THE RELAXATION MATRIX RECONSTRUCTED FROM AN INCOMPLETE SET OF 2D-NOE DATA : STATISTICS AND LIMITS

Patrice Koehl and Jean François Lefèvre

GCMMS, Institut de Biologie Moleculaire et Cellulaire
15, rue Descartes. 67084 Strasbourg CEDEX. France

Among the different techniques involved in the field of macromolecule structure determination, NMR spectroscopy has recently proven to be a powerful tool for defining the folding of a protein (for review, see Wüthrich, 1989) as well as for qualitative studies of DNA fragments in solution (Van de Ven and Hilbers, 1988).Usually, structural informations are derived from NOE (Nuclear Overhauser Effect) experiments.The difficulty, particularly in macromolecules, arises from the fact that the NOE develops by magnetisation transfer within a dipole network (the so called spin diffusion) and depends on fluctuations in the position and the length of the interproton vector (due to internal motion, Jardetzky, 1981). Hence, dynamics and structural informations are imbedded in NOE intensities and need to be deconvoluted.

Along with the development of 2D spectroscopy, a simple method for analysing NOE effects was proposed (Kumar et al., 1981). It consisted of directly estimating the distances between pairs of protons within the molecule from the intensities of the NOEs obtained by one or two dimensional experiments, at short mixing times in order to avoid the spin diffusion phenomenon (i.e. indirect pathways for magnetisation transfers). These sets of distances allowed reconstruction using different types of approach : application of distance constraints alone (Braun and Go, 1985, Altman & Jardetzky, 1990), mixed energetic and geometric constraints minimisation (Kaptein et al, 1985), restrained molecular dynamics (Nilsson et al., 1986). In these methods, the internal motions of the molecule are taken in account by giving some uncertainty in the determination of distances. Weak NOE's are associated with large uncertainties which can account either for large distances or for short distances in a very mobile region. The hope here is that the uncertainty given can be supplemented by the tighter constraints. In fact, these approaches have proven to give consistent folding of the peptidic chain of several proteins (Wüthrich, 1989). However, when very subtle features of the structure (like bent helices, twisted beta sheets or side chains position and dynamics) have to be determined, a refinement procedure has to be applied on these models of structure.

In parallel to protein structure determination, an increasing effort has been put in DNA fragments and oligonucleotide structural studies by means of NMR. Since the general features of these molecules are already established (double stranded helical molecules, either right or left handed) the purpose of these high resolution studies is precisely to refine the structure and to determine either sequence specific structure variations, and/or the dynamics of these molecules. In this case, a more precise analysis of the spin diffusion and the dynamics effects on NOE's is needed. Several methods have been proposed to take in account the indirect transfers of magnetisation. These methods are based on two different ideas.

One approach is to calculate the NOE intensities (for a one or two dimensional NOE experiment), using the full relaxation matrix derived from a molecular model. The system of Bloch equations which describes the magnetisation transfers can be solved numerically, either by stepwise integration of the differential equation (Lefèvre et al., 1987), or by evaluating the matrix exponentiation involved in their general solution (Keepers & James, 1984; Olejniczak et al., 1986). The structure of the molecule is then iteratively modified, using a non linear least square optimizer, till an agreement is reached with the experimental NOEs. These optimisers either use numerically evaluated gradients (see Lefèvre et al., 1987, or the programme COMATOSE, Borgias and James, 1988) or, as described recently, analytical expressions of the derivatives of the NOE according to the coordinates of the protons (Yip & Case, 1989)

Another approach is to directly backtransform the NOE matrix in order to

obtain the relaxation matrix (it should be noted that this method can only be applied in the case of 2D experiments, as the equation for a 1D case is transcendental in Γ, the relaxation matrix). This scheme requires diagonalisation of the NOE matrix, which is only possible in the case of small molecules, since the overlap of resonances does not permit quantitative estimation of all possible NOEs for large molecules. To overcome this problem (in the case of large oligonucleotides), Kaptein and coworkers have proposed to combine the incomplete set of experimental NOEs with a set of theoretical NOEs, calculated from a structural model according to the Bloch equations. These two sets of NOEs allow to build a complete, but mixed NOE matrix. The relaxation matrix obtained after its backtransformation is used to reconstruct an improved structure, using restrained molecular dynamics refinement, which then serves as a new initial model for the dynamics of the iterative procedure (IRMA algorithm, Boelens et al, 1988, 1989).

Both methods however have the drawback to require a model for the dynamics of the molecule in order to estimate the relaxation rate constants from a given structure. Usually, the molecules are supposed to be rigid, and the same correlation time is used for all interproton vectors. The influence of such a systematic error, as internal motions have been shown to be important in DNA (Hogan & Jardetzky, 1979; Levy et al., 1983), on the quality of the structures deduced with the two approaches described above has not yet been studied.

To circumvent this problem, we recently proposed a new approach (encoded in the GOLF program) in which the longitudinal relaxation rate constants ρ and the cross relaxation rate constants σ (which in fact are the primary NMR parameters) are determined first from the experimental NOEs. The main idea of this method is to deconvolute the problem of indirect transfers of magnetisation in a NOE experiment before getting in the details of the determination of the structure and dynamics of the molecule, which comes in the analysis of the relaxation rate constants (ρ) and (σ).

The basic mathematics of all these methods are the same : they involve a general least square optimiser to fit the experimental values to a "model" that depends on adjustable parameters. The efficiency of such procedures has to be evaluated through a statistical test of the goodness of the fit. In this paper, we want to address this problem for the GOLF method, by determining the number of NOE mixing times that should be taken in account in order to obtain a "good" fit.

GOLF : a general strategy for reconstructing the relaxation matrix

a. Nuclear Overhauser effects are non linear functions of the relaxation parameters

The phenomological Bloch equations (Bloch, 1946) adapted by Solomon for spin-spin interactions (Solomon, 1955) describe the time dependance of the longitudinal magnetisation M_i of a spin i :

$$\frac{dM_i}{dt} = -\rho_i(M_i - M_{i0}) - \sum_{j \neq i} \sigma_{ij}(M_j - M_{j0}) \quad (1)$$

where ρ_i is the longitudinal relaxation rate constants between spins i and j, σ_{ij} is the cross relaxation rate constant between spins i and j, and M_{i0} is the z-magnetisation at equilibrium of spin i

For homonuclear proton relaxation, the expression of ρ_i and σ_{ij} are well known :

$$\rho_i = 56.92 \sum_{j \neq i} \left(\frac{1}{r_{ij}}\right)^6 [J(0) + 3J(\omega) + 6J(2\omega)] \quad (2)$$

$$\sigma_{ij} = 56.92 \left(\frac{1}{r_{ij}}\right)^6 [6J(2\omega) - J(0)] \quad (3)$$

$J(\omega)$ are the spectral density functions, which depend on the resonance frequency and on an apparent correlation time for the interproton vector motion (τ_c) :

$$J(\omega) = \frac{\tau_c}{1 + \omega^2 \tau_c^2} \quad (4)$$

The coefficient 56.92 was evaluated for the distance r_{ij} between protons i and j in Å and τ_c in ns.

It should be noted that :

i) equation (1) assumes that only direct dipole-dipole interactions are considered, and does not include possible cross correlation (i.e. multi-spin effects). This has been shown to be valid in the case of 1D and 2D NOE experiments applied to macromolecules (Bull; 1987).

ii) for a given structure, the longitudinal relaxation time is usually calculated by using equation (2). This calculation includes only dipolar interaction with the neighbouring protons attached to the molecule. For macromolecules, this approximation might not be valid anymore at long mixing times (> 300 ms), because of the background relaxation produced by the remote or the solvent protons; for this reason, in our fitting procedure of the relaxation matrix, we keep ρ as an independent parameter, and we do not try to get structural or dynamic informations from its value.

iii) equation (4) only considers a single correlation time for the vector; in a macromolecule in which internal motions are present, it is then important to define a different apparent correlation time for each interproton vector.

The relaxation rate constants can be obtained experimentally from the buildup of NOE. This effect measures the change of magnetisation of a spin i when a neighbouring spin j has been displaced from its equilibrium :

$$NOE_{ij}(\tau_m) = \frac{M_i(\tau_m) - M_{i0}}{M_{i0}} \qquad (5)$$

τ_m is the mixing time during which the magnetisation transfer is allowed. Including (5) into (1) and vectorizing all the corresponding equations lead to a general equation which describes the kinetics of the NOE buildup :

$$d\,NOE\,(t) / dt = - \Gamma\, NOE\,(t) \qquad (6)$$

where NOE is the matrix of NOE for a 2D experiment and Γ is the so called relaxation matrix containing the ρ values on its diagonal and σ values as off diagonal elements.

The solutions of equation (6) are in the case of a 2D-NOE experiment:

$$NOE\,(\tau_m) = \exp\,(-\Gamma\,\tau_m)\,NOE\,(0) \qquad (7)$$

where NOE(0) is the 2D-NOE matrix at zero mixing time (diagonal matrix, each diagonal element corresponding to the number of equivalent spins since no magnetisation has been transfered yet)

b. An iterative method for the determination of the relaxation matrix

Because of missing or overlapping peaks on a 2D experiments (especially on the diagonal), deducing the relaxation parameters directly from the NOE values is far from being straightforward. Hence, we resorted to a non linear fitting method in order to determine the relaxation parameters which minimise the deviation (in terms of a Ki2) between observed (NOE_{obs}) and calculated (NOE_{cal}) NOEs. The method has been extensively described elsewhere (Koehl and Lefevre; 1990) and is summarized in figure 1. Ki2 is defined classically by:

$$Ki2 = \sum_i \frac{[NOE_{cal}(i) - NOE_{obs}(i)]^2}{err(i)^2} \qquad (8)$$

The sum is made over the number of observed NOE, and err(i) is the estimated error on $NOE_{obs}(i)$.

c. Creating an experimental 2D case

In the scope of this paper, we will only consider the application of GOLF to nucleic acids. In the case of a cytosine residue, the spin system considered for the NOE calculations was limited to H1', H2', H2",H3', H4', H6 and H5. We did not include the H5', H5" protons on the basis that they do not belong to the major pathways of spin diffusion or longitudinal relaxation when H6, H5, H1', H2', H2" are observed.

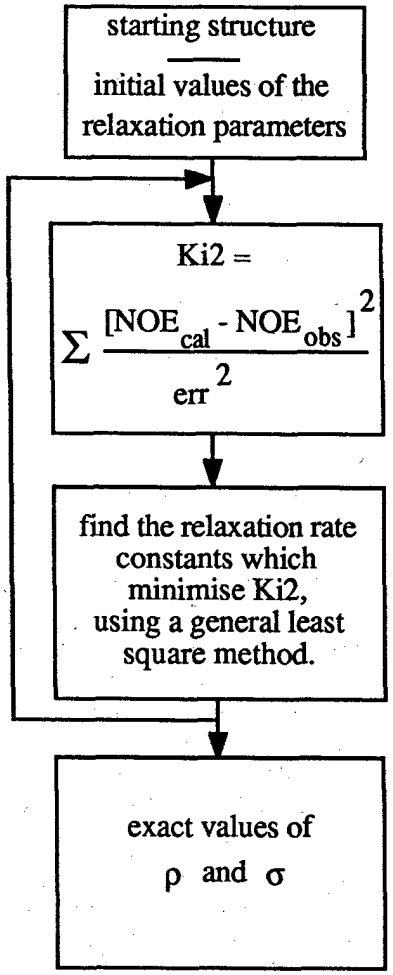

Figure 1 : General scheme for the refinement procedure of the relaxation rate matrix

The reference structure So for this residue was defined from standard B DNA parameters. The relaxation rate constants for this spin system were evaluated using equations (2) and (3), with the same correlation time for each proton-proton vector. Using these values, sets of 2D NOE data were computed using equation (7), for the different mixing times needed in each case. Only a limited number of these NOE were conserved as experimental NOEs for the cytosine residue, as presented in table 1. These values correspond to the NOEs that are usually deduced from a NOESY map of a DNA fragment.

Table 1 : the inter proton NOEs indicated by 1 were calculated for several mixing times, and used in the refinement procedure of the relaxation matrix of a modelised cytosine.

	H1'	H2'	H2"	H3'	H6	H5
H1'	1	1	1		1	1
H2'	1				1	1
H2"	1				1	1
H3'						
H6	1	1	1		1	1
H5	1	1	1		1	1

d. A statistical test to assess the quality of the procedure

A total of 18 parameters were iteratively adjusted using GOLF : the longitudinal relaxation rates for H1', H2', H2", H5 and H6 and the cross relaxation rates corresponding to the same protons, along with the H1'-H3', H2'-H3' and H2"-H3' sigmas. The other relaxation rates were kept constant, equal to their initial values.

These 18 parameters were refined using 15 different starting structures, corresponding to the nucleotide constructed from standard A, B or Z parameters, and for each using 5 possible correlation times (from 2ns to 6ns) to calculate their corresponding relaxation rates. After completion of each refinement, i.e. when a minimum of Ki2 (defined in equation 8) is reached, the corresponding sigmas and rhos are averaged over all trials, and a standard deviation and an error are calculated for each sigma according to:

$$SD = \sqrt{\frac{\sum_{j=1}^{n} (\sigma_{av} - \sigma_j)^2}{n(n-1)}} \qquad (9)$$

$$ER = \frac{SD}{\sigma_{av}} \qquad (10)$$

σ_{av} is the averaged sigma value, and σ_j is the value obtained from the refinement j, the sum being over the 15 refinements from the 15 initial structures. Finally the averaged sigma values are checked against the "real" sigma values (corresponding to the structure So) using a global CHI2 value defined as :

$$CHI2 = \sum_{j=1}^{N} \left(\frac{(\sigma_{av} - \sigma_{real})^2}{\sigma^2_{real}} \right) \qquad (11)$$

the sum being over the 18 refined sigmas. SD, ER and CHI2 are used as statistical tests of the refinement: SD and ER define the dispersion of the refined parameters, starting from different initial values, and CHI2 gives an indication of the quality of the refinements.

RESULTS

a. *NOE values at several mixing times compensate for missing NOE values.*

Experimental 2D NOE data for an oligonucleotide often are incomplete, due to peak overlaps in crowded regions of the NOE map (especially on the diagonal). We addressed the question of whether knowing some NOE values at several mixing times could compensate for the missing values to deconvolute the full spin diffusion process. Hence, we refined a cytosine residue according to the method described above, with the "experimental" NOEs summarized in table 1 obtained for 1, 2, 3, 4 or 5 mixing times (chosen between 100, 200, 300, 400 and 500 ms). Figure 2 represents the evolution of the global CHI2, which checks the refined cross and longitudinal relaxation rates versus the "real" values, as the number of experimental data is increased by adding NOE values at other mixing times. It clearly indicates that it is necessary to record NOE at several mixing times in order to get reliable values for the refined relaxation parameters (see also table 2). It should be noted however that we did not expect reliable informations on the cross relaxation rate constants which include the H3' proton (table 2), even when 5 mixing times are considered, since they have little influence on the set of NOE described in table 1. Including

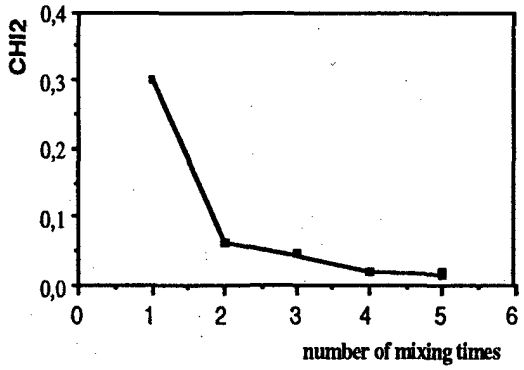

Figure 2 : Evolution of the CHI2 as a function of the number of mixing times which were included in the "experimental" data.

all the parameters in the refinement did not improve these results (data not shown). Hence, increasing the number of experimental values does not overdetermine the problem with respect to the number of parameters to be refined, as all these experimental values are not independent from each other.

Table 2 : Refined versus "real" cross relaxation rate constants for the isolated cytosine residue. Results are shown for 1 and 5 mixing times. ER values were calculated using equation (10)

Protons	σ_{real}	σ_{av}	ER (%)	σ_{av}	ER (%)
		1 mixing time		5 mixing times	
H1'-H2'	-0.35	-0.23	20.5	-0.36	6.3
H1'-H2"	-1.79	-1.82	3.6	-1.83	2.4
H1'-H3'	-0.07	-0.03	19.7	-0.04	22.5
H1'-H6	-0.11	-0.14	8.7	-0.10	3.2
H2'-H2"	-6.90	-6.73	10.0	-6.85	3.4
H2'-H3'	-1.37	-1.48	11.1	-1.30	8.1
H2'-H6	-7.06	-5.63	7.2	-7.09	2.0
H2"-H3'	-0.61	-0.43	12.7	-0.77	19.0
H2"-H6	-0.16	-0.37	28.0	-0.23	15.0
H6-H5	-0.90	-0.89	0.3	-0.90	0.2

b. Each nucleotide of a DNA fragment can be treated separately

The method presented above for an isolated nucleotide cannot be directly translated for the complete analysis of a DNA oligonucleotide. The major problem resides in the fact that the number of parameters to be refined increases as the square of the number of spins included in the system. The problem therefore easily reaches levels out of the range of reliable minimisation routine. Moreover, we believe that it is possible to extract a partial system of spins from the full molecule, and to cut down the dimension of the problem. For example, each nucleotide of a DNA molecule can be refined separately. To assess this statement, we considered a trinucleotide of sequence d(ACG). The full relaxation matrix for this molecule was calculated using equations (2) and (3) for the relaxation rate constants and with the same correlation time (3ns) for each spin-spin vector. A set of 2D NOE was calculated using this relaxation matrix at 5 mixing times (100, 200, 300, 400, 500 ms) for the reasons mentioned above. From these values, only an incomplete set of NOEs was extracted as "experimental" NOEs for the cytosine residue, according to table 1. The refinements were then performed as if the cytosine was isolated.The CHI2 (calculated according to equation 11) was equal to 0.08 for the cross relaxation rate constants and 0.27 for the longitudinal relaxation rate constants. Obviously, the latter values are not correct, since they compensate for the loss of magnetisation related to the protons of the 2 neighbouring residues, included in the calculation of the "experimental" NOEs, but not in the refinement process. However, the main sigma values were driven back to their "real" values. This validates the process of treating each residue separately in the quantitative analysis of a DNA fragment.

Conclusion

The statistical tests performed on simulated NOE data have validated the method of reconstructing the relaxation matrix from an incomplete set of NOEs. They show that it is necessary to measure these available NOEs at several mixing times. However, one should notice that in this method, relaxation rate constants involving spins which do not participate to the spin diffusion are not calculated reliably, unless all NOEs are available.

Another important aspect of the simulation presented in this paper is the fact that it is possible to refine the relaxation rate constants within a selected subset of a spin system. First, this cut down the problem of manipulating large assembly of spins, second it allows to refine only parts of this assembly that are of particular interest. For example, this method can be applied to study the local structure and dynamics in a macromolecule and the effect of changing a residue on these parameters, without resorting to the determination of the structure of the whole molecule.

Obviously, the method can be transposed in order to refine the structure and the dynamics of proteins or selected domains in proteins. The corresponding works are in progress.

References

Altman, R.B. & Jardetzky, O. (in press) in "Methods in Enzymology, 177, Nuclear Magnetic Resonance, Part B : Structure and Mechanisms", Academic Press, NY

Bloch, F., 1946, "Nuclear Induction", Phys. Rev, 70:460

Boelens, R., Koning, T.M.G., & Kaptein, R., 1988, "Determination of Biomolecular Structures from Proton-Proton NOE's using a Relaxation Matrix Approach", J. Mol.Struct., 173:299

Boelens, R. Koning, T.M.G., Van der Marel, G.A., Van Boom, J.H. & Kaptein, R., 1989, "Iterative Procedure for Structure Determination from Proton-Proton NOEs Using a Full Relaxation Matrix Approach : Application to a DNA octamer", J. Magn. Reson, 82:290

Borgias, B.A. & James, T.L., 1988, "Comatose, A Method for Constained Refinement of Macromolecular Structure Based on Two Dimensional Nuclear Overhauser Effect Spectra", J. Magn. Reson., 79:493

Braun, W & Go, N., 1985, "Calculation of Protein Conformations by Proton-Proton Distance Constraints, A New Efficient Algorithm", J. Mol. Biol, 186:611

Bull, T.E., 1987, "Cross Correlation and 2D NOE Spectra", J. Magn. Reson, 72:397

Hogan, M.E. & Jardetzky, 0., 1979, "Internal Motion in DNA", Proc. Natl. Acad. Sci USA, 76:6341

Jardetzky, O., 1980, "On the Nature of Molecular Conformations Inferred from High Resolution NMR", Biochim. Biophys. Acta, 621:227

Kaptein, R. Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F., 1985, "A Protein Structure from Nuclear Magnetic Resonance Data Lac Repressor Headpiece", J. Mol. Biol, 182:170

Keepers, J.W. & James, T.L., 1984, "A Theoretical Study of Distance Determinations from NMR Two-Dimensional Nuclear Overhauser Effect Spectra", J. Magn. Reson, 54:404

Koehl, P. & Lefèvre, J.F., 1990, " The Reconstruction of the Relaxation Matrix from an Incomplete Set of Nuclear Overhauser Effects", J. Magn. Reson (in press)

Kumar, A., Wagner, G., Ernst, R.R. & Wüthrich, K., 1981, "Buildup Rates of the Nuclear Overhauser Effect Measured by Two Dimensional Proton Magnetic Resonance Spectroscopy : Implications for Studies of Protein Conformation", J. Am. Chem. Soc, 103:3654

Lefèvre, J.F., Lane, A.N. & Jardetzky, O., 1987, " Solution Structure of the Trp Operator of Escherichia Coli Determined by NMR", Biochemistry, 26:5076

Levy, G.C., Marchetti, P.S., Ejchart, A., Levy, L.F., Kumar, A., Hilliard, P.R. & Rill, R.L., 1983, "Understanding DNA Conformational Dynamics : Answering Questions and Questioning Answers", J. Biomol. Struct Dyn, 1:795

Nilsson, L., Clore, G.M., Gronenborn, A.M., Brünger, A.T. & Karplus, M., 1986,"Structure Refinement of Oligonucleotides by Molecular Dynamics with Nuclear Overhauser Effect Interproton Distance Restraints : Application to 5'd(CGTACG)$_2$, J. Mol. Biol., 188:455

Olejniczak, E.T., Gampe, R.T. & Fesik, S.W., 1986, "Accounting for Spin Diffusion in the Analysis of 2D NMR data", J. Magn. Reson, 67:28

Solomon, I., 1955, "Relaxation Processes in a System of Two Spins", Phys. Rev, 99:559

Van de Ven, F.J.M. & Hilbers, C.W., 1988, "Nucleic Acids and Nuclear Magnetic Resonance", E. J. Biochem., 178:1

Wüthrich, 1989, "Protein Structure Determination in Solution by Nuclear Magnetic Resonance Spectroscopy", Science, 243:45

Yip, P. & Case, D.A., 1989, "A new method for Refinement of Macromolecular Strcutures based on Nuclear Overhauser Effect Spectra", J. Magn. Reson, 83:643