Journal of Biomolecular NMR, **16:** 23–28, 2000.
KLUWER/ESCOM

Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data

Patrice Dosset, Jean-Christophe Hus, Martin Blackledge* & Dominique Marion
Institut de Biologie Structurale – Jean-Pierre Ebel, C.N.R.S.-C.E.A., 41 rue Jules Horowitz, F-38027 Grenoble Cedex, France

Received 5 October 1999; Accepted 12 November 1999

Key words: anisotropic tumbling, heteronuclear relaxation, minimization, rotational diffusion, simulated annealing

Abstract

A novel program has been developed for the interpretation of ^{15}N relaxation rates in terms of macromolecular anisotropic rotational diffusion. The program is based on a highly efficient simulated annealing/minimization algorithm, designed specifically to search the parametric space described by the isotropic, axially symmetric and fully anisotropic rotational diffusion tensor models. The high efficiency of this algorithm allows extensive noise-based Monte Carlo error analysis. Relevant statistical tests are systematically applied to provide confidence limits for the proposed tensorial models. The program is illustrated here using the example of the cytochrome c′ from *Rhodobacter capsulatus*, a four-helix bundle heme protein, for which data at three different field strengths were independently analysed and compared.

Introduction

Heteronuclear relaxation is now widely used to describe the backbone dynamics of proteins (Palmer 1997; Kay, 1998). In the most common analysis the angular reorientation of the dominant spin relaxation interactions is sampled by measuring R_1 and R_2 autorelaxation rates and the heteronuclear ^{1}H-^{15}N NOE at each available NH pair in the molecule. For folded globular proteins a general method of interpretation has been adopted which parametrizes all internal angular motion, irrespective of its true geometric character, in terms of a generalized amplitude, S^2, and a characteristic internal correlation time τ_i (Lipari and Szabo, 1982a,b). If necessary, further parameters can be introduced either by evoking internal motion on two timescales (Clore et al. 1990), or by incorporating slow motion on the sub-millisecond timescale due to chemical or conformational exchange.

The time autocorrelation function employed to describe angular reorientation using this model is affected by both internal and overall motion, implying that both contributions require detailed attention to accurately reproduce the measured relaxation rates. In the simplest analysis, overall motion is assumed to be isotropic and this contribution is described by a single exponential. In the case of rotational diffusion anisotropy the time correlation is a multiexponential function, whose relative weighting depends on the orientation of the different relaxation interactions with respect to the principal components of the diffusion tensor. The exact relationship between measured relaxation rates and the time correlation and spectral density functions for the different motional models can be found elsewhere (Woessner, 1962; Tjandra et al., 1995).

Not surprisingly, the interpretation of heteronuclear relaxation using the model-free approach assuming isotropic overall motion can artificially evoke fictive contributions to the local internal motion if the rotational diffusion tensor is actually significantly anisotropic (Schurr et al., 1994; Cordier et al., 1998). The determination of the rotational diffusion tensor is therefore a prerequisite in the analysis of internal dynamics using heteronuclear relaxation – once

*To whom correspondence should be addressed. E-mail: martin@rmn.ibs.fr

rotational diffusion has been characterized, the derived tensor can then be incorporated into a model-free analysis of local motion. The determination of the overall rotational diffusion tensor can also provide important information concerning the overall form of globular molecules, and the relative domain orientation in modular proteins (Brüschweiler et al., 1995) or different components in strongly interacting biomolecular complexes (Tjandra et al., 1997).

It has recently been shown that rotational diffusion can be accurately characterized using the relaxation rate ratio R_2/R_1 (Tjandra et al., 1995; Lee et al., 1998), a parameter which becomes independent of internal dynamics in the fast motion limit while remaining highly sensitive to overall diffusion. Here we present TENSOR, a program designed for the efficient determination of the rotational diffusion tensor from three-dimensional structural coordinates and ^{15}N relaxation data using a simulated annealing algorithm. A graphical interface allows the examination of the orientation of tensorial components with respect to the three-dimensional coordinates of the molecule, and provides a diagrammatic representation of the statistical behaviour of the system.

Methods

The diffusion parameters are extracted by minimizing the target function:

$$\chi^2 = \sum_n \left\{ \left[\left(R_2^{meas}/R_1^{meas} \right) - \left(R_2^{calc}/R_1^{calc} \right) \right]_n / \sigma_n \right\}^2 \quad (1)$$

where σ is the uncertainty in the experimental R_2/R_1 ratio. The parametric space describing the orientation and amplitude of the tensorial components requires the sampling of 4 $\{D_\perp, D_\parallel, \theta, \phi\}$ and 6 $\{D_{xx}, D_{yy}, D_{zz}, \alpha, \beta, \gamma\}$ dimensional space for the axially symmetric and fully anisotropic descriptions.

The use of the simpler axially symmetric model in the case of real rhombicity results in the presence of two orthogonal minima, describing prolate and oblate approximations to the real tensor (Blackledge et al., 1998). Depending on the values of θ and ϕ, equally valid minima can be steep-sided in an otherwise flat χ^2 space (the value corresponding to the isotropic approximation) and therefore difficult to localise using a grid-search type of minimisation algorithm, or broad and consequently much easier to find. The values of θ and ϕ are of course dependent on the reference frame, usually taken from the database coordinates, and therefore completely arbitrary. This makes the identification of appropriate models difficult and time-consuming if simple minimization methods are used.

The program uses a highly efficient algorithm, specifically designed to search the parametric space defining the particular tensorial models by random variation of the 4 or 6 parameters. The principal difference between this algorithm and a classical simulated annealing (Metropolis et al., 1953) is that, once a plateau χ^2 level is reached, the temperature is regulated to extensively sample this level until significant minima, corresponding to the more complex tensorial models, are found. This temperature regulation is achieved using a fuzzy logic algorithm (Leondes, 1997). Once a significant minimum is found, a Levenberg–Marquardt minimization (Press et al., 1988) is used to avoid inefficient sampling and the relevant transformations are performed to place the minimum within the reference frame ($D_{xx} < D_{yy} < D_{zz}$, $-\pi < \alpha, \beta, \gamma < +\pi$). Model R_2/R_1 values are then calculated from this minimum and used as the centre of the noise simulations for the Monte Carlo error analysis which are fitted in an identical manner.

In order to reduce the possibility of contributions from internal motion in this analysis, it is important to select residues whose interaction vectors are likely to be in the librational motional limit. This is achieved either by analysis of the actual relaxation rates with respect to the mean values for the whole molecule (Tjandra et al., 1995), or in our case by identifying those residues present in secondary structural elements which are assumed to be similarly rigid.

While the fully anisotropic tensorial description is clearly the most generally applicable, care must be taken not to evoke a more complex model than the data allow. To avoid overinterpretation, the experimental χ^2 is compared to a χ^2 distribution of fits to simulated datasets from Monte Carlo sampling of Gaussian distributions. This allows the confidence level of the proposed models to be determined. The significance of the improved fit obtained from the more complex model is also calculated from the random statistical χ^2 improvement gained by fitting the noise simulations from the best axially symmetric tensor to a fully anisotropic model. Distinction between the different diffusion models is achieved by calculating the function F to test the justification of including more parameters into the fit (Cordier et al., 1998). If the F

Table 1a. Diffusion tensor parameters from R_2/R_1 data at 800 MHz (63 vectors)

Tensor[a]	D_{xx}[b] ($10^7 s^{-1}$)	D_{yy} ($10^7 s^{-1}$)	D_{zz} ($10^7 s^{-1}$)	α[c] (deg)	β/ϕ (deg)	γ/θ (deg)	χ^2_{exp}	$\chi^2_{0.05}$[d]
Oblate	1.22 ± 0.05	1.90 ± 0.05	1.90 ± 0.05	–	38.0 ± 4.7	44.8 ± 4.5	64.0	77.4
Prolate	1.52 ± 0.01	1.52 ± 0.01	1.96 ± 0.06	–	-74.3 ± 3.7	-35.1 ± 3.3	71.6	77.1
Asymm	1.27 ± 0.05	1.81 ± 0.06	2.01 ± 0.07	-50.7 ± 7.3	-73.7 ± 7.3	-23.9 ± 10.8	55.5	74.4

[a]Data are fitted to the crystal structure (pdb code 1nbb; Tahirov et al., 1996). All of the available NH relaxation rates from helical residues were used in the fit. Parameter uncertainties are taken from mean and standard deviations of 1000 Monte Carlo simulations. The isotropic model gives $\tau_c = (10.56 \pm 0.03)$ ns with $\chi^2_{exp} = 119$ compared to $\chi^2_{0.05} = 80$ (see footnote d), so this model is rejected.
[b]The fitted parameters $D_\perp$ and $D_\parallel$ are shown here as D_{xx}, D_{yy} and D_{zz} for direct comparison with the generally asymmetric diffusion tensor.
[c]Angles (α, β, γ) describe the orientations of **D** in the NMR structure frame. $D_{xx} < D_{yy} < D_{zz}$. For axial symmetry the polar angles θ and ϕ are given.
[d]$\chi^2_{0.05}$ refers to the ($\alpha = 0.05$) confidence limit for the fit derived from 1000 Monte Carlo simulations.

Table 1b. Tensor analysis from 600 MHz data (64 vectors)

Tensor[a]	D_{xx}[b] ($10^7 s^{-1}$)	D_{yy} ($10^7 s^{-1}$)	D_{zz} ($10^7 s^{-1}$)	α[c] (deg)	β/ϕ (deg)	γ/θ (deg)	χ^2_{exp}	$\chi^2_{0.05}$[d]
Prolate	1.43 ± 0.01	1.43 ± 0.01	2.03 ± 0.04	–	-77.5 ± 1.7	-38.1 ± 2.0	77.2	77.8
Asymm	1.30 ± 0.04	1.60 ± 0.04	2.05 ± 0.04	-20.7 ± 10.1	-77.8 ± 2.2	-34.8 ± 1.9	69.3	75.2

[a]The isotropic model gives $\tau_c = (11.05 \pm 0.03)$ ns with $\chi^2_{exp} = 281$ compared to $\chi^2_{0.05} = 82$, so this model is rejected.
[b–d]See Table 1a.

Table 1c. Tensor analysis from 400 MHz data (49 vectors)

Tensor[a]	D_{xx}[b] ($10^7 s^{-1}$)	D_{yy} ($10^7 s^{-1}$)	D_{zz} ($10^7 s^{-1}$)	α[c] (deg)	β/ϕ (deg)	γ/θ (deg)	χ^2_{exp}	$\chi^2_{0.05}$[d]
Prolate	1.45 ± 0.26	1.45 ± 0.26	1.95 ± 0.27	–	-77.2 ± 7.5	-30.9 ± 6.2	52.1	61.8
Asymm	1.31 ± 0.22	1.60 ± 0.21	2.05 ± 0.52	-33.8 ± 24.8	-78.1 ± 8.85	-28.2 ± 17.5	50.4	58.9

[a]The isotropic model gives $\tau_c = (11.10 \pm 0.04)$ ns with $\chi^2_{exp} = 92$ compared to $\chi^2_{0.05} = 62$ so is rejected.
[b–d]See Table 1a.

Table 1d. F statistic calculated from datasets using 6-parameter fitting of data simulated from 4-parameter model

λ	F[a]		
	800 MHz	600 MHz	400 MHz
Simulated			
0.1	2.31	2.34	2.44
Experimental	4.31	3.30	0.70

[a]F-test values between the best 4-parameter minima and the 6-parameter model are shown.

statistic is greater than a given threshold, the improved model is statistically relevant. This may also be expressed as a probability p that the improvement is due to random fluctuation. Uncertainties in the parameterisation of the diffusion tensor are also determined using these Monte Carlo simulations, based on the experimental error in the relaxation data. Both χ^2 and F distributions calculated by TENSOR can be inspected graphically (Figure 1b).

Dipolar and chemical shift anisotropy interactions are often assumed to be collinear, an approximation which is certainly not generally applicable, but which simplifies the analysis considerably. Following a recent investigation of the effects of noncollinearity on relaxation measurements (Fushman & Cowburn, 1999), we have incorporated the possibility of changing this angle from its default value of 0°. For simplicity, we assume that the same angle prevails for each residue (or that an average interaction angle is used) and that the ^{15}N CSA tensor is axially symmetric and in the C_{i-1}-N_iH_i plane. The internal motion

a

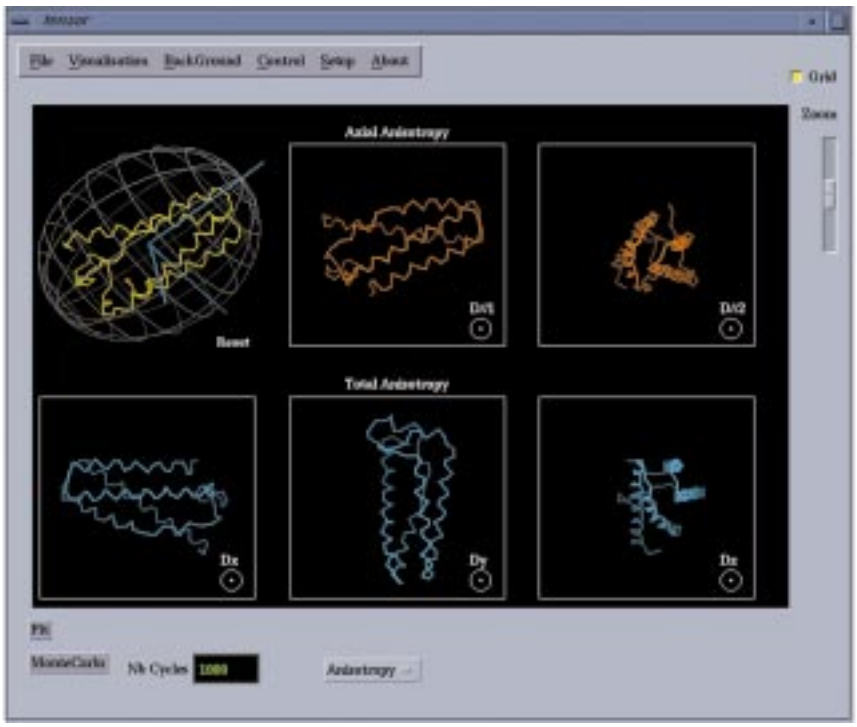

Figure 1. 1. Representation of tensor orientation and relative component amplitude determined from ^{15}N relaxation data using the progam TENSOR. (a) 800 MHz relaxation data. The top left-hand panel shows the rotational diffusion tensor as an ellipse. The principal components of the tensor are drawn with respect to the coordinates of the N atoms. This representation may be manipulated in three dimensions to examine the orientation and amplitude of the components relative to the molecule. The two panels to the right show the molecule oriented so that the unique axes of the axially symmetric models are along the z-axis (out of the page). The three lower panels show the molecule oriented so that the principal axes of the tensor are along the z-axis. (b) Graphical representation of the statistical analysis performed by TENSOR. Top and bottom left-hand panels: χ^2 distribution based on 1000 Monte Carlo noise simulations for the axially symmetric models. The x-axes show the χ^2 value, the y-axes the relevant population. Orange and blue lines represent the 0.05 and 0.10 confidence limits calculated from the distributions and the yellow lines indicate the experimental χ^2. The top right-hand panel shows the χ^2 distribution for the 6-parameter fit of the fully anisotropic model. The bottom right-hand panel shows the F-distribution calculated for the random statistical improvement obtained when simulations derived from the best axially symmetric model are fit with the 6-parameter model. The experimental F is again shown in yellow.

was assumed to be identical for the two interactions, a hypothesis which is certainly not justified, even in the case of crankshaft-only motions, but one which only weights the relative significance of the dipolar and csa dependent contributions which are in any case defined by the experimentally ill-defined interaction constants ($\sigma_{\parallel}$-$\sigma_{\perp}$) and ($1/r^6_{NH}$).

Results and discussion

As an example, the analysis of the rotational diffusion tensor of the cytochrome c′ from *Rhodobacter capsulatus*, a four-helix bundle heme protein, has been performed using three independent relaxation data sets collected at 400 MHz, 600 MHz and 800 MHz ^1H frequency. Longitudinal and transverse relaxation rates were determined as previously described (Farrow et al., 1994; Cordier et al., 1998); using mixing times of (10, 30, 50, 70, 90, 110, 130, 150, 210, 350) ms, (10, 20, 30, 50, 70, 90, 130, 150, 200, 290) ms and (5, 15, 30, 50, 70, 90, 130, 170, 250) ms for R_2 and (10, 40, 80, 120, 200, 360, 500, 900, 1300, 1800) ms, (10, 40, 80, 120, 240, 400, 800, 1200, 1600, 2500) ms and (10, 60, 100, 180, 360, 660, 800, 1200, 1760, 3000) ms for R_1 for 400 MHz, 600 MHz and 800 MHz, respectively. The experimental relaxation values are available from the authors as Supplementary material. Table 1 summarises the parametrization of the tensor as determined using the data from each field separately. The eigenvalues and orientations found for the three datasets are in general very similar. Nevertheless, the precision of the data varies quite significantly due to available signal-to-noise and spectrometer performance (only 49 helical

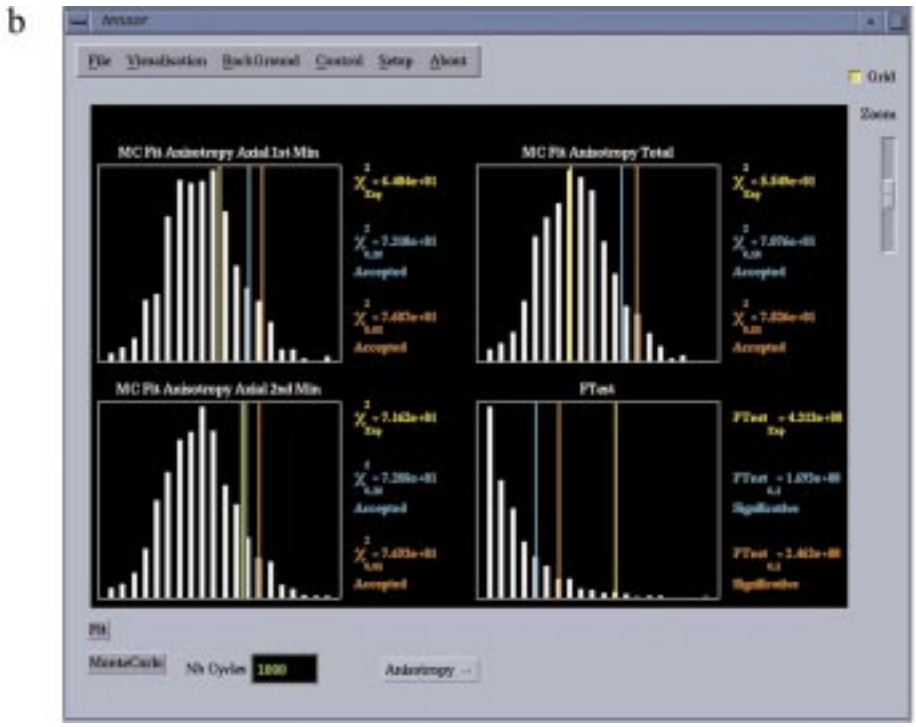

Figure 1. (continued).

R_1,R_2 pairs are available for the 400 MHz data), which clearly has implications for the confidence levels of the various tensor descriptions. In the case of the data measured at 800 MHz, both axially symmetric tensorial models are significant, implying real rhombicity. This is reflected in the statistical improvement ($F_{exp} > F_{0.1}$) of the fully anisotropic model compared to these two models. For the 600 MHz dataset the improvement is also significant, while for the 400 MHz data, the fully anisotropic model has an acceptable χ^2, but the improvement compared to the axially symmetric model is not significant. In general, however, the tensor descriptions from the three analyses are very similar. Two of the angles defining the orientations are very closely reproduced, while the third (the α Euler angle) and the intermediate tensor component D_{yy} are statistically different between the 800 MHz data and one of the other datasets. These differences may be due to the poor sampling of angular space available for this molecule, where NH vectors taken into consideration are present in four nearly parallel helices, or to small temperature variations between experiments performed on the three spectrometers. A detailed analysis of the structure and dynamics of this molecule will be presented elsewhere.

A fit of each of the three anisotropic models for one dataset containing 63 R_2/R_1 ratios takes a total of 7 s on a single processor R10000 SGI 180 MHz instrument. All Monte Carlo simulations for the three models and 4×1000 subsequent fits took 120 min on the same machine.

Conclusions

In conclusion, a highly efficient and robust program, capable of rapidly finding the necessary parametric description of the rotational diffusion tensor from ^{15}N relaxation, is now available from our laboratory. Uncertainties in the parameters and confidence levels in the different tensorial descriptions are determined using extensive Monte Carlo simulations based on the experimentally derived error in the relaxation rates. The program is demonstrated here using three independent datasets determined at different field strengths for the cytochrome c′ from *Rhodobacter capsulatus.*

Software availability

TENSOR is currently available from our website at (http://www.ibs.fr/ext/labos/LRMN/welcome_en.htm).

Acknowledgements

This work was supported by the Commisariat à l'Energie Atomique and the Centre National de la Recherche Scientifique. This is publication number 738 of the Institut de Biologie Structurale.

References

Blackledge, M., Cordier, F., Dosset, P. and Marion, D. (1998) *J. Am. Chem. Soc.*, **120**, 4538–4539.

Brüschweiler, R., Liao, X. and Wright, P. (1995) *Science*, **268**, 886–889.

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.

Cordier, F., Caffrey, M., Brutscher, B., Cusanovich, M., Marion, D. and Blackledge, M. (1998) *J. Mol. Biol.*, **281**, 341–361.

Farrow, N.A., Muhandiram, R., Singer, A.U., 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.

Fushman, D. and Cowburn, D. (1999) *J. Biomol. NMR*, **13**, 139–147.

Kay, L.E. (1998) *Nat. Struct. Biol.*, **5**, 513–517.

Lee, L.K., Rance, M., Chazin, W.J. and Palmer III, A.G. (1997) *J. Biomol. NMR*, **9**, 287–298.

Leondes, C.T. (1997) *Fuzzy Logic and Expert Systems Applications*, Academic Press, San Diego, CA.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.*, **104**, 4546–4558.

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

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) *J. Chem. Phys.*, **21**, 1087–1094.

Palmer III, A.G. (1997) *Curr. Opin. Struct. Biol.*, **7**, 732–737.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988) *Numerical Recipes in C, The Art of Scientific Computing*, Cambridge University Press, Cambridge.

Schurr, J.M., Babcock, H.P. and Fujimoto, B.S. (1994) *J. Magn. Reson.*, **B105**, 211–224.

Tahirov, T.H., Shintaro, M., Meyer, T.E., Cusanovich, M., Higuchi, Y. and Yasuako, N. (1996) *J. Mol. Biol.*, **259**, 467–479.

Tjandra, N., Feller, S.E., Pastor, R.W. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 12562–12566.

Tjandra, N., Garrett, D.S., Gronenborn, A.M., Bax, A. and Clore, G.M. (1997) *Nat. Struct. Biol.*, **4**, 443–449.

Woessner, D.E. (1962) *J. Chem. Phys.*, **37**, 647–654.