relax and its applications for the study of internal and domain motions in biomolecules

Edward d'Auvergne and Christian Griesinger

Department of NMR-Based Structural Biology, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany

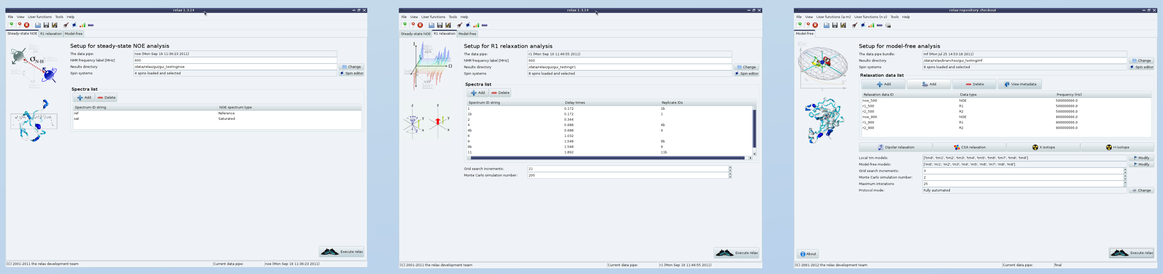
Introduction

The program relax (http://www.nmr-relax.com) is designed for molecular dynamics studies through the analysis of experimental NMR data[1,2]. It is widely used for the Lipari-Szabo model-free analysis of relaxation data, implementing a new protocol which reverses the logic of previous studies and solves many of the problems faced in the past[2,3]. The software is also used for exponential curve-fitting for determining the R_1 and R_2 relaxation rates, reduced spectral density mapping (RSDM), consistency testing of relaxation data, for dynamic studies of ensembles via the N-state model analyses, and for visualising domain motions using the new frame order theory.

Herein, the recent developments and applications of relax are demonstrated. A new major release series of relax is also presented – the 2.x versions. The first 2.0 point release was only recently published in June 2012. This is an evolution of relax from the older code base into a highly flexible toolkit and API for studying the dynamics of proteins, RNA, DNA, polysaccharides, and organic molecules using all types of NMR data.

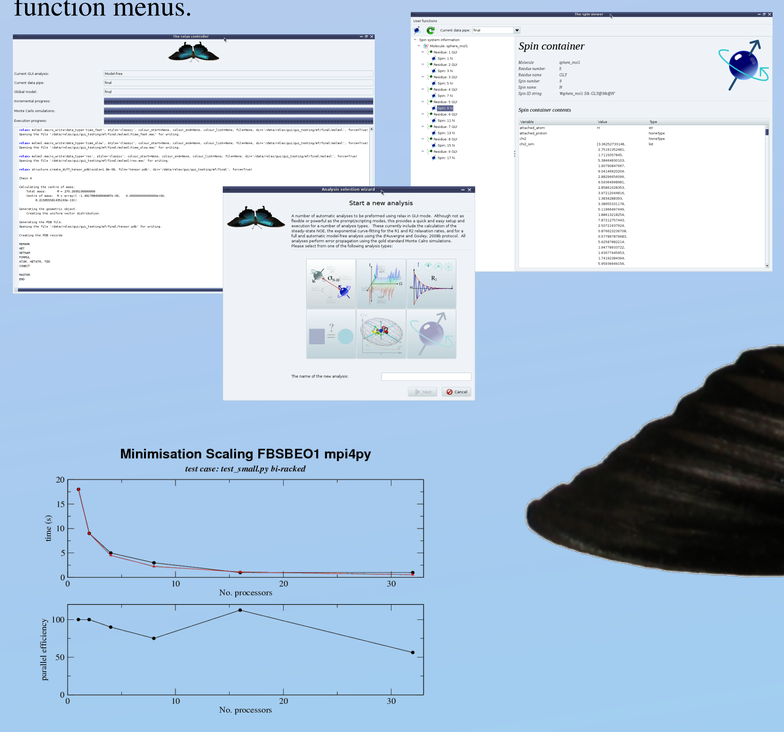
The relax toolkit design facilitates the prototyping and implementation of new dynamics analysis types based on NMR data, being a powerful alternative to computation software such as Mathematica, Matlab, and Maxima. relax has the advantage in that it has a strong focus on NMR, with import and export capabilities for most NMR data types and inbuilt handling of biological molecules. It uses numpy and scipy to provide numerous mathematical and analytical tools and minfx (http://gna.org/projects/minfx/) for optimisation. The relax core provides a NMR and dynamics-based data storage model, handling NMR data input and output, 2D graphing and molecular visualisations via PyMOL and MOLMOL. As relax is open source, licensed under the General Public Licence (GPL), everyone is free download and customise the code to their needs and can use it to implement new analyses.

Graphical User Interface



In the past, relax has been mainly used in the prompt/scripting mode. Today many new users prefer to use the Graphical User Interface (GUI), the implementation of which was lead by Michael Bieri[4]. In the new relax 2 versions, the GUI has been completely rewritten. The redesign allows most of the flexibility of the prompt and scripting modes to be available via the GUI.

The aim of the GUI is to present an easy to use, black box like interface to users. As shown in the screenshots above, it currently supports a number of relax's automatic analysis protocols including the R_1 and R_2 relaxation rate exponential curve-fitting, the steady-state NOE value and error calculation, and a modern model-free analysis protocol[2,3]. As the other analysis types (consistency testing, N-state model, RSDM, frame order) are not standardised or have black box protocols developed, these best accessed via the prompt/scripting modes, though GUI operation is still possible through the user function menus.

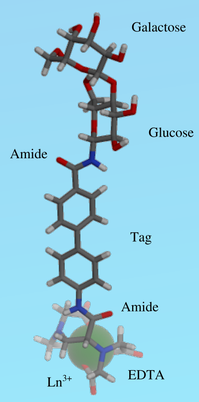


The multi-processor framework

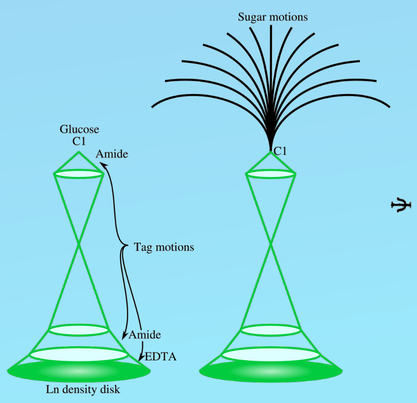
Sitting within relax is Gary Thompson's multi-processor framework. This easy to use package allows parallelised calculations within relax to be run on clusters or grids of computers. The uniprocessor and mpi4py Message Passing Interface (MPI) processor fabrics are currently supported. The figure above demonstrates the scaling efficiency for model-free analyses where the red line is the expected run times with perfect scaling efficiency and the black is relax's scaling efficiency. Access to this framework is a distinct advantage over computational packages or self-written tools and is relatively easy to hook into for the implementation of new analyses.

Lactose motions

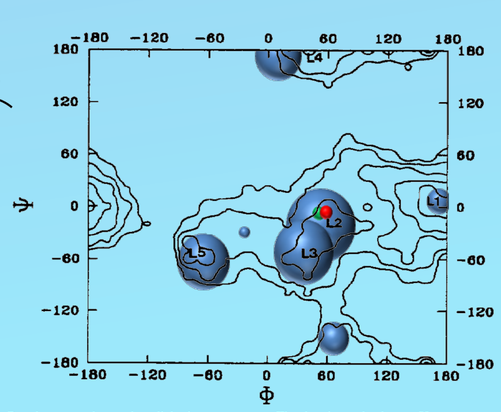
As an example of the use of the N-state model analysis in relax, the conformations and mobility of the lactose sugar molecule was studied[5]. Each conformation in the ensembles or single structures analysed represented one of the N states. The question posed was: Is the lactose glycosidic bond mobile in solution as predicted by adiabatic energy maps or in a single state in the L2 low energy region as proposed by the published X-ray[6] and NMR[7] structures? A paramagnetic lanthanide-based tag was used to internally align the lactose and PCS and RDC data was measured. Using AIC model selection[8] to compare many different ensembles and single structures, it was found that lactose samples all five low energy states in solution[5].



The tagged system.



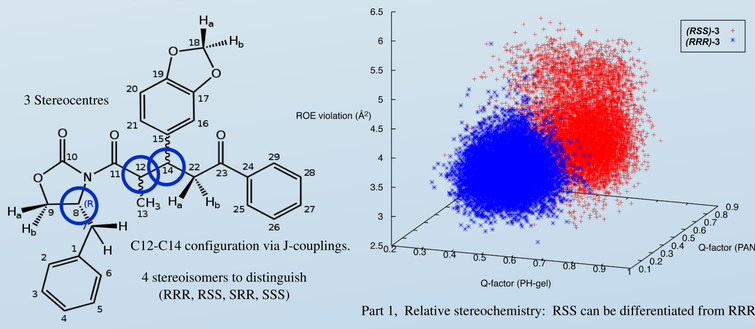
Modelling of all internal dynamics with carbon C1 as the frame of reference.



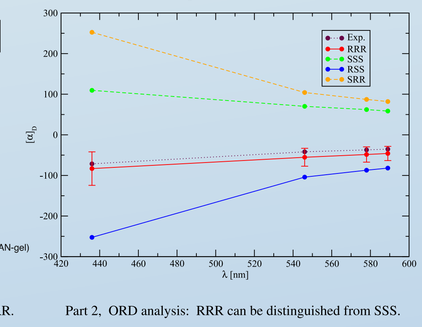
Results mapped onto the adiabatic energy map. The data is: red sphere – X-ray structures; green sphere – NMR structures; blue spheres – best ensemble, the sphere diameter is proportional to the probability of each state.

Stereochemistry

Another demonstration of the use of the N-state model analysis in relax is for determining the absolute stereochemistry of organic molecules in solution[9]. NOESY data and RDC data from external alignment is used for the principle analysis. The procedure consists of two steps. The first step is the determination of relative stereochemistry by evaluating large numbers of ensembles using the NMR data (using the N-state model in relax). The second step is to use the selected NMR ensembles from the first step together with chiroptical data to determine the absolute stereochemistry.



Part 1, Relative stereochemistry: RSS can be differentiated from RRR.



Part 2, ORD analysis: RRR can be distinguished from SSS.

relax - BMRB integration

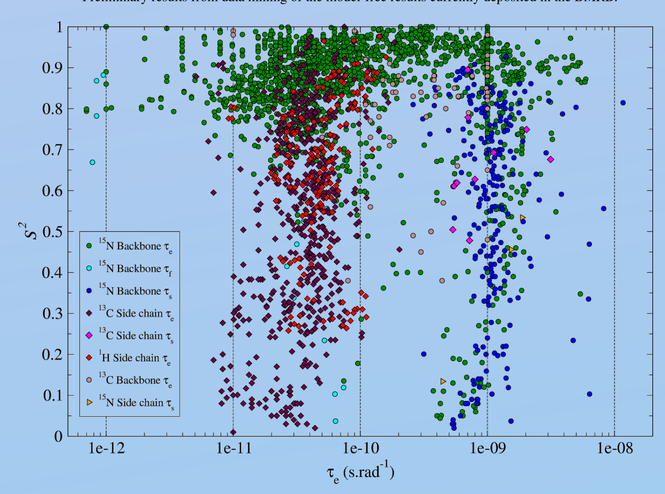
A problem with model-free analyses of NMR relaxation data is that the link between the fast picosecond to nanosecond motions and the biologically relevant functions is not well understood. A major issue inhibiting this understanding is the lack of access to the basic data. This is in sharp contrast to chemical shifts and 3D structures. Access to base data would open up avenues for new perspectives and understandings of macromolecular motions.

To ameliorate the problem of the paucity of available dynamics data, reading and writing of NMR-STAR files has been implemented in relax. This means that relax can be used to read old Modelfree4, Dasha and relax results files from published model-free analyses for easy conversion to BMRB format. The resultant NMR-STAR file can be deposited via the standard ADIT-NMR interface.

In addition, relax's ability to read the NMR-STAR data opens up interesting opportunities for data mining. In the figure to the bottom right, for example, a distinctive pattern of dynamics is clearly visible in the plot of the order parameter S^2 verses the internal correlation times τ_e. Such a plot of model-free data clearly shows, in a model-independent way, distinct modes of motion throughout the backbone and sidechains of proteins.

Comparison of the number of published results verses those deposited in the BMRB.

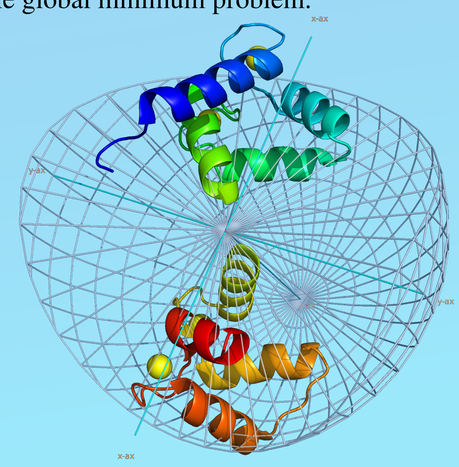
Preliminary results from data mining of the model-free results currently deposited in the BMRB.



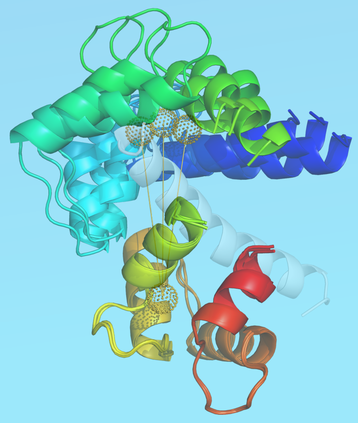
Frame Order

The Frame Order theory is a new, yet to be published analysis for determining the inter-domain dynamics of a multi-domain molecular system. It uses RDC and PCS data from all spins via the internal alignment of a single domain. Its aim it to convert the data into a statistical mechanical model of the motion. By modelling the domain motions via the ball and socket joint, also known as spherical joint, the theory translates the NMR data into a number of simple models. This includes simple rotors with a restricted single torsional mode of motion about an axis, pivoted motion in a straight line, isotropic cone motions (with a restricted torsional motion, torsionless, or free-rotor motions), and pseudo-elliptic cone motions (again with the torsion restricted, torsionless, or free).

This theory is implemented in relax, though the use of PCS data is currently a work in progress. Preliminary results for the calmodulin (CaM) signalling molecule using RDC data is shown below. From preliminary testing it can be seen that RDCs contain mainly information about the amplitude of motions and are relatively insensitive to the direction or orientation of the modes of motion. In contrast the PCS contains orientational information but barely any amplitude information. Combining the two significantly improves the dynamic picture from the frame order modelling. And the addition of the PCS converts the problem from a multi-local minima problem with thousands of minima into a single global minimum problem.



The preliminary frame order results for free CaM using published RDC data[10]. The model chosen is the pseudo-ellipse. The amplitudes of motion are comparable with those previously published. The directions of the motions need to be refined using PCS data.



The domain motions of the CaM-IQ peptide complex. The three structures from the crystal[11] are superimposed showing the linear motion sweeping over the target peptide (grey) of ~10 degrees. The preliminary frame order results indicate that the motion is that of a rotor, as the crystal structure suggests, and that the motion is also in the same direction. However the amplitude of motion from the RDC data is about 1.25 times greater than that seen in the crystal.

References

[1] d'Auvergne and Gooley, *JBNMR* 2008, **40**, 107
[2] d'Auvergne and Gooley, *JBNMR* 2008, **40**, 121
[3] d'Auvergne and Gooley, Mol. BioSyst. 2007, 3, 483
[4] Bieri, d'Auvergne and Gooley, *JBNMR* 2011, **50**, 147
[5] Erdélyi, d'Auvergne, Navarro-Vázquez, Leonov, Griesinger, Chem. Eur. J. 2011 17, 9368
[6] Jimenez-Barbero et al., *JACS* 2005, **127**, 3589.
[7] Noordick et al., *Krystallogr.* 1984, **168**, 59.
[8] d'Auvergne and Gooley, *JBNMR*, 2003, **25**, 25.
[9] Sun, d'Auvergne, Reinscheid, Dias, Andrade, Rocha, Griesinger, *Chem. Eur. J.* 2011, **17**, 1811.
[10] Bertini et al., *Proc. Natl. Acad. Sci. U.S.A.* 2004, **101**, 6841.
[11] Petegem et al., *Nat. Struct. Mol. Biol.*, 2005, **12**, 1108.