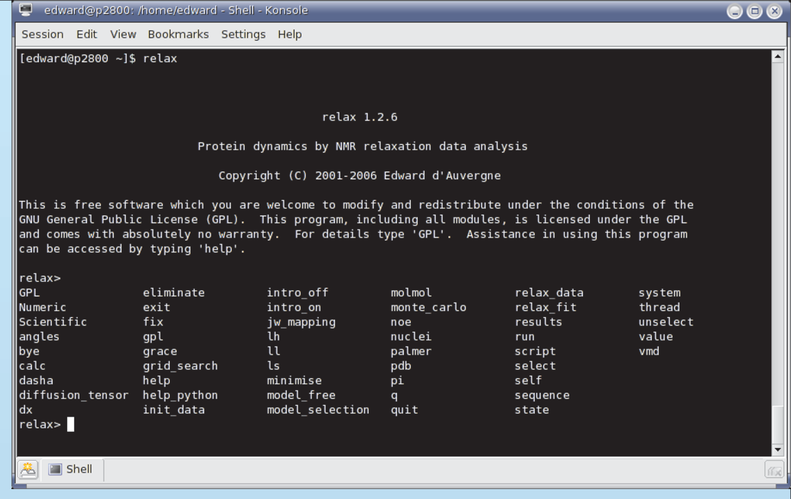
The program relax

Improving optimisation and model-free analysis in reverse

Edward d'Auvergne and Paul Gooley

Department of Biochemistry and Molecular Biology, Bio21 Institute of Biotechnology and Molecular Science, University of Melbourne, Victoria 3010, Australia



Model-free model elimination

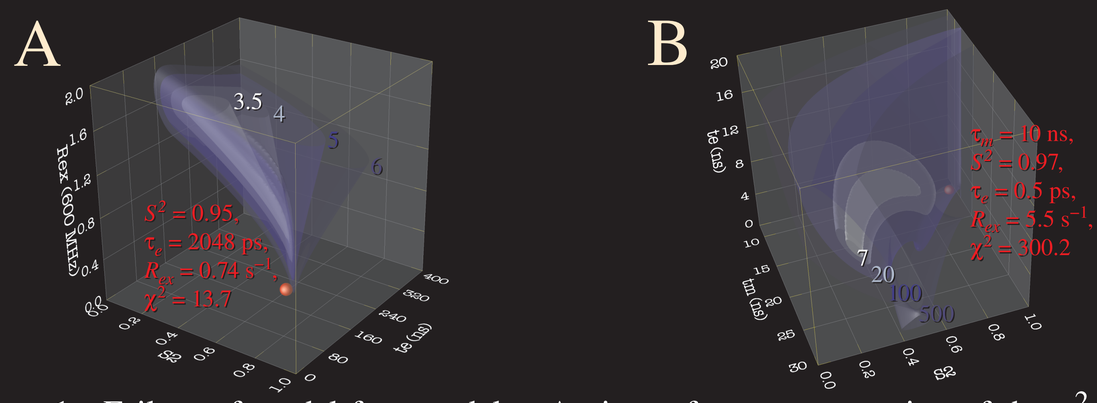


Figure 1: Failure of model-free models. An isosurface representation of the χ^2 space where the red sphere is placed at the true parameter values and the relaxation data has been randomised. In A) failure is caused by noise whereas in B) failure is due to the absence of the R_{ex} parameter. Both are selected over all other models.

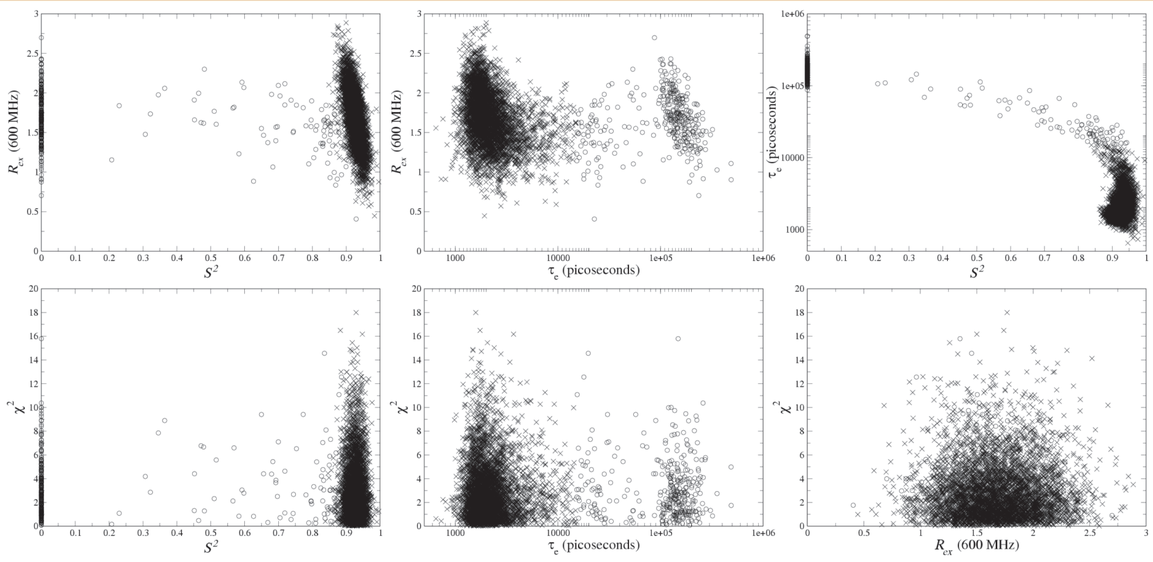


Figure 2: The failure of Monte Carlo (MC) simulations. Plotted are the results of 5000 simulations whereby the true parameters are $\{S^2 = 0.93, \tau_e = 2048$ ps, $R_{ex} = 1.64$ s$^{-1}\}$.

As failed model-free models frequently have low χ^2 values they are often selected to represent the dynamics of the residue. Hence the final result can be far from the truth. The related failure of MC simulations causes a significant overestimation of the parameter errors. Therefore failed models and simulations should respectively be eliminated prior to model selection and during error analysis.

Model-free optimisation

After testing a large number of minimisation algorithms – including many line search, trust region, and conjugate gradient techniques as well as the simplex and Levenberg-Marquardt algorithms – the most efficient and reliable local optimisation algorithm within the convoluted model-free space was found to be Newton minimisation.

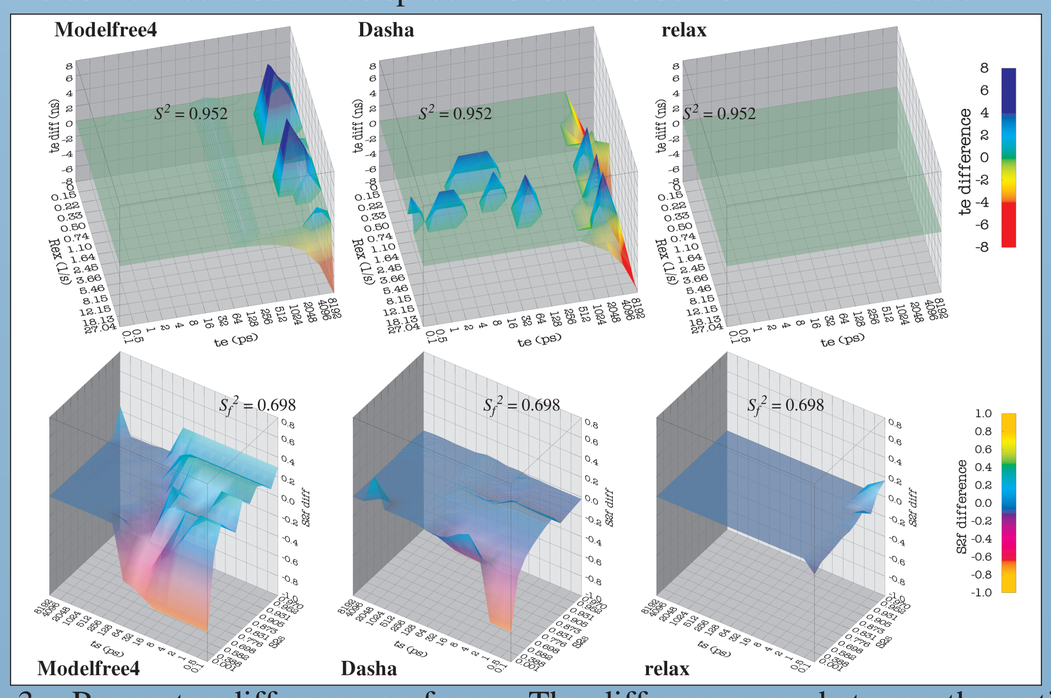


Figure 3: Parameter difference surfaces. The differences are between the optimised value and true value of each grid point. If the minimum has been found for all points the surface should be flat with a height of zero. Positive and negative heights correspond to over and under-estimation respectively.

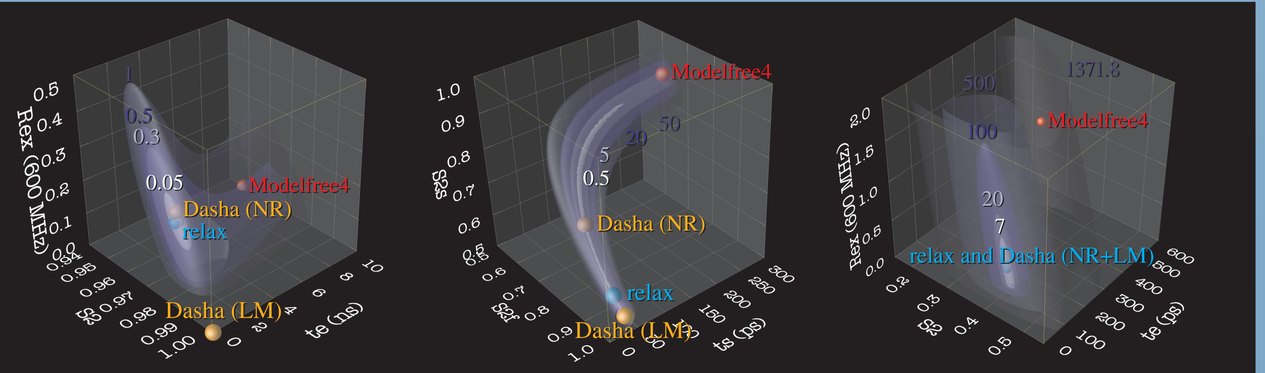


Figure 4: Optimisation failure. The cyan sphere represents the true parameter values. Four types of failure include the singular matrix failure of the Levenberg-Marquardt algorithm, low precision, the inability to slide along the limits, and a bug in Modelfree4.

Model-free set theory

Model-free analysis is a problem intricately linked to the Brownian rotational diffusion of the molecule. The diffusion parameters influence the values of the model-free parameters and vice versa. Model selection is also influenced by the tensor. These properties create a multi-universe, multi-minima optimisation and model selection problem. The entirety of the problem is encapsulated by the universal set

$$\mathfrak{U} = \bigcup_{i=1}^{l \cdot m \cdot n} \mathfrak{S}_i,$$

where l is the number of residues, m is the number of model-free models, n is the number of diffusion models, and $\mathfrak{S}$ is the global model of all model-free and diffusion parameters. The solution within $\mathfrak{U}$, the universal solution, can be formulated as

$$\widehat{\mathfrak{U}} = \hat{\theta} \in \left\{ \mathfrak{S} : \min_{\hat{\theta} \in \mathfrak{U}} \Delta_{\text{K-L}}(\hat{\theta}) \right\}, \quad \text{s.t. } \hat{\theta} = \arg\min \left\{ \chi^2(\theta) : \theta \in \mathfrak{S} \right\},$$

where θ is the parameter vector and $\Delta_{\text{K-L}}$ is the Kullback-Leibler discrepancy.

New optimisation protocol

The current approach to finding the universal solution is: start with a diffusion tensor estimate, optimise the model-free models, select the best model, and optimise the global model. These steps are repeated until convergence. Here the reverse is proposed – the model-free parameters are optimised in the absence of any global model parameters, the best model is selected, and finally the diffusion tensor is optimised.

Five category of global model $\mathfrak{S}$ (defined above) can be constructed: each residue having its own local τ_m (MI), spherical diffusion (MII), the prolate spheroid (MIII), the oblate spheroid (MIV), and the ellipsoid (MV). Firstly model MI is optimised and the best model-free models are selected. The parameters for MII to MV are copied from MI, the local τ_m removed, and finally the diffusion tensor is optimised. Then, iteratively until convergence, the model-free models are optimised, failed models are eliminated, the best selected, and the global model is optimised.

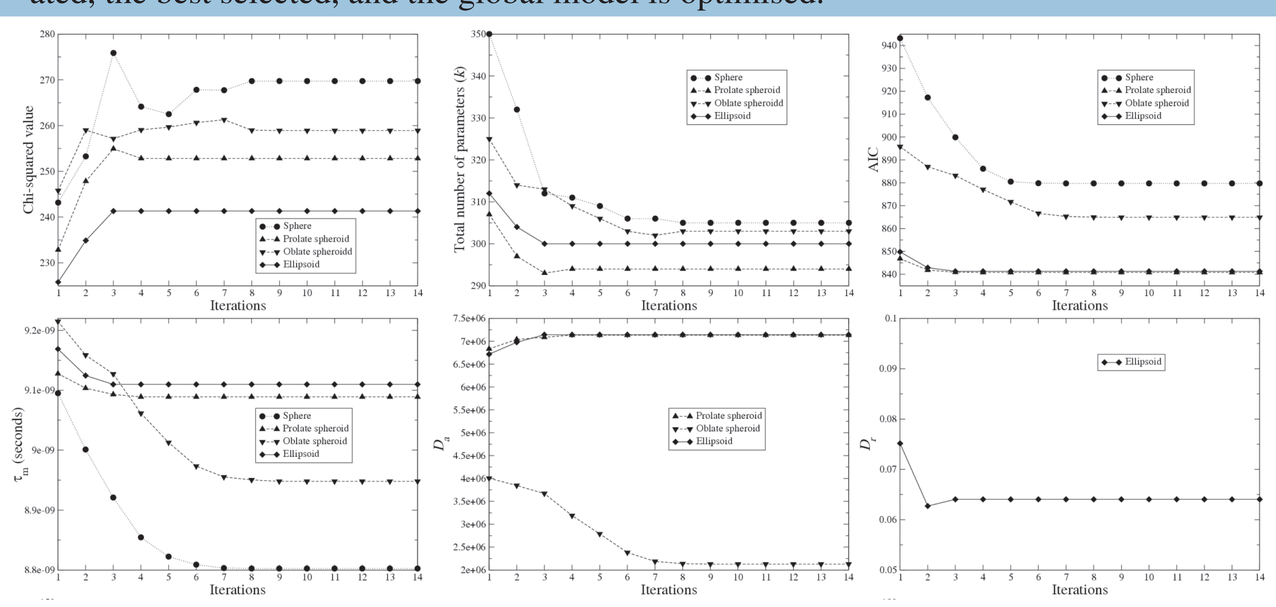


Figure 5: Global statistics and parameters for the new optimisation protocol applied to the OMP relaxation data of Gitti et al., (2005) using the X-ray structure 1F35. Each glyph corresponds to a fully minimised global model $\mathfrak{S}$.

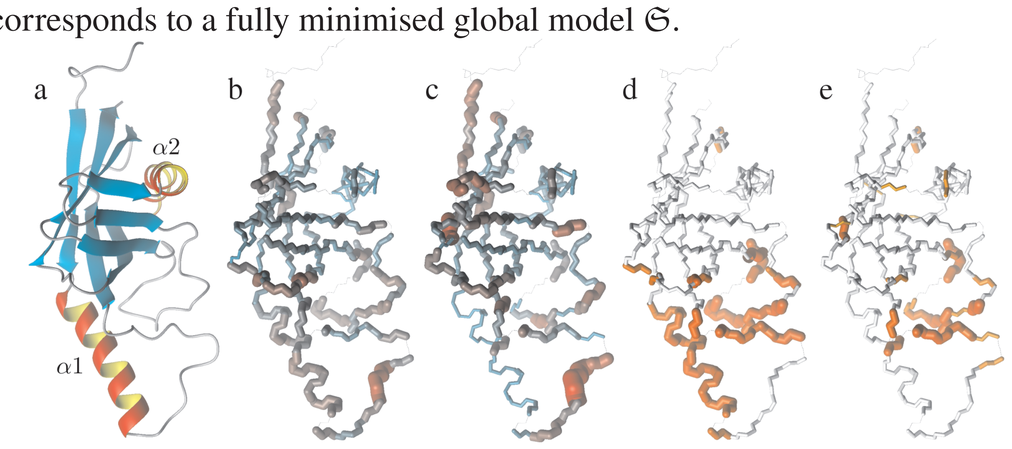


Figure 6: Comparison of the original OMP results with the reanalysis. In (b) the original and (c) the new order parameters are mapped onto the structure. The R_{ex} values of (d) the original results and (e) the reanalysis are also shown.

Gitti, R. K., Wright, N. T., Margolis, J. W., Varney, K. M., Weber, D. J., and Margolis, F. L. (2005) Backbone dynamics of the olfactory marker protein as studied by ^{15}N NMR relaxation measurements. *Biochemistry* **44**(28), 9673–9679.

relax

The program relax is new software designed for the analysis of NMR relaxation data. Written as a flexible modular collection of data analysis tools, its features include:

- Numerous optimisation algorithms.
- Numerous model selection techniques.
- Data visualisation using Molmol, Grace, or OpenDX.
- Error analysis through Monte Carlo simulation.
- Exponential relaxation curve fitting to determine the R_1 and R_2 relaxation rates.
- Calculation of the steady-state NOE.
- Reduced spectral density mapping.
- The implementation of all aspects of model-free analysis.
- GPL open source licence (freely downloadable from http://nmr-relax.com).