the chi-squared statistic and criterion values. The chi-squared statistic indicates how close the model is to the experimental data (R) and is inversely proportional to the likelihood of the model (Edwards, 1972). The comparison of the relative criterion values between the contending models is related to the concept of model selection uncertainty (Linhart and Zucchini, 1986; Jin et al., 1998) where, if two models have nearly identical AIC values they can be considered to describe the true data equally well. Caution must then be taken in reaching a conclusion. If the models are nested, in most cases the additional parameter is on the border of being insignificant. If the two model-free models are non-nested this may be an indication that a more complex model is required, one which contains the parameters of both models, although this is only possible provided the number of parameters is less than or equal to the number of data points in the relaxation data set. To reach valid conclusions, careful analysis of AIC values, chi-squared statistics, and model-free parameter values is required.

The fundamental aim of the current technique (Mandel et al., 1995) is to find the correct description of the underlying motion from a set of contending models, and if the true model is not within that set, no model will be selected. Yet in this analysis, the current model selection is shown to not select a model when the true model is actually within the set. The black and white concept of classifying a model as either correct or incorrect is counter intuitive to the philosophy of model selection due to the shades of grey introduced by the underlying complexity of the true model, the parametric restrictions of nested models, and the randomisation of data by experimental noise. For the frequentist and Bayesian model selection techniques the aim is not to find the correct description of the motion, but to select the model which best describes the data. This is why these techniques always select a model. The one special case where the frequentist and Bayesian methods cannot select a model is when the chi-squared value for all models is infinite. This occurs in certain rare cases in the Double Motion grids when both the S_f^2 and S_s^2 values are 0.001 and τ_s values are low and is due to the artificial truncation of a value in the error set to zero causing infinite chi-squared values (Tables 2, 4, and 6; Figure 2b). Although in this comparison the truth is very simple, being perfect model-free motions, the frequentist and Bayesian methods were designed with the assumption that the underlying model may never be fully reflected in the data, therefore their application to real NMR relaxation data is completely justifiable.

For model-free analysis using the current technique (Mandel et al., 1995) Monte Carlo simulations for each model are required to generate the appropriate chi-squared and F-distributions for hypothesis testing. An additional advantage of using AIC model selection is that these simulations are no longer necessary. A single model-free minimisation and resultant chi-squared statistic per nucleus is all that is needed. The parameter error estimates can be determined following model selection in a final minimisation run with Monte Carlo simulations. Assuming that the simulations would only be used for the five model-free model runs and no final optimisations would be carried out, the use of AIC would only require a fifth of the computation time of the current technique (Mandel et al., 1995). If, in addition, the diffusion tensor parameters are optimised using a final run and are followed by multiple iterations of the fitting, model selection, and final optimisation, all of this can be accomplished without the use of Monte Carlo simulations. As well as speeding up model-free analysis by decreasing the computation time, the use of AIC would also simplify analysis by removing the need for multiple statistical tests and the adherence to a complex flow diagram for model selection. All that is needed for AIC is the calculation and comparison of five criteria using the simple formula $\chi^2 + 2k$. Because AIC is able to compare non-nested models, the model-free models can be directly compared to other non model-free motional

models as long as the dependence of model-free parameters on the global rotational diffusion tensor is not overlooked. This makes it the perfect tool for comparison of the model-free theory with alternative theories such as the Slowly Relaxing Local Structure (SRLS) (Tugarinov et al., 2001). In addition, as the SRLS theory contains multiple models which are discriminated using hypothesis testing, based on the current model-free technique, the use of AIC for SRLS model selection will increase the accuracy of these results as well. AIC can easily be extended to other relaxation applications including the decision of whether an isotropic, an axially symmetric, or a fully anisotropic diffusion tensor best describes the global tumbling of the protein. In this case the AIC criterion is identical to equation 16 where the chi-squared statistic is simply the sum of the chi-squared values over all relaxation sets and k is the sum of the number of global rotational diffusion parameters and the total number of model-free parameters describing the system. A possible consequence of this application may be the effective discrimination between chemical exchange effects and anisotropic tumbling (Osborne and Wright, 2001; Pawley et al., 2001).

Conclusions

Two major flaws of the current model selection technique (Mandel et al., 1995), which is almost universally used in the model-free analysis of protein dynamics, have been identified by the study of synthetic data. These are, not selecting a model when one ought to be selected, and under-fitting. The consequence of not selecting a model is that a proportion of the protein backbone will not be described using model-free parameters. The consequences of the under-fitting are that the final model-free results are skewed with S^2 overestimated and τ_e and R_{ex} underestimated, and if two internal motions exist these are not separated. The principle of parsimony states that the simplest model which fits well should be selected and since under-fitting results in a model which is too simple, the current technique (Mandel et al., 1995) breaks this principle. As a result the protein falsely appears more rigid than it really is.

The performance of the current model-free model selection technique was compared to an older model-free technique (Farrow et al., 1994), to various frequentist model selection techniques including AIC, AICc, bootstrap methods, cross-validation, the realised and expected discrepancies, and to the Bayesian method BIC. Synthetic noisy relaxation data covering all current model-free motions was used for the comparison and it is concluded that the most accurate model-free results for experimental relaxation data can be realised using AIC model selection. AIC is the best implementation of Occam's razor and the principle of parsimony. Its use will increase the accuracy, speed, and simplicity of model-free analysis.

Acknowledgements

We thank Andrew J. Perry, Dr James D. Swarbrick, and Haydyn D.T. Mertens for valuable discussions. This work was supported by the ARC (to PRG).

References

Abragham, A. (1961) *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford.

Akaike, H. (1973) In *Information Theory and an Extension of the Maximum Likelihood Principle*, Petrov, B.N. and Csaki, F. (Eds.), Proceedings of the 2nd International Symposium on Information Theory, Budapest, Academiai Kiado, pp. 267–281.

Andrec, M., Inman, K.G., Weber, D.J., Levy, R.M. and Montelione, G.T. (2000) *J. Magn. Reson.*, **146**, 66–80.

Andrec, M., Montelione, G.T. and Levy, R.M. (1999) *J. Magn. Reson.*, **139**, 408–421.

Burnham, K.P. and Anderson, D.R. (1998) *Model Selection and Inference: A Practical Information-Theoretic Approach*, Springer-Verlag, New York.

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

Edwards, A.W.F. (1972) *Likelihood*, Cambridge University Press, London.

Farrow, N.A., Muhandiram, R., Singer, A.N., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) *Biochemistry*, **33**, 5984–6003.

Hurvich, C.M. and Tsai, C-L. (1989) *Biometrika*, **76**, 297–307.

Jin, D., Andrec, M., Montelione, G.T. and Levy, R.M. (1998) *J. Biomol. NMR*, **12**, 471–492.

Korzhnev, D.M., Orekhov, V.Y. and Arseniev, A.S. (1997) *J. Magn. Reson.*, **127**, 184–191.

Kullback, S. and Leibler, R.A. (1951) *Ann. Math. Stat.*, **22**, 79–86.

Linhart, H. and Zucchini, W. (1986) *Model Selection*, John Wiley and Sons, New York.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.*, **104**, 4546–4559.

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.*, **104**, 4559–4570.

Mandel, A.M., Akke, M. and Palmer, A.G. (1995) *J. Mol. Biol.*, **246**, 144–163.

Millet, O., Loria, J.P, Kroenke, C.D., Pons, M. and Palmer, A.G. (2000) *J. Am. Chem. Soc.*, **122**, 2867–2877.

Osborne, M.J. and Wright, P.E (2001) *J. Biomol. NMR*, **19**, 209–230.

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

Pawley, N.H., Wang, C., Koide, S. and Nicholson, L.K. (2001) *J. Biomol. NMR*, **20**, 149–165.
Schwarz, G. (1978) *Ann. Stat.*, **6**, 461–464.
Tugarinov, V., Liang, Z., Shapiro, Y.E., Freed, J.H. and Meirovitch, E. (2001) *J. Am. Chem. Soc.*, **123**, 3055–3063.
Zucchini, W. (2000) *J. Math. Psychol.*, **44**, 41–61.