
Proceedings of the XXI International Meeting on Radio and Microwave Spectroscopy
RAMIS 2005, Poznań-Będlewo, Poland, April 24–28, 2005

Quantitative Analysis of Transferred Nuclear Overhauser Effects in Complex Spin Systems by Full Relaxation Matrix Analysis

J. CZAPLICKI* AND A. MILON

Institut de Pharmacologie et de Biologie Structurale, CNRS
205, route de Narbonne, 31077 Toulouse, France

A computer program was developed for studying transferred nuclear Overhauser effects in complex spin systems. It permits quantitative analysis of nuclear Overhauser effects observed in biologically important systems, such as ligands interacting with transmembrane receptors in the presence of lipid bilayers. The full generalized relaxation matrix approach takes into account the local mobility, spin equivalence, finite exchange rates, and spectral overlap. The program can be used either to simulate theoretical nuclear Overhauser effect buildup curves or to fit a relaxation matrix of a given model to experimental data. Selected examples illustrate the program's performance.

PACS numbers: 82.56.Ub, 82.20.Wt, 07.05.Kf

1. Introduction

NMR spectroscopy is an important tool in studying structure and dynamics of biological molecules [1]. It has evolved to the point where it is possible to study structures of ligands, bound to their receptors from the exchange-mediated nuclear Overhauser effect (transferred NOE) [2]. Typically, ligand is a small molecule, whose highly resolved spectrum contrasts with the broad or even unobserved spectrum of a slowly tumbling macromolecule. Complexation implies immobilization of the ligand in the binding site of the receptor, which leads to the quenching of the ligand's signal. For low affinities, there is fast exchange between the ligand's free and bound states, leading to a transfer of information from the bound state

*corresponding author; e-mail: cgeorge@ipbs.fr

to the free one, where it is recorded. Transferred NOE spectra are analyzed in the same way as ordinary NOE spectra, with the same approximations (linearization of buildup curves for short mixing times, using reference data to determine unknown distances). A new parameter which appears in the case of transferred NOE is the exchange rate between the bound and free states. In principle it can be measured by NMR or other methods [3–5].

Our research activities focus on the studies of neuropeptides, which are implicated in the perception of pain and control of motility by binding to opioid receptors [6]. The ligand conformation in its binding site is a very important piece of information in structural model refinements. Neuropeptides interacting with opioid receptors in the presence of lipid bilayers represent a complex macromolecular system, in which molecules are free to adopt varying conformations and appear in several states (ligand molecules in the free state, interacting with lipids and complexed with receptors). Exchange rates and populations of molecules in each state can be inferred from a basic set of parameters, such as concentrations of system components and diffusion constants [7]. In order to extract information about the structure of the bound ligand from experimental data collected on samples of this type, a rigorous quantitative analysis of NOE data is necessary. We opted for the full relaxation matrix treatment of the system. Recent examples of the relaxation matrix application concern systems composed of two molecules (ligand/enzyme) in two or three states [8–11]. We generalized these approaches by creating a program, capable of handling several molecules exchanging among multiple states with arbitrary rate constants. The software, ANAGRAM (analysis of NOE amplitudes by the generalized relaxation matrix), handles local mobility, spin equivalence and spectral overlap. Given a definition of a spin system, it can simulate theoretical data. More importantly, it performs fitting to experimental data. Herein we present a description of the methods and algorithms used in the program.

2. Theory

The time evolution of the matrix of NOE peak intensities \mathbf{A} can be described by the differential equation

$$\frac{d\mathbf{A}}{dt} = -\mathbf{G}\mathbf{A}, \quad \mathbf{G} = \mathbf{R} - \mathbf{K}, \quad (1)$$

where \mathbf{R} represents the relaxation matrix of the spin system, and \mathbf{K} is the matrix of chemical or conformational exchange rates. \mathbf{G} is the generalized relaxation matrix of the spin system. The solution of Eq. (1) is obtained by computing the eigenvalues λ_i and the matrix of eigenvectors χ of the generalized relaxation matrix \mathbf{G} for a given mixing time τ :

$$\mathbf{A}(\tau) = \mathbf{A}_0 \exp(-\mathbf{G}\tau) = \mathbf{A}_0 \chi e^{-\lambda\tau} \chi^{-1}, \quad (2)$$

where $e^{-\lambda\tau}$ is a diagonal matrix with elements $e^{-\lambda_i\tau}$ and \mathbf{A}_0 is the matrix of ini-

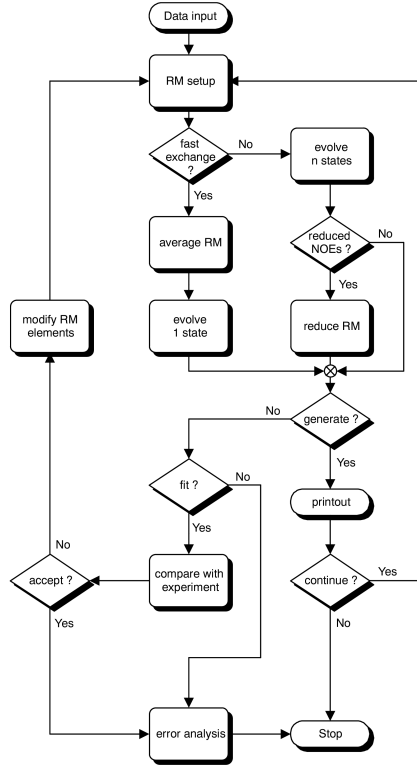


Fig. 1. Flowchart of the ANAGRAM program.

tial magnetization values, usually normalized to unity in thermal equilibrium and multiplied by the corresponding spin populations. Figure 1 shows the flowchart of the program. The input parameters define the spin system with its structure and dynamics, and the details pertaining to the working mode (simulation or fit). The relaxation matrix \mathbf{R} as given by [12–15] contains the dipolar relaxation rates for all spins. In general, for k states, the matrix \mathbf{R} has $Nk \times Nk$ elements, where N is the total number of spins in all molecules in any given state. For isotropically tumbling isolated spin pairs, the off-diagonal elements σ_{kl} are given by

$$\sigma_{kl} = \frac{B}{r_{kl}^6} [6J_2(\tau_{kl}) - J_0(\tau_{kl})]. \quad (3)$$

The constant $B = 0.1\gamma^4\hbar^2$, r_{kl} is the distance between nuclei k and l , and the spectral density functions

$$J_n(\tau_{kl}) = \frac{\tau_{kl}}{1 + (n\omega\tau_{kl})^2} \quad (4)$$

are Lorentzian functions, depending on the Larmor frequency ω and on the rota-

tional correlation time τ_{kl} of the internuclear vector r_{kl} . The diagonal elements of the relaxation matrix ρ_{kk} are the spin-lattice relaxation rates, given by

$$\rho_{kk} = \sum_{l \neq k} \rho_{kl} + \rho_k^*, \quad (5)$$

where the ρ_k^* are the relaxation leakage terms, representing spin relaxation due to non-dipolar mechanisms, and

$$\rho_{kl} = \frac{B}{r_{kl}^6} [J_0(\tau_{kl}) + 3J_1(\tau_{kl}) + 6J_2(\tau_{kl})]. \quad (6)$$

Local motions are taken into account by averaging the relaxation rates of the affected spins according to the $\langle r^{-6} \rangle$ law

$$\langle \sigma_{kl} \rangle = \frac{1}{M} \sum_{i=1}^M \sigma_{kl}^{(i)} \quad (7)$$

or motions slower than the overall tumbling of the molecule. For fast motions, such as those of rapidly rotating methyl protons, the $\langle r^{-3} \rangle^2$ should be used. The full expression was given by Tropp [16]:

$$\sigma_{kl} = B \{ [6J_2(\tau_c) - J_0(\tau_c)] \mathcal{F} + [6J_2(\tau_e) - J_0(\tau_e)] \mathcal{G} \},$$

$$\frac{1}{\tau_e} = \frac{1}{\tau_c} + \frac{1}{\tau_m}, \quad (8)$$

where τ_c is the correlation time of the vector between a spin and the center of the methyl group, τ_m is the time constant characterizing the jumps of methyl protons, and τ_e is the effective correlation time for the local movement of the inter-proton vector. The functions \mathcal{F} and \mathcal{G} are given by [17, 18]:

$$\mathcal{F} = \frac{4\pi}{5} \sum_{m=-2}^2 \left\langle \left| \frac{Y_m^2(\Phi)}{r^3} \right| \right\rangle^2, \quad \mathcal{G} = \frac{4\pi}{5} \sum_{m=-2}^2 \left\langle \left| \frac{Y_m^2(\Phi)}{r^3} \right|^2 \right\rangle - \mathcal{F}. \quad (9)$$

Y_m^2 are normalized spherical harmonics of the second order, while Φ represents the set of spherical coordinates (θ_{ij}, ϕ_{ij}) of the internuclear vector r_{ij} . The brackets denote averaging over all equivalent sites of the methyl protons. The intra-methyl interaction between two protons is given by Eq. (3), with the spectral density functions

$$J_n(\omega) = \frac{1}{4} \frac{\tau_c}{1 + (n\omega\tau_c)^2} + \frac{3}{4} \frac{\tau_e}{1 + (n\omega\tau_e)^2}, \quad (10)$$

where τ_e is given by Eq. (8). The rotational correlation time τ_{kl} reflects the local mobility of the internuclear vector r_{kl} , linking spins k and l .

The elements k_{ij} of the exchange matrix \mathbf{K} are rates, describing the migration of spins from one state to another. In practice, we often work with populations which may be estimated from the concentrations of substrates at equilibrium. For

a special case of the two-state system, there are two independent rates and two populations, given by

$$p_A = \frac{k_{BA}}{k_{AB} + k_{BA}}, \quad p_B = \frac{k_{AB}}{k_{AB} + k_{BA}}. \quad (11)$$

Generating NOE data matrix for a given set of mixing times τ requires calculating the term $\exp(-\mathbf{G}\tau)$, which consists in multiple diagonalizations. Similar time-consuming procedures are involved in fitting to experimental data, which can be best described as the search of the global minimum of a selected error function. This is a difficult problem and no single strategy exists, which guarantees a solution. We provide several different algorithms, assuring flexibility in choosing the strategy, adapted to any particular problem. The user can choose between the hybrid matrix method [19], conjugate gradient algorithm [20] with matrix gradients calculated according to Yip [21], and the simulated annealing approach as described by Goffe et al. [22]. Only the last of the three algorithms offers the possibility to reach the global minimum of the minimized error function. Consequently, we have spent considerable time trying to optimize its performance.

Simulated annealing (SA) is a stochastic minimization algorithm, randomly sampling the parameter space, and based on the probabilistic Monte Carlo method [23]. The procedure is equivalent to simulating thermal fluctuations with jumps occurring with the Boltzmann probabilities at temperature T . The temperature is reduced according to the selected cooling protocol. In the simplest case, it is a geometric progress, but our preferred choice is the Gaussian type of the cooling scheme

$$T_i = T_0 e^{-\left(\frac{i-1}{\alpha}\right)^2}, \quad (12)$$

where i is the index of the current temperature, and α determines the cooling rate. The advantage of this model is that in the beginning the temperature falls off slowly, permitting a thorough search of the parameter space, then it drops faster when the algorithm has sampled the function landscape adequately, and slows down in the end, permitting the refinement of results. At each temperature, the parameter space is sampled uniformly, with efficiency depending, among other factors, on the ranges of variability selected for fitted parameters. They may differ by orders of magnitude, rendering the sampling difficult. To alleviate the problem, we contract the parameter space by applying a logarithmic transformation. If the lower and upper bounds of parameters are denoted by N -component vectors \mathbf{L} and \mathbf{U} , each point \mathbf{x} from the interval $[\mathbf{L}, \mathbf{U}]$ can be transformed into $\boldsymbol{\xi}$ and vice versa

$$\boldsymbol{\xi} = \ln(\mathbf{x} - \mathbf{L} + c), \quad \mathbf{x} = e^{\boldsymbol{\xi}} + \mathbf{L} - c. \quad (13)$$

An arbitrary positive constant c is added to avoid calculating $\ln(0)$ when $\mathbf{x} = \mathbf{L}$. As a result, the sampling of the contracted parameter facilitates locating promising areas. The disadvantage of the SA algorithm comes from the fact that to find the

global minimum, adequate sampling of the parameter space is necessary, which implies large numbers of generated points, and hence long execution times.

The minimization of the error function permits the determination of the cross-relaxation rates for all spins in the system. The next required step is the conversion of the relaxation matrix elements into corresponding internuclear distances. This procedure works well only for isolated spin pairs. When local dynamics becomes important, the above relations give only apparent distances, therefore it is important to estimate distance errors correctly. Finally, thus obtained distances are used as restraints in molecular dynamics simulations, leading to a three-dimensional structure of the studied compound.

It is worth noting that acquisition of full buildup curves is essential for a successful reconstruction of the relaxation matrix [24–26]. The effects of spin diffusion seen at long mixing times permit to resolve ambiguities from the analysis of usually undetermined systems.

3. Software parallelization and performance

Analysis of large spin systems leads to excessively long execution times. Modern processors are characterized by high numbers of floating point operations (FLOPS) per unit time, but the analysis of a moderately complex spin system with adequate sampling of the parameter space requires months to years on a monoprocessor computer. Hence, it is necessary to distribute the computational task among many processors, trying to achieve maximal time gain by way of their organization and synchronization. Our program has been developed on an SGI Origin 2000 supercomputer, a shared memory system equipped with 64 mixed processors ($40 \times$ R12K at 300 MHz and $24 \times$ R14K at 500 MHz) with the crossbar architecture. The software has been parallelized under OpenMP [27], using the coarse-grained parallelism, assuring a quasi-linear Amdahl speedup factor for the SA algorithm. The search trajectory in the parameter space was divided equally among the processors, using a first-come-first-served queue system. Individual results are pooled after each iteration and the best candidate becomes the starting point for the next iteration. Our tests concerned the compounds currently studied in our laboratory, whose description will appear elsewhere. A typical sample of the results in terms of execution times and speedup factors is presented in Fig. 2 (see captions for details). Single-processor execution times depend on the number of variables N according to the well-known N^3 relation.

Originally, the program was developed under the UNIX operating system. The code performs best on shared-memory machines. On the Intel platform, the parallelization can no longer offer a significant advantage, since a single memory bus creates bottlenecks in memory access. However, since the structure of the software is modular, porting the code to a distributed-memory machine, such as a Linux cluster, should pose no particular problem. This work, as well as the parallelization under message passing interface (MPI), is underway.

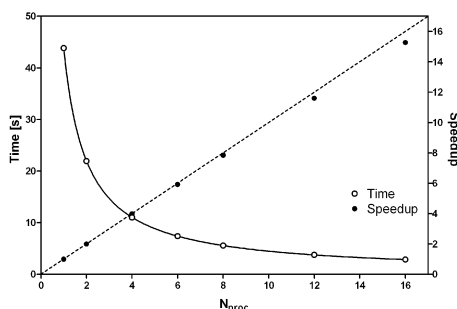


Fig. 2. Execution times and speedup factors for a test system of 52 spins in 1 state. The benchmarks were performed on SGI Origin 2000 computer, using a homogeneous set of R12K processors at 300 MHz. The system is represented by a 52×52 matrix with 703 independent variables. This number reflects solely the unique cross-relaxation rates (the duplicates, found in groups of equivalent spins were excluded). The execution times were normalized by dividing the result of the UNIX system “time” command (the sum of the user and system CPU time) by the number of processors N_{proc} , by the number of iterations, and by the number of mixing times used in each data set. The dashed line shows the ideal linear speedup. The degree of parallelization of the code is 99.7%, as determined from the fitting to the Amdahl law.

4. Conclusions

ANAGRAM permits analysis of multimolecular and multistate spin systems, taking into account the effects of local mobility, spin equivalence, finite exchange rates, and spectral overlap. Future versions of the software will permit to analyze results of experiments involving heteronuclear NOEs. Also, anisotropy of molecular reorientations will be included.

Acknowledgments

We acknowledge financial support from CNRS (programme IMABIO) and Région Midi-Pyrénées for the NMR equipment. The computational part of this work was made possible by the allocation of resources from the regional scientific grouping Calmip.

References

- [1] G.M. Clore, C.D. Schwieters, *Current Opinion in Structural Biology* **12**, 146 (2002).
- [2] G.M. Clore, A.M. Gronenborn, *J. Magn. Reson.* **48**, 402 (1982).
- [3] M. Andrec, J.H. Prestegard, *J. Biomol. NMR* **9**, 136 (1997).
- [4] R. Ishima, D.A. Torchia, *J. Biomol. NMR* **14**, 369 (1999).

- [5] Z. Zolnai, N. Juranic, D. Vikić-Topic, S. Macura, *J. Chem. Inf. Comput. Sci.* **40**, 611 (2000).
- [6] J.-C. Meunier, *Biochimie* **68**, 1153 (1986).
- [7] J. Czaplicki, A. Milon, *J. Chim. Phys.* **95**, 196 (1996).
- [8] E.V. Curto, H.N.D. Moseley, N.R. Krishna, *J. Comp. Aided Mol. Design* **10**, 361 (1996).
- [9] V. Jayalakshmi, N. Rama Krishna, *J. Magn. Reson.* **155**, 106 (2002).
- [10] D. Li, E.F. DeRose, R.E. London, *J. Biomol. NMR* **15**, 71 (1999).
- [11] R.E. London, *J. Magn. Reson.* **141**, 301 (1999).
- [12] J.W. Keepers, T.L. James, *J. Magn. Reson.* **57**, 404 (1984).
- [13] R.E. London, M.E. Perlman, D.G. Davis, *J. Magn. Reson.* **97**, 79 (1992).
- [14] F. Ni, Y. Zhu, *J. Magn. Reson. B* **102**, 180 (1994).
- [15] H. Moseley, E. Curto, N.R. Krishna, *J. Magn. Reson. B* **108**, 243 (1995).
- [16] J. Tropp, *J. Chem. Phys.* **72**, 6035 (1980).
- [17] T.M.G. Koning, R. Boelens, R. Kaptein, *J. Magn. Reson.* **90**, 111 (1990).
- [18] S. Edmondson, *J. Magn. Reson. B* **103**, 222 (1994).
- [19] S.P. Edmondson, *J. Magn. Reson.* **98**, 283 (1992).
- [20] W.H. Press, B. Flannery, S. Teukolsky, W.T. Vetterling, *Numerical Recipes in Fortran*, Cambridge University Press, New York 1992.
- [21] P.F. Yip, *J. Biomol. NMR* **3**, 361 (1993).
- [22] W.L. Goffe, G.D. Ferrier, J. Rogers, *J. Econometrics* **60**, 65 (1994).
- [23] D. Vanderbilt, S.G. Louie, *J. Comp. Phys.* **56**, 259 (1984).
- [24] T.E. Malliavin, M.A. Delsuc, J.Y. Lallemand, *J. Biomol. NMR* **2**, 349 (1992).
- [25] A.K. Suri, R.M. Levy, *J. Magn. Reson. B* **106**, 24 (1995).
- [26] H. Moseley, W. Lee, C. Arrowsmith, N.R. Krishna, *Biochemistry* **36**, 5293 (1997).
- [27] R. Chandra, L. Dagum, D. Kohr, D. Maydan, J. McDonald, R. Menon, *Parallel Programming in OpenMP*, Morgan Kaufmann, San Diego 2000.