relax: the analysis of biomolecular kinetics and thermodynamics using NMR relaxation dispersion data

Morin, Sebastien; Linnet, Troels Emtekær; Lescanne, Mathilde; Schanda, Paul; Thompson, Gary S.; Tollinger, Martin; Teilum, Kaare; Gagne, Stephane; Marion, Dominique; Griesinger, Christian; Blackledge, Martin; d’Auvergne, Edward J.

Published in:
Bioinformatics

DOI:
10.1093/bioinformatics/btu166

Publication date:
2014

Document version
Early version, also known as pre-print

Citation for published version (APA):
relax: the analysis of biomolecular kinetics and thermodynamics using NMR relaxation dispersion data

Sébastien Morin, Troels E. Linnet, Mathilde Lescanne, Paul Schanda, Gary S. Thompson, Martin Tollinger, Kaare Teilm, Stéphane Gagné, Dominique Marion, Christian Griesinger, Martin Blackledge, and Edward J. d’Auvergne

1PROTEO, Université Laval, Québec G1V 0A6, Canada. 2International AIDS Society HQ, CH-1202 Geneva, Switzerland. 3Dept. of Biology, University of Copenhagen, DK-2200, Denmark. 4Institut de Biologie Structurale, Grenoble F-38027, France. 5Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK. 6Institute of Organic Chemistry & CMBI, University of Innsbruck, A-6020, Austria. 7NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany.

ABSTRACT

Nuclear Magnetic Resonance (NMR) is a powerful tool for observing the motion of biomolecules at the atomic level. One technique, the analysis of relaxation dispersion phenomenon, is highly suited for studying the kinetics and thermodynamics of biological processes. Built on top of the relax computational environment for NMR dynamics is a new dispersion analysis designed to be comprehensive, accurate and easy to use. The software supports more models, both numeric and analytic, than current solutions. An automated protocol, available for scripting and driving the GUI, is designed to simplify the analysis of dispersion data for NMR spectroscopists. Decreases in optimisation time are granted by parallelisation for running on computer clusters and by skipping an initial grid search by using parameters from one solution as the starting point for another – using analytic model results for the numeric models, taking advantage of model nesting, and using averaged non-clustered results for the clustered analysis.

Availability: The software relax is written in Python with C modules and is released under the GPLv3+ licence. Source code and precompiled binaries for all major operating systems are available from http://www.nmr-relax.com.

Contact: edward@nmr-relax.com

Biological macromolecules are intricate machines and their functions are closely related to their motions. These motions can be studied experimentally at the atomic level by NMR spectroscopy. Many important biological processes occur on the ms to ms timescale and, for atoms exchanging between different states, NMR relaxation dispersion can be observed. By studying this exchange process, kinetic and thermodynamic information can be obtained.

For exchanging atoms, their nuclear spin magnetisation is described by the Bloch-McConnell equations (McConnell, 1958). Using experimental data the solution to these equations reveals both populations of the molecular states (thermodynamics) and rates of exchange between them (kinetics). Though the general solution valid for all motions remains intractable, analytic solutions with restricted motions are available and are frequently used. The equations can also be solved numerically.

Two NMR dispersion methods are used for analysing motions: single, zero, double, or multiple quantum (SQ, ZQ, DQ, MQ) CPMG (Carr and Purcell, 1954; Meiboom and Gill, 1958); or R1,ρ (Deverell et al., 1970). Combined SQ, ZQ, DQ, and MQ data will be labelled as multiple-MQ or MMQ data. Various models are used to analyse different data and motions. The simplest is that of no motion (No Rex). For SQ CPMG-type experiments, analytic models include the original Luz and Meiboom (1963) multiple-site fast exchange models (LM63), the Carver and Richards (1972) and population-skewed Ishima and Torchia (1999) 2-site models for most time scales (CR72, IT99), and the Tollinger et al. (2001) 2-site very slow exchange model (TSMFK01). The CR72 model has been extended by Korzhnev et al. (2004) for MMQ data. For R1,ρ-type data analytic equations include the Meiboom (1961) 2-site fast exchange model for on-resonance data (M61), extended by Davis et al. (1994) to off-resonance data (DPL94), and the Trott and Palmer (2002) and Miloushev and Palmer (2005) 2-site models for non-fast and all time scales (TP02, MP05). Different numeric solutions (NS) can be designed for SQ or MMQ data.

Diverse software solutions exist for analysing relaxation dispersion data including CPMGFit (http://www.palmer.hs.columbia.edu/software/cpgfit.html), cpmgfit (available upon request from Dmitry Korzhnev), CATIA (Hansen et al., 2008), Nessy (Bieri and Gooley, 2011), Guardd (Kleckner and Foster, 2012), ShereKhan (Mazur et al., 2013), and Glove (Sugase et al., 2013). The software relax (d’Auvergne and Gooley, 2008) is a platform for studying molecular dynamics using experimental NMR data and can be used as a numerical computing environment. Herein support for relaxation dispersion within relax is presented. Distributed as part of
Table 1. Comparison of model support for different dispersion software

<table>
<thead>
<tr>
<th></th>
<th>CPMG-type</th>
<th>( R_{1,\rho} )-type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No-Rex</td>
<td>DL63 2-site</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DL63 3-site</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TSMFK01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS SQ 2-site</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMQ CR72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMQ MMQ 2-site</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMQ MMQ 3-site</td>
</tr>
<tr>
<td>CPMGFit</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>cpmg_fit</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CATIA</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NESSY</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>GUARDD</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ShereKhan</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>GLOVE relax</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

This table shows the number of dispersion models supported by relax, with the most comprehensive package supporting the greatest number of dispersion models and NMR data types. The number of dispersion models supported by relax is extensive (Table 1). This allows for detailed comparisons between modern numeric and traditional analytic approaches. Different user interfaces (UIs) can be employed to analyse dispersion data including the prompt, scripting, and graphical user interface (GUI). The scripting UI enables the greatest flexibility and allows for most analysis protocols to be replicated. By implementing a novel automated analysis and providing an easy to use GUI based on this auto-analysis, the study of dispersion data is much simplified.

The setup of the auto-analysis includes defining the molecular system, loading the dispersion data directly from peak lists, clustering atoms with the same kinetics, modifying the list of dispersion models, and setting up Monte Carlo (MC) simulations for error propagation. Execution involves sequential optimisation of the models, fixed model elimination rules to remove failed models and failed MC simulations increasing both parameter reliability and accuracy (d’Auvergne and Gooley, 2006), and a final run whereby AIC model selection is used to judge statistical significance (Akaike, 1973; d’Auvergne and Gooley, 2003). The optimisation is designed for absolute accuracy and robustness but, as this can take time, it has been parallelised at the spin cluster and MC simulation level to run on computer clusters using OpenMPI. Three additional methods are used to speed up calculations, all designed to skip the computationally expensive grid search. The first is model nesting – the more complex model starts with the optimised parameters of the simpler. The second is model equivalence – when two models have the same parameters. For example the CR72 model parameters are used as the starting point for the CPMG numeric models resulting in a huge computational win. The third is for spin clustering – the analysis starts with the averaged parameter values from a completed non-clustered analysis.

The dispersion analysis in relax is implemented in Python using NumPy and the GUI using wxPython. Optimisation using the Nelder-Mead simplex and log-barrier constraint algorithms from the minfx library (https://gna.org/projects/minfx/) removes the need for numerical gradient approximations which add a second numeric layer to the NS models. Data visualisation is via the software Grace.

ACKNOWLEDGEMENT

Nikolai Skrynnikov is thanked for his generous feedback and code contributions for implementing many of the numeric models and Flemming Hansen and Dmitry Korzhnev for kindly providing their software and published dispersion data.

REFERENCES


