Appl. Magn. Reson. 9, 581–588 (1995)

Applied Magnetic Resonance

Letter to the Editor

Processing of Heteronuclear NMR Relaxation Data with the New Software DASHA

Abstract. The new program DASHA is an efficient implementation of common data processing steps for the protein internal dynamic analysis. The "model-free" parameters and their uncertainties (Lipari G., Szabo A.: J. Am. Chem. Soc. **104**, 4546–4559 (1982) can be calculated from an arbitrary combination of experimental data sets (i.e. heteronuclear ^{1}H-^{15}N or ^{1}H-^{13}C relaxation times and NOE values at different spectrometer frequencies). Anisotropy of the molecular rotational diffusion could be also taken into account without introduction of the new adjustable parameters into the spectral density function $J(\omega)$, provided the structure of the molecule is known. Parameters of chemical (conformational) exchange can be estimated from the CPMG spin-lock frequency dependences (Bloom *et al.*: J. Chem. Phys. **42**, 1615–1624 (1965); Orekhov *et al.*: Eur. J. Biochem. **219**, 887–896 (1994). The program can be used both in the interactive and batch modes. It has sophisticated PostScript plotting facilities.

The analysis of heteronuclear relaxation times and NOE values became a commonly used procedure for the elucidation of protein internal dynamics (for a recent review see [1]). A major step in this analysis is the extraction of the "model-free" [2, 3] parameters (i.e. order parameters and internal correlation times) from relaxation data. Generally, calculations are performed in the assumption of isotropic overall rotational diffusion of the molecule. The rotational anisotropy could also be taken into account, provided spatial structure of the molecule is available [4]. In some cases transverse relaxation times depend on the CPMG [5] or continuous spin-lock frequency. It could be used for the elucidation of the internal dynamics in milli-microsecond time scales [6]. In this paper we represent a new software package DASHA that provides a flexible tool for NMR dynamic analysis. In the following consideration we will represent software environment and types of input and output data. Then the principles of evaluation of dynamic parameters and their uncertainties will be discussed. Software is written in a standard ANSI "C", and it currently runs under UNIX on Silicon Graphics and SUN workstations. The package consists of the two independent modules DASHA and DIFFC. Both modules can be regarded as the sets of flexible interactive tools for handling of input data, calculation parameters and results. The modules have also on-line help pages.

Calculation of the dynamic parameters and specifically their uncertainties require extensive nonlinear minimizations. Calculations can be fulfilled in the interactive mode on relatively powerful (remote) computer or in the batch mode. The batch script is the sequence of all necessary commands collected in an ASCII file. The results of batch calculations can be analyzed in the interactive mode using "backup restore" concept, which allows one to save current conditions and load them into another program session. User needs an interactive mode and graphic facilities for the handling and visualization of input data and results and for the adjustment of the calculation parameters. Package has no real graphic display. Only facilities for PostScript file generation are implemented in the DASHA module. The resulting parameters and their uncertainties as well as the input data can be flushed into the ASCII files or plotted versus residue number in the PostScript format. The PostScript files can easily be viewed on most graphic stations (SUN, Silicon Graphics, etc.) by the standard software.

The primary input for the "model-free" analysis by the DASHA module are the ASCII files with heteronuclear relaxation times, NOE values and their uncertainties. Before loading the experimental data user should select a heteronuclear pair. By now there are two pairs available, protein backbone ^{15}N-^{1}H and $^{13}C^{\alpha}$-^{1}H. T_1, T_2 and NOE values in an arbitrary combination and for all available spectrometer frequencies can be loaded into the module using read command. The experimental data can also be set or modified interactively for a particular residue during program session. For the analysis of conformational exchange using CPMG or continuous spin-lock frequency dependences (see below) a set of ASCII files including T_2 values for different spin-lock frequencies should be loaded. If "model-free" calculation is made using anisotropic type of spectral density function (see below), loading of the binary file with correlation times and coefficients of anisotropic spectral density function produced by DIFFC module is necessary.

As the input for DIFFC module the protein structure in the format of Protein Data Bank is used. The results of hydrodynamic calculations are in the ASCII format. DIFFC also generates the binary file with the correlation times and coefficients of anisotropic spectral density function. This binary file can be used as the input for DASHA module.

Calculation of the parameter set, ζ for "model-free" or conformational exchange modes (i.e. order parameters, correlation times of the internal motions and conformational exchange contribution to the transverse relaxation times, or the parameters of conformational exchange respectively) is performed in the DASHA module by the nonlinear minimization of the penalty function χ^2 [7]

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(V_i^{\text{th}}(\zeta) - V_i^{\text{exp}}\right)^2}{\left(\Delta V_i^{\text{exp}}\right)^2}, \qquad (1)$$

where $V_i^{th}(\zeta)$ and V_i^{exp} are the theoretical [3, 6, 8, 9] and experimental values respectively and ΔV_i^{exp} is the uncertainty of the experimental value. Index i runs over the experimental data set (i.e. T_1, T_2 and heteronuclear NOE measured at all available spectrometer frequencies for "model-free" analysis and T_2 values measured for different spin-lock frequencies for the "conformational exchange" analysis). An arbitrary subset of the ζ parameters can be fixed to the particular values prior to the minimization if necessary.

To keep the particular adjustable parameter ζ_k within the user specified limits ζ_k^{min}, and ζ_k^{max} (e.g., default minimal and maximal values for the order parameters are 0 and 1, respectively) during minimization, the restrictive potentials $U(\zeta_k)$ are added to the penalty function of Eq. (1):

$$U(\zeta_k) = \begin{cases} (\zeta_k - \zeta_k^{min})^6, & \text{if } \zeta_k < \zeta_k^{min}, \\ 0, & \text{if } \zeta_k^{min} < \zeta_k < \zeta_k^{max}, \\ (\zeta_k - \zeta_k^{max})^6, & \text{if } \zeta_k > \zeta_k^{max}. \end{cases}$$

In the final calculation of χ^2 value (at the end of the minimization), $U(\zeta_k)$ terms are not included.

The user specified number of the Monte-Carlo trials with V_i^{exp} randomly taken from the interval $V_i^{exp} \pm \Delta V_i^{exp}$ is used for the evaluation of the ζ_k uncertainties. Minimal and maximal values of the particular ζ_k parameter, obtained by this procedure, are taken as corresponding lower and upper estimates.

Conformational exchange processes could significantly contribute, (Δ_{ex}), to the measured transverse relaxation times, T_2, in spin-lock experiments [6, 9]:

$$\frac{1}{T_2} = \Delta_{ex} + \frac{1}{T_2^*}, \tag{2}$$

where T_2^* is the "chemical exchange free" part of the transverse relaxation time stemmed from the internal motions faster than the overall rotation correlation time and from the overall rotation of the molecule.

There are several possibilities for the evaluation of conformational exchange contribution. If the exchange rate is not very high relative to the available frequency of the spin-lock or CPMG pulse repetition rate (i.e. exchange rate constant is less then 10^3–10^4 s^{-1}), it is straightforward to make use of Δ_{ex} dependence on the spin-lock frequency [6, 9–11]. Apart from the evaluation of Δ_{ex} this approach provides also chemical shifts dispersion and exchange rate constants. Alternatively Δ_{ex} could be obtained in the "model-free" analysis [5] as an adjustable parameter in nonlinear minimization [12]. This method is especially useful [13] if transverse relaxation times are available for several spectrometers. Then

only one adjustable parameter can account for the Δ_{ex} at all spectrometer frequencies, because

$$\Delta_{ex}^{k} = \Delta_{ex}^{i}\left(\frac{\nu^{k}}{\nu^{i}}\right)^{2} , \tag{3}$$

where indices i, and k correspond to different spectrometer frequencies ν^{i}, and ν^{k}.

The spectral density function, which determines the relaxation parameters is given by:

$$J(\omega) = 2\int_{0}^{\infty}\cos(\omega t)C(t)\,\mathrm{d}t , \tag{4}$$

where $C(t)$ is the autocorrelation function of a relevant vector in the laboratory frame. Both for isotropic and anisotropic rotational diffusion of the molecule we assume $C(t)$ to be the product of two correlation functions corresponding to internal motions, $C_I(t)$, and overall rotation of the molecule, $C_R(t)$:

$$C(t) = C_R(t)C_I(t) . \tag{5}$$

The validity of this factorization can be rigorously shown if internal motions and rotation of the molecule are independent and rotation of the molecule is isotropic. The later condition does not hold for the anisotropic rotation of the molecule. However, as reported in [2], Eq. (5) can still be regarded as good approximation.

Relaxation data can be analyzed using three types of autocorrelation function, $C_I(t)$, for the internal motions [2, 3]. First type assumes that internal motions are very fast, i.e., two orders faster as compared to the overall rotation of the molecule. Then

$$C_I(t) = \mathbf{S}^2 , \tag{6}$$

where $\mathbf{S}^2$ is the square of the generalized order parameter. If internal motions are from one to two orders faster than overall rotation, then

$$C_I(t) = \mathbf{S}^2 + (1-\mathbf{S}^2)\exp\left(-\frac{t}{\tau_e}\right) , \tag{7}$$

where τ_e is the effective correlation time for the internal motions. If the rates of the internal motions are comparable with rate of the overall rotation, then

$$C_1(t) = S_f^2 S_s^2 + (1 - S_s^2) \exp\left(-\frac{t}{\tau_s}\right), \quad (8)$$

where τ_s is the effective correlation time for internal motions on the intermediate time scale between overall rotation correlation time and fast limit (20 ps), S_f^2 and S_s^2 are the order parameters for the motions on fast and intermediate time scales.

The DASHA module can perform the calculations both for the isotropic and anisotropic correlation function of the overall rotation of the molecule. The isotropic correlation function is given by:

$$C_R(t) = \exp\left(-\frac{t}{\tau_R}\right), \quad (9)$$

where τ_R is the overall rotation correlation time.

If the overall rotation of the molecule is isotropic and backbone dynamics of most of the residues can be described by Eqs. (6) and (7), τ_R can be calculated in the DASHA module from the averaged ratio of T_1/T_2 [12]. However, if most of the residues are involved into intermediate (nanosecond) time scale motions, the ratio T_1/T_2 gives an underestimated value of τ_R [14]. For the characterization of anisotropic rotation of the molecule this method cannot be used either. Alternatively, τ_R (or effective τ_R for the overall anisotropic rotation of the molecule) can be estimated as an adjustable parameter during the minimization of the penalty function, $\Gamma(\tau_R)$ as in [11]:

$$\Gamma(\tau_R) = \sum \chi_j^2, \quad (10)$$

where χ_j^2 is the penalty function for the particular nuclei, calculated for the same τ_R. Index j runs over all nuclei in the consideration. It should be noted that both for isotropic and anisotropic (see below) overall rotation of the molecule only one adjustable parameter τ_R is necessary if the spatial structure of the molecule is specified.

The DIFFC module was developed to calculate rotation correlation times and coefficients of the correlation function for the overall anisotropic rotation of molecules with known spatial structure. The general approach of interpretation of nuclear spin relaxation in rigid molecules undergoing anisotropic rotational Brownian diffusion was developed by Woessner [15]. This approach implies that the overall rotation correlation times and coefficients depend on the rotation diffusion rates. These rates can be calculated using the Brenner's theory of trans-

lation-rotation dynamics [16] and the beads model approximation [17–19]. The beads model approximation assumed the molecule as a rigid body, represented by a number of frictional points with particular radii, called beads or subunits, with hydrodynamics interactions described by the Oseen tensor or its modifications. In the DIFFC facility the user defined atoms are used as the positions of the frictional points.

The correlation function $C_R(t)$ which is relevant for the nuclear spin relaxation in rigid molecules undergoing anisotropic rotational Brownian diffusion was represented by Woessner [15]:

$$C_R(t) = \sum_{n=1}^{5} C_n \exp\left(-\frac{t}{\tau_n}\right) , \tag{11}$$

i.e., the overall rotation correlation function is a sum of five exponents with appropriate correlation times, τ_n, and coefficients, C_n, which are given by:

$$\tau_2^{-1} = D_1 + 4D_2 + D_3 ,$$

$$\tau_3^{-1} = D_1 + D_2 + 4D_3 ,$$

$$\tau_4^{-1} = 6\{D + (D^2 - D'^2)^{1/2}\} ,$$

$$\tau_5^{-1} = 6\{D - (D^2 - D'^2)^{1/2}\} ,$$

$$D = \frac{1}{3}(D_1 + D_2 + D_3) ,$$

$$D' = \left\{\frac{1}{3}(D_1D_2 + D_1D_3 + D_2D_3)\right\}^{1/2} ,$$

$$C_1 = 6m^2n^2 , \qquad C_2 = 6l^2n^2 , \qquad C_3 = 6l^2m^2 ,$$

$$C_4 = d - e , \qquad C_5 = d + e ,$$

$$d = \frac{1}{2}\{3(l^4 + m^4 + n^4) - 1\} ,$$

$$e = \frac{1}{6}\{\varepsilon_1(3l^4 + m^2n^2 - 1) + \varepsilon_2(3m^4 + l^2n^2 - 1) + \varepsilon_3(3n^4 + m^2l^2 - 1)\} ,$$

$$\varepsilon_i = \frac{D_i - D}{(D^2 - D'^2)^{1/2}} , \qquad i = 1, 2, 3 , \tag{12}$$

where D_1, D_2 and D_3 are the eigenvalues of rotation diffusion tensor or appropriate eigenvalues of reaction diffusion tensor, and l, m and n are the direction cosines of the vector between nuclei in a coordinate frame fixed to the principal axes of diffusion ellipsoid. D_1, D_2, D_3, l, m and n values are calculated within DIFFC module using the known spatial structure of the protein.

If the particular model for the spatial structure is chosen (i.e. positions of beads and their radii) only one adjustable parameter is necessary to calculate all correlation times and coefficients of the anisotropic correlation function (Eq. (11)). In DIFFC this adjustable parameter is the effective viscosity. In the DASHA module the effective overall rotation correlation time is used for this purpose. In both cases changing of one of those parameters leads to the equal rescaling of all correlation times in Eq. (11).

In conclusion the software package DASHA provides a flexible interactive tool for molecular dynamics analysis based on NMR data. Model free approach and chemical exchange analysis could be easily fulfilled. Anisotropy of the overall rotation diffusion of the molecule can be taken in to account by the hydrodynamic calculations. Theoretical NMR data can be simulated for the different models of molecular dynamics. We hope that new package could contribute to the NMR elucidation of the protein internal dynamics. The DASHA software is available upon request from the authors (electronic mail address in Internet is aars@siobc.msk.su or aars@siobc.ras.ru).

Acknowledgement

This work was supported by a grant (MIP000) from the International Science Foundation.

References

[1] Wagner G.: Curr. Op. in Struct. Biol. **3**, 748–754 (1993)
[2] Lipari G., Szabo A.: J. Am. Chem. Soc. **104**, 4546–4559 (1982)
[3] Clore G.M., Szabo A., Bax A., Kay L.E., Driscoll P.C., Gronenborn A.M.: J. Am. Chem. Soc. **112**, 4989–4991 (1990)
[4] Barbato G., Ikura M., Kay L.E., Pastor R.W., Bax A.: Biochemistry **31**, 5269–5278 (1992)
[5] Carr H.Y., Purcell E.M.: Phys. Rev. **94**, 630–638 (1954)
[6] Bloom M., Reeves L.W., Wells E.J.: J. Chem. Phys. **42**, 1615–1624 (1965)
[7] Palmer A.G., Rance M., Wright P.E.: J. Am. Chem. Soc. **113**, 4371–4380 (1991)
[8] Abragam A. in: The principles of Nuclear Magnetism. England, Oxford: Clarendon Press 1961.
[9] Dewerell C., Morgan R.E., Strange J.H.: Mol. Phys. **18**, 553–559 (1970)
[10] Reeves L.W. in: Dynamic Nuclear Magnetic Resonance Spectroscopy (Jackman L.M., Cotton F.A., eds.), pp. 83–130. New York, San Francisco, London: Academic Press 1985.
[11] Orekhov V.Yu., Pervushin K.V., Arseniev A.S.: Eur. J. Biochem. **219**, 887–896 (1994)
[12] Kay L.E., Torchia D.A., Bax A.: Biochemistry **28**, 8972–8979 (1989)
[13] Barchi J.J., Grasberger B., Gronenborn A.M., Clore G.M.: Protein Science **3**, 15–21 (1994)

[14] Orekhov V.Yu., Pervushin K.V., Korzhnev D.M., Arseniev A.S.: J. Biomol. NMR **6**, 113–122 (1995)
[15] Woessner D.E.: J. Chem. Phys. **37**, 647–654 (1962)
[16] Happel J., Branner H. in: Low Reynolds Number Hydrodynamics. Leyden: Nordoff 1973.
[17] Garsia de la Torre J., Bloomfield V.A.: Biopolymers **16**, 1747–1763 (1977)
[18] Garsia de la Torre J., Bloomfield V.A.: Biopolymers **16**, 1765–1778 (1977)
[19] Garsia de la Torre J., Bloomfield V.A.: Q. Rev. Biophys. **14**, 81–139 (1981)

V. Yu. Orekhov, D. E. Nolde,
A. P. Golovanov, D. M. Korzhnev,
A. S. Arseniev

Shemyakin and Ovchinnikov Institute
of Bioorganic Chemistry,
Russian Academy of Sciences,
ul. Miklukho-Maklaya 16/10,
Moscow, 117871 Russia

Received October 26, 1995