A toolkit for dynamics and how to relax the BMRB

Edward d'Auvergne[1*], Chris MacRaild[2*], Sébastien Morin[3*], Gary Thompson[4*], Eldon Ulrich[5]

* relax developers.
(1) Max Planck Institute for Biophysical Chemistry, Goettingen, Germany.
(2) Structural Biology Division, The Walter And Eliza Hall Institute of Medical Research, Melbourne, Australia.
(3) Department of Biochemistry and Microbiology and PROTEO, Laval University, Canada.
(4) Institute of Molecular and Cellular Biology, University of Leeds, UK.
(5) Biological Magnetic Resonance Data Bank, University of Wisconsin-Madison, US.

relax

The program relax [1, 2, 3] (http://nmr-relax.com) is designed for the study of the dynamics of proteins or other molecules through the analysis of NMR relaxation data. It is a flexible open source toolkit focused on experimental molecular dynamics, as investigated by NMR. The infrastructure built into relax can be taken and modified to suit any new analysis type. It is controlled via special user functions either by the built in python prompt or by powerful python scripting.

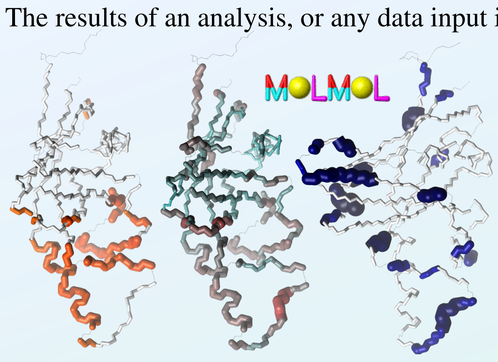
Supported NMR theories

A number of data analysis techniques are supported by relax (with others such as relaxation dispersion currently being added) including:

- Model-free analysis.
- Reduced spectral density mapping.
- Exponential curve fitting (for finding the R1 and R2 relaxation rates).
- Steady-state NOE calculation.

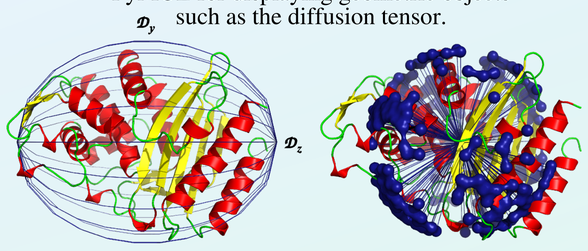
Data visualization

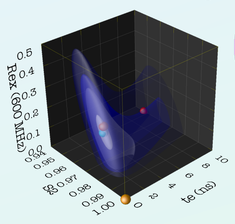
The results of an analysis, or any data input into relax, can be visualised using a number of programs:



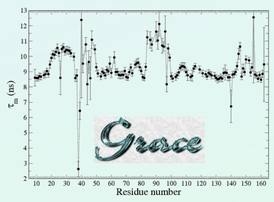
MOLMOL for 1D data representation via macros, such as S^2 values.

PyMOL for displaying geometric objects such as the diffusion tensor.





OpenDX for mapping and viewing the chi-squared space for optimisation surface visualisation.



Grace for all combinations of 2D data.

Data analysis tools

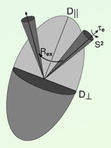
Many tools are implemented as modular components for reuse in new types of data analysis (e.g. SRLS):

- Numerous high-precision optimisation algorithms (the minfx library, https://gna.org/projects/minfx/).
- Model selection techniques for determining which model best represents the data.
- Model elimination - the removal of failed models prior to model selection.
- Monte Carlo simulations for error analysis when using non-linear fitting.

Interfacing with other programs

relax can create the input files, execute in-line, and then read the output of the following programs effectively using these as replacement model-free optimisation engines:

- Modelfree.
- Dasha.



The power of open source

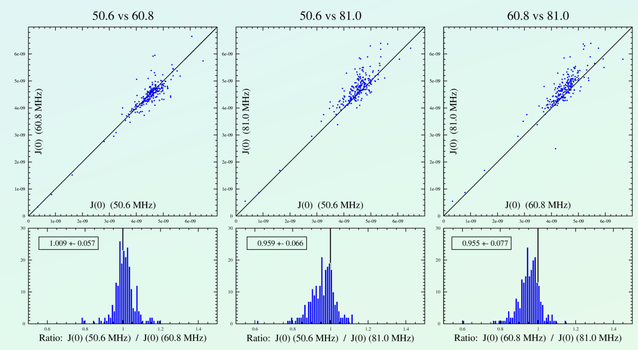
The relax source code is licenced under the GNU General Public Licence (GPL). This ensures that the program will never disappear, all resources will be permanently reachable at http://gna.org/projects/relax. It also allows everyone to modify, fix, or expand the code to suit their needs. The benefits include a solid framework and toolkit for rapid development using one of the simplest programming languages for non-coders to learn. As the program is freely available to all NMR spectroscopists on all operating systems, the new analysis will be easily accessible for others to test and use, maximizing its distribution and reach.

The 1.3 relax versions

Three years old, the advanced 1.3 relax development line adds many new abilities to the program including the support for RNA, DNA, small organic molecules, molecular complexes, and multiple spins per protein residue. The numerous improvements and additions are too great to list, but include PyMOL visualizations, data consistency testing, multi processor support, and better PDB file reading. Almost all parts of the program have been rewritten for this 1.3 line redesign and would not have been possible without the open source collaborative design, input, and coding with Chris MacRaild and the safety net of Gary Thompson's test suite framework.

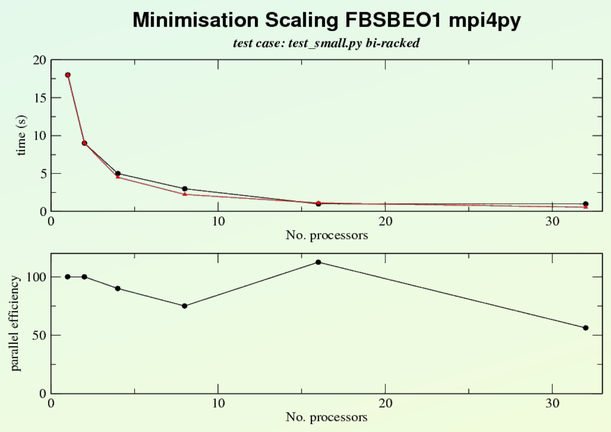
Relaxation data consistency

The open source framework has also allowed a new data analysis type to be added by Sébastien Morin [4]. This analysis checks the consistency of experimentally measured, multiple-field relaxation data, allowing the quality of the data to be assessed.



Multi-processor branch

Although not yet released, sitting within the relax source code repository is a special branch developed by Gary Thompson. This is the fully functional multi-processor code allowing relax to execute on clusters through the MPI messaging interface. The figure to the right demonstrates the scaling efficiency where the red line is the expected run times with perfect scaling efficiency and the black is relax's scaling efficiency.

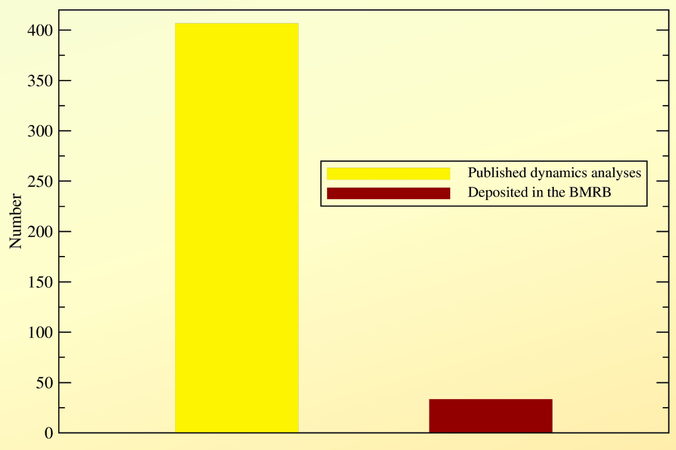


The CST branch

Still a work in progress, another branch is being developed by Pavel Kadeřávek and Petr Novák for handling the non-colinearity of the rhombic chemical shift tensor and the multiple dipolar interaction mechanisms in RNA and DNA.

Making dynamics useful - data deposition

A real problem with dynamics data from NMR, specifically fast pico to nanosecond motions, is understanding the link between motions and molecular function. One cause of this problem is an absence of access to basic data preventing further analysis to be performed. In contrast to dynamics, the utility of chemical shifts and macromolecular 3D structures is assured by their presence within the BMRB and PDB databases respectively. Without this required deposition, published structures and shifts would practically be simply pretty pictures. Compared to the 100% deposition rates for shifts and structures, the deposition of dynamics data is relatively meager:



To counter this problem, the Indiana Dynamics Database (IDD) was set up in 2000 by Assc. Prof. Martin Stone as a resource for NMR spectroscopists to deposit model-free or other dynamics data. Unfortunately this early attempt at making dynamics data accessible to researchers was shut down a number of years ago due to a lack of interest and paucity of submitted data.

Relaxing the BMRB to simplify deposition

A collaboration between the BMRB and relax is in progress with the goals of simplifying the BMRB deposition process for NMR relaxation data analyzed using relax and of improving the flexibility and completeness of the NMR-STAR data model for these kinds of data in a software independent way.

BMRB NMR-STAR dictionary v3.1

The NMR-STAR dictionary (ontology) is being redesigned on the basis of theory to better represent NMR relaxation and model-free dynamics data in a software independent way. Changes include the creation of new data categories such as 'General_relaxation' for better deposition, archiving, and distribution of new types of relaxation data, and the renaming of the old saveframe 'Order_parameter' to 'Model_free' to allow for SRLS theory and the Jens Meiler model-free order parameters of RDC data, and numerous other additions to improve the value and utility of the deposited data. These changes and additions will be released soon as v3.1 of the NMR-STAR dictionary.

The modifications to relax

A number of changes are occurring within relax to facilitate data deposition. Once a model-free analysis has been performed, creating a single file for BMRB submission containing all relaxation and model-free data will be as simple as loading the final results file into the program and then executing a single user function called bmrb.write(). Missing data, such as information about temperature control settings, etc., will be identified in the process. To support this simplicity, a new python API layer between relax and the STAR format parsing library pystarlib (http://code.google.com/p/pystarlib/) is being created to support the NMR-STAR dictionary.

With relax's support for reading Modelfree4 and Dasha output, together with support for the creation of NMR-STAR dictionary v3.1 files, deposition of relaxation and model-free data to the BMRB will become a very quick and simple operation.

[1] d'Auvergne, E. J. and Gooley, P. R. (2008). Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. J. Biomol. NMR, 40(2), 107-119.
[2] d'Auvergne, E. J. and Gooley, P. R. (2008). Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. J. Biomol. NMR, 40(2), 121-133.
[3] d'Auvergne, E. J. (2006). Protein dynamics: a study of the model-free analysis of NMR relaxation data. Ph.D. thesis, Biochemistry and Molecular Biology, University of Melbourne. (*Available in paperback form from amazon.com*).
[4] Morin, S. and Gagné, S. M. (2009). NMR dynamics of PSE-4 beta-lactamase: An interplay of ps-ns order and us-ms motions in the active site. Biophys. J. *in press*.