

J-Bio NMR 078

# CROSREL: Full relaxation matrix analysis for NOESY and ROESY NMR spectroscopy

Bas R. Leeftang<sup>a,\*</sup> and Loes M.J. Kroon-Batenburg<sup>b</sup>

<sup>a</sup>*Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands*

<sup>b</sup>*Bijvoet Center, Department of Crystal and Structural Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands*

Received 20 December 1991

Accepted 12 June 1992

*Keywords:* ROESY; Relaxation matrix analysis; Anisotropic motion; Internal motion; Molecular dynamics; HOHAHA correction

---

## SUMMARY

A method is proposed for quantitative analysis of ROESY peak intensities, to which corrections are applied for their offset dependence and for direct HOHAHA effects. Additionally the effects of anisotropic and internal motion can be assessed. This method has been implemented for full relaxation matrix analysis in the CROSREL program. Although CROSREL is applicable to NOESY data, its use for ROESY peak intensities has been evaluated here, because of its innovative character in this respect. The agreement between calculated and experimental intensities is expressed by a weighted residual  $R_w$  factor, similar to X-ray crystallography. The merits of the program have been tested on methyl(*d*3)  $\beta$ -cellobioside, for which a ROESY build-up series has been acquired, and for which extensive MD simulations have been performed. It is concluded that correction for direct HOHAHA effects is obligatory for the analysis of ROESY data. Extension of the model for methyl  $\beta$ -cellobioside with internal and anisotropic motion, as was derived from MD data, did not improve the results obtained for assumed isotropic tumbling of a rigid model. It has been shown that ROESY peak intensities can be analysed successfully by the CROSREL program.

---

## INTRODUCTION

A most valuable NMR parameter for conformational analysis of biomolecules is the proton-proton ( $^1\text{H}$ - $^1\text{H}$ ) NOE (Wüthrich, 1986). For large molecules it is most conveniently measured by multi-dimensional NOE spectroscopy (NOESY), wherein NOEs are obtained as cross peaks (Jeener et al., 1979; Ernst et al., 1987). The NOE intensity depends on the rate of cross-relaxation of the protons, which in turn depends on fluctuations in the orientation and on the length of the

---

\* To whom correspondence should be addressed at the current address: Bijvoet Center, Department of NMR Spectroscopy, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

interproton vectors (Solomon, 1955; Noggle and Schirmer, 1971). The experimental NOE cross-peak intensity can be either negative or positive, depending on the product of the Larmor frequency,  $\omega$ , and the rotational correlation time,  $\tau_c$ , and on the possible occurrence of spin-diffusion. For molecules that have a  $\tau_c$  near the critical value of  $\sqrt{5}/(2\omega)$  no NOEs can be measured, and NOESY is not applicable. However, NOE type information for this group of molecules can be obtained from Rotating frame nuclear Overhauser Enhancement Spectroscopy (ROESY) (Bothner-By et al., 1984; Bax and Davis, 1985). Until now quantitative analysis of ROESY data has been regarded to be very difficult in terms of distances, because the cross-peak intensity is also influenced by non-relaxation factors that are not always easily accounted for, namely the offset dependence, and multi-spin and Homonuclear Hartmann-Hahn (HOHAHA) effects. In this paper a method is described for the analysis of both ROESY and NOESY experiments. The main emphasis, however, is on the ROESY part, because of its innovative character.

Several methods have been presented to relate experimental NOE cross-peak intensities to a model, in terms of interproton distances (Bothner-By and Noggle, 1979; Kumar et al. 1981; Keepers and James, 1984; Olejniczak et al. 1986; Scarsdale et al. 1986; Boelens et al. 1988; Breg et al. 1989; Summers et al. 1990). Although these methods all focus on the interpretation of NOESY spectra, they are in principle also applicable to the examination of ROESY data. However, the effects of the ROESY offset dependence and HOHAHA transfer must be taken into account. One of the simplest approaches is the analysis of initial cross-peak intensity build-up rates, which leads in a direct way to relative distances, and which can be calibrated using known distances. However, the initial rate approach does not account for multi-spin effects, such as spin diffusion. Therefore, the effect of spin diffusion is minimized by measuring at short mixing times, with the disadvantage, however, of low sensitivity (Olejniczak et al. 1984; Clore and Gronenborn, 1985). The use of non-linear fitting functions overcomes this problem to some extent (Fejzo et al. 1989). Methods that use a full relaxation matrix are preferred, because high signal-to-noise spectra obtained at long mixing times can be used.

Here a full relaxation matrix analysis method is presented to examine ROESY and NOESY spectra, in which multi-spin effects are fully accounted for, that corrects for all offset dependencies (ROESY), and that applies a correction for estimated HOHAHA effects (ROESY). Additionally, with the proposed method it is possible to incorporate internal and anisotropic motion effects. The program will be referred to as CROSREL (CROSs RELaxation).

The consistency of some scaling methods, and the effect of HOHAHA correction, supplied by the program and the influence of anisotropic and internal motion, were investigated with a ROESY build-up series of methyl(*d*3)  $\beta$ -cellobioside.)

## THEORY

### *Basic NOESY and ROESY theory*

Cross-relaxation in multi-spin systems can be described by the generalized Bloch equations. The time dependence of the peak intensities in 2D NOESY or 2D ROESY spectra is given by (Macura and Ernst, 1980):

$$\mathbf{A}(\tau_m) = \exp[-\mathbf{R} \tau_m] \mathbf{M}_0 \quad (1)$$

$\mathbf{M}_0$  is the magnetization defined as the diagonal matrix of peak intensities at  $\tau_m = 0$ .  $\mathbf{A}$  is the NOE/ROE matrix.  $\mathbf{R}$  is the relaxation matrix that consists of diagonal elements  $R_{ii} = \rho_i$  and off-diagonal elements  $R_{ij} = \sigma_{ij}$ . The NOE type relaxation rate  $\rho_i^{\text{NOE}}$  for proton  $i$  is given by (Solomon, 1955; Noggle and Schirmer, 1971):

$$\rho_i^{\text{NOE}} = C \cdot \sum_{j \neq i}^n [6J_2(\omega) + 3J_1(\omega) + J_0(\omega)] \quad (2)$$

and the NOE type cross-relaxation rates  $\sigma_{ij}^{\text{NOE}}$  are:

$$\sigma_{ij}^{\text{NOE}} = C \cdot [6J_2(\omega) - J_0(\omega)] \quad (3)$$

The ROE type relaxation rate  $\rho_i^{\text{ROE}}$  for spins resonating at the carrier frequency is given by (Bothner-By et al. 1984):

$$\rho_i^{\text{ROE}} = C \cdot \sum_{j \neq i}^n [3J_2(\omega) + 4.5J_1(\omega) + 2.5J_0(\omega)] \quad (4)$$

and the ROE type cross-relaxation rates  $\sigma_{ij}^{\text{ROE}}$  by:

$$\sigma_{ij}^{\text{ROE}} = C \cdot [2J_0(\omega) + 3J_1(\omega)] \quad (5)$$

The contribution of leakage relaxation rate  $R_L$ , which is usually added to Eq. 5: C2 and 4, has been omitted here, but will be included later in Eq. 8. In Eqs. 2 – 5,  $C = 0.1 \gamma^4 \hbar^2 (\mu_0 / 4\pi)^2$ . Under the assumption of isotropic tumbling of a rigid molecule with a correlation time  $\tau_c$ , the spectral density function is described by:

$$J_n(\omega) = r_{ij}^{-6} \cdot \frac{\tau_c}{1 + n^2 \omega^2 \tau_c^2} \quad (6)$$

wherein  $\omega$  is the proton Larmor frequency and  $r_{ij}$  is the distance between protons  $i$  and  $j$ .

In general the resonance frequencies will not be identical to the carrier frequency in the ROESY experiment, and the spins will be locked along a tilted spin-lock axis during the mixing time. The angle  $\theta_i$  of the tilted spin-lock axis for spin  $i$  and the  $z$ -axis is dependent on the spin-lock field strength and on the position of the resonance relative to the carrier frequency, and is defined by:

$$\tan \theta_i = \frac{\gamma \mathbf{B}_1}{(\omega_i - \omega_0)} \quad (7)$$

wherein  $\omega_0$  and  $\omega_i$  are the carrier frequency and the resonance frequency of spin  $i$ , respectively, and  $\gamma \mathbf{B}_1$  is the  $rf$  field strength, in which  $\gamma$  is the  $^1\text{H}$  gyromagnetic ratio and  $\mathbf{B}_1$  is the spin-lock field. When  $\theta_i$  values deviate from  $90^\circ$ , magnetization relaxes partially via the ROE and partially via the NOE pathway. Therefore, the NOE and ROE type relaxation and cross-relaxation rates in  $\mathbf{R}$  have

to be mixed, according to the distribution of both relaxation types (Griesinger and Ernst, 1987).

$$\rho_i = \sin^2(\theta_i) \cdot \rho_i^{\text{ROE}} + \cos^2(\theta_i) \cdot \rho_i^{\text{NOE}} + R_L \quad (8)$$

$$\sigma_{ij} = \sin(\theta_i) \cdot \sin(\theta_j) \cdot \sigma_{ij}^{\text{ROE}} + \cos(\theta_i) \cdot \cos(\theta_j) \cdot \sigma_{ij}^{\text{NOE}} \quad (9)$$

Although the leakage relaxation rate  $R_L$  is in principle unique for each spin, in practice it is included in the relaxation matrix as a common correction factor for the diagonal elements  $\rho_i$ .

The exponential matrix equation, in Eq. 1 can be expanded in a power series

$$\exp[-\mathbf{R}\tau_m] = 1 - \mathbf{R}\tau_m + 0.5\mathbf{R}^2\tau_m^2 + \dots \quad (10)$$

The method of initial rate analysis is based on Eq. 10, since for sufficiently short mixing times the first two terms dominate and the peak intensity  $A_{ij}(\tau_m)$  becomes  $-\sigma_{ij} \cdot \tau_m$  and builds up almost linearly with the mixing time. The slope of this curve can accurately be determined in a fitting procedure (Fejzo et al. 1989).

The exponential matrix equation, Eq. 1, can also be solved numerically (Keepers and James, 1984). This implies that the NOE or ROE matrix  $\mathbf{A}$  can be calculated from the relaxation matrix  $\mathbf{R}$ , derived from a given molecular model.

#### Offset dependence of ROESY spectra

ROESY peak intensities are decreased, as a function of the difference of the resonance frequencies with the carrier frequency, (1) during the cross-relaxation period as described before, and (2) by loss of magnetization due to inefficiency of the spin-locking process (Bax and Davis, 1985). For cross peaks the ROE contribution has an offset dependence, produced during cross-relaxation, (1), of  $\sin(\theta_i) \cdot \sin(\theta_j)$ , whereas the NOE contribution has an offset dependence of  $\cos(\theta_i) \cdot \cos(\theta_j)$ . For diagonal peaks these values are  $\sin^2(\theta_i)$  and  $\cos^2(\theta_i)$  for the ROE and NOE contributions, respectively. The spin-locking process itself causes additional loss of magnetization, (2), by two consecutive projections, marked p1 and p2 in Fig. 1. After a  $90^\circ$   $x$ -pulse, the initial magnetization, aligned along the  $y$ -axis, is projected, (p1), onto the effective spin-lock  $z'$ -axis, making an angle,

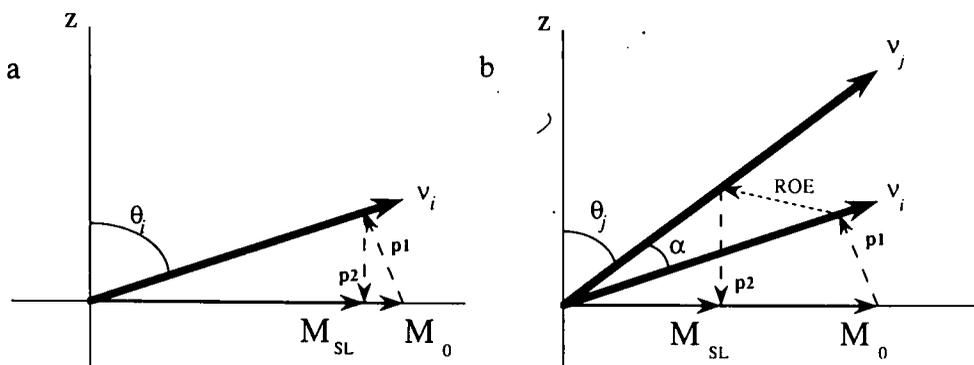


Fig. 1. Graphical representation of the offset dependence of ROESY diagonal peaks (a) and cross peaks (b). The spin-locking process is not fully efficient, due to two projections, marked p1 and p2. The offset dependence during the relaxation process is marked ROE. The angles  $\theta_i$ ,  $\theta_j$ , and  $\alpha$  have been indicated.

( $\theta_i$ ), with the  $z$ -axis. During acquisition only the amount of the magnetization is detected which is obtained after back projection, ( $p_2$ ), of the locked magnetization onto the  $xy$ -plane. Therefore, the ROE and NOE contributions to the cross peaks are attenuated by other factors,  $\sin(\theta_i) \cdot \sin(\theta_j)$  and  $\cos(\theta_i) \cdot \cos(\theta_j)$ , respectively, independent of the length of the mixing time. Within the initial rate approximation, the separate contributions of each relaxation pathway cannot be distinguished. However, since ROESY is generally used in cases in which the NOE contribution is negligible, the offset dependence is dominated by the ROE contribution. For short mixing times (initial rate) the overall correction factor for cross peaks is  $\sin^2(\theta_i) \cdot \sin^2(\theta_j)$ , and for diagonal peaks it is  $\sin^2(\theta_i)$ , since in Eq. 10,  $\rho_i \cdot \tau_m$  is still small compared to the initial magnetization of 1. Within a full relaxation matrix approach the offset dependence can be applied correctly for any mixing time and for any molecule, for which the NOE contribution is not negligible.

### HOHAHA correction

The magnetization transfer during the ROESY spin-lock pulse via scalar couplings (Hartmann–Hahn transfer) can be estimated, and has been described (Bax, 1988). The effective field experienced by spin  $i$  in the rotating frame equals  $\nu_i = \sqrt{(\omega_i - \omega_0)^2 + (\gamma \mathbf{B}_1)^2}$ . The angle  $\alpha_{ij}$  between the effective spin-lock axes of spins  $i$  and  $j$ , respectively, is given by  $\alpha_{ij} = |\theta_i| - |\theta_j|$ , and is displayed in Fig. 1. The Hartmann–Hahn mismatch is expressed in  $2\phi_{ij}$  defined by:

$$\tan(2\phi_{ij}) = \frac{(1 + \cos(\alpha_{ij})) \cdot J_{ij}}{2(\nu_i - \nu_j)} \quad (11)$$

wherein  $J_{ij}$  is the scalar coupling constant of spins  $i$  and  $j$ . HOHAHA type magnetization transfer has an oscillatory character which dampens during the spin-lock period. The transfer frequency is defined as:

$$2q_{ij} = \sqrt{(\nu_i - \nu_j)^2 + (1 + \cos(\alpha_{ij}))^2 \cdot J_{ij}^2/4} \quad (12)$$

Both the Hartmann–Hahn mismatch and the transfer frequency depend on the applied field strength, on the coupling constant, and on the positions relative to the carrier frequency. In order to minimize HOHAHA transfer a low spin-lock field strength is used (e.g. 2500 Hz), and the carrier frequency is positioned outside the spectral region of interest. It has been found, however, that only the spectral region for which the angle  $\theta \geq 60^\circ$  can be analysed quantitatively (Farmer II and Brown, 1987). In the two-spin case, when the HOHAHA oscillation has dampened completely ( $2q_{ij} < 1/\tau_m$ ), the relative peak intensities  $H_{ij}$  caused by HOHAHA are estimated by:

$$\begin{aligned} H_{ii} &= \frac{1}{2} \sin^2(\theta_i) (1 + c_{ij}^2) \\ H_{jj} &= \frac{1}{2} \sin^2(\theta_j) (1 + c_{ij}^2) \\ H_{ij} &= \frac{1}{2} \sin(\theta_i) \sin(\theta_j) s_{ij}^2 \end{aligned} \quad (13)$$

wherein  $c_{ij} = \cos(2\phi_{ij})$  and  $s_{ij} = \sin(2\phi_{ij})$ . This approximation is reliable if  $|\phi| < 15^\circ$  and when the oscillation has dampened.

### Internal motion

In the theory of relaxation of non-equilibrium magnetizations some frequently used simplifications, with respect to the spectral density functions and  $r$ -dependence (cf. Eq. 6), have been made. It has been assumed that the reorientational motion of the molecule is isotropic and that both the distance between the interacting protons and the orientation of the  $^1\text{H}-^1\text{H}$  vector are fixed with respect to a molecular axes system. In general, however, this will not be the case. A more complete description of spectral densities is necessary in order to assess the importance of internal and overall anisotropic motions. The Hamiltonian of two interacting  $^1\text{H}$  spins contains the orientation of the  $^1\text{H}-^1\text{H}$  vector with respect to the laboratory axes system. Owing to this orientation one proton can cause fluctuations in the magnetic field at the position of the other proton. When these fluctuations are in order of the Larmor frequency, they will cause transitions between spin-states, which lead to relaxation (Bloembergen et al. 1948; Solomon, 1955). These fluctuations are described by the time-correlation function:

$$C(t) = \langle P_2^0(\vec{\mu}(0) \cdot \vec{\mu}(t)) \rangle_0 = \langle P_2^0(\cos(\Delta\beta)) \rangle_0 = \langle \frac{1}{2}(3 \cos^2(\Delta\beta) - 1) \rangle_0 \quad (14)$$

wherein  $\vec{\mu}$  is a unit vector describing the orientation of the  $^1\text{H}-^1\text{H}$  vector in the laboratory frame, and  $\Delta\beta$  is the angle between  $\vec{\mu}(0)$  and  $\vec{\mu}(t)$ .  $P_2^0$  is the second-order associate Legendre polynomial. For a rigid isotropically rotating molecule, this function is frequently assumed to have a single exponential form:  $C(t) = \exp(-t/\tau_o)$ , wherein  $\tau_o$  is the correlation time of the overall rotation of the molecule. From the previous discussion it is clear that the fluctuations are not only accomplished by overall rotation of the molecule but also by internal motions that influence the orientation and the length of the individual  $^1\text{H}-^1\text{H}$  vectors. The general form of the correlation function in terms of the orientation of the molecule in the laboratory frame and the orientation of the  $^1\text{H}-^1\text{H}$  vector in the molecule-fixed frame is given by (Steele, 1976; Tropp, 1980):

$$C(t) = 5 \cdot \sum_{rr'=-2}^2 \left\langle \frac{D_{r0}^2(\zeta_0, \eta_0) D_{r'0}^{*2}(\zeta_t, \eta_t)}{r_0^3 r_t^3} \right\rangle \cdot \langle D_{mr}^2(\Omega_0) D_{mr'}^{*2}(\Omega_t) \rangle \quad (15)$$

which can be reduced, by assuming a random distribution of the initial  $\Omega_0$ , to:

$$C(t) = \frac{4\pi}{5} \sum_{rr'=-2}^2 \left\langle \frac{Y_{2r}(\zeta_0, \eta_0) Y_{2r'}^*(\zeta_t, \eta_t)}{r_0^3 r_t^3} \right\rangle \cdot \langle D_{rr'}^{*2}(\delta\Omega) \rangle \quad (16)$$

wherein  $D_{mr}^2$  are Wigner rotation matrices (Steele, 1976),  $\Omega_t$  represents the Euler angles that transform the laboratory frame into the molecule-fixed frame at time  $t$ ,  $Y_{2r}$  are the second order spherical harmonic functions, and  $\zeta_t$  and  $\eta_t$  are the polar angles defining the orientation of the  $^1\text{H}-^1\text{H}$  vector with respect to the molecule fixed axes system.

The expression in Eq. 16 can be rewritten in terms of an internal and overall correlation function:  $C_I(t)$  and  $C_O(t)$ , respectively.

$$C(t) = [(\langle r^{-6} \rangle - S^2 \langle r^{-3} \rangle^2) \exp(-t/\tau_i) + S^2 \langle r^{-3} \rangle^2] C_O(t) = C_I(t) C_O(t) \quad (17)$$

wherein  $S^2 = \sum_r \left| \left\langle \left( \frac{4\pi}{5} \right)^{1/2} Y_{2r}(\zeta, \eta) \right\rangle \right|^2$ .

The correlation function in Eq. 17 has been described by Lipari and Szabo (1982), and in a different form also by Tropp (1980).  $S^2$  is the generalized order parameter which is the measure of the degree of spatial restriction of the internal reorientational motion with correlation time  $\tau_i$ . Note that in Lipari and Szabo's definition,  $S^2$  also contains the distance average, whereas in Eq. 17 the orientations and distances are averaged separately, because the orientations are assumed to fluctuate less than the distances. In the case of isotropic rotation of the molecule with correlation time  $\tau_o$ , the overall correlation function has a single exponential form:  $C_o(t) = \exp(-t/\tau_o)$ . If  $S^2 = 1$  the  $^1\text{H}-^1\text{H}$  vector is rigidly fixed to the molecule. If  $S^2 = 0$  the  $^1\text{H}-^1\text{H}$  vector is completely unrestricted and no distinction can be made between the two correlation times, so that again the motion is described by a single exponential:  $\exp(-t/\tau_c)$  with  $\tau_c^{-1} = \tau_o^{-1} + \tau_i^{-1}$ . The spectral densities that are used in the rate matrix are defined as the Fourier transform of the correlation function:

$$J_n(\omega) = \frac{1}{2} \int_{-\infty}^{\infty} C(t) e^{i\omega t} dt \quad (18)$$

and are in principle different for each individual  $^1\text{H}-^1\text{H}$  vector. After substitution of  $C(t)$  in Eq. 18 by the expression of Eq. 17, this becomes:

$$J_n(\omega) = (\langle r^{-6} \rangle - S^2 \langle r^{-3} \rangle^2) \frac{\tau_c}{1 + n^2 \omega^2 \tau_c^2} + S^2 \langle r^{-3} \rangle^2 \frac{\tau_o}{1 + n^2 \omega^2 \tau_o^2} \quad (19)$$

In order to make use of Eq. 19 the quantities  $S^2$ ,  $\tau_i$  and  $\tau_o$  have to be determined. In theory these values can be obtained from MD simulations.  $C(t)$  can be calculated using Eq. 14 for each  $^1\text{H}-^1\text{H}$  vector and then be fitted to Eq. 17. In practice, however, this cannot easily be accomplished, as is evident from the following cases using:

$$C(t) = [(\langle r^{-6} \rangle - S^2 \langle r^{-3} \rangle^2)] \exp(-t/\tau_c) + S^2 \langle r^{-3} \rangle^2 \exp(-t/\tau_o) \quad (20)$$

**(1)**  $\tau_i < \tau_o$

These conditions apply to a slowly tumbling macromolecule with rapid internal motions. The correlation function  $C(t)$  as determined from MD simulations will typically drop rapidly from a value of 1.0 to the plateau value of  $S^2$  due to the internal reorientations, but will not decrease any further on the MD time scale. Therefore, the form of  $C(t)$  allows the determination of  $S^2$  and possibly  $\tau_i$ , although noise makes the latter less accurate, but will not reveal  $\tau_o$ . In this limit the correlation function can be approximated by

$$C(t) = \langle r^{-3} \rangle^2 S^2 \exp(-t/\tau_o) \quad (21)$$

$S^2$  can be determined from a plot of  $C(t)$  and  $\tau_o$  is usually implicitly determined from a fit of NOEs of known  $^1\text{H}-^1\text{H}$  distances. This approach has been used by Koning et al. (1990).

**(2)**  $\tau_i \approx \tau_o$

This situation applies to a small molecule with fast internal motions. From the calculated  $C(t)$ ,  $S^2$  cannot be recognized because the function continues to drop due to the overall rotation. Therefore, both  $S^2$  and  $\tau_i$  cannot be determined. Two situations can be distinguished.

(a) If  $S^2$  is near 1.0 (0.6–1.0), which is the case for most fast internal motions, the last term in Eq. 20 is the most important so that  $C(t)$  and  $J_n(\omega)$ , are best described using  $\langle r^{-3} \rangle^2$ :

$$C(t) = \langle r^{-3} \rangle^2 [(1 - S^2) \exp(-t/\tau_c) + S^2 \exp(-t/\tau_0)] \quad (22)$$

It turns out that the term in square brackets can be fitted to a single  $\exp(-t/\tau_{fit})$  and that the corresponding  $J_n(\omega)$  is quite accurate:

$$J_n(\omega) = \langle r^{-3} \rangle^2 \frac{\tau_{fit}}{1 + n^2 \omega^2 \tau_{fit}^2} \quad (23)$$

The  $\tau_{fit}$  is an average correlation time that accounts for overall and internal motions of the  $^1\text{H}-^1\text{H}$  vector.

(b) If  $S^2$  is near 0.0 (0.0–0.4), the correlation function has a  $\langle r^{-6} \rangle$  character and can be written as:

$$C(t) = \langle r^{-6} \rangle [(1 - S^2) \exp(-t/\tau_c) + S^2 \exp(-t/\tau_0)] \quad (24)$$

Again, the latter term can be fitted to  $\exp(-t/\tau_{fit})$  and the spectral density is described by:

$$J_n(\omega) = \langle r^{-6} \rangle \frac{\tau_{fit}}{1 + n^2 \omega^2 \tau_{fit}^2} \quad (25)$$

The situations (a) and (b) deviate in the amount of internal mobility, which leads to a different distance averaging scheme in Eq. 23 and Eq. 25, respectively.

### (3) $\tau_i \gg \tau_0$

This situation applies to a small molecule, tumbling fast, with a hindered internal motion.  $S^2$  and  $\tau_i$  cannot be determined because the correlation drops to zero before the internal correlation function becomes effective. In other words, the two terms of  $S^2$  in Eq. 20 tend to cancel and:

$$C(t) = \langle r^{-6} \rangle \exp(-t/\tau_0) \quad (26)$$

From the above discussion it follows that spectral densities should be used for the individual  $^1\text{H}-^1\text{H}$  vectors. The specific internal mobility can be accounted for by means of either  $S^2$  or  $\tau_{fit}$  values, and by means of distance averaging by either  $\langle r^{-3} \rangle^2$  or  $\langle r^{-6} \rangle$ .

### Anisotropy

Up to now it has been assumed that the overall rotation of the molecule is isotropic, i.e.  $C_o(t)$  can be described by a single exponential. Incorporation of anisotropic rotation has been described mainly for symmetrical top molecules (Tropp, 1980; Lipari and Szabo, 1982). Rotation of a molecule of arbitrary shape can be described by 3 rotation correlation times with respect to the principal axes of inertia. The correlation function then becomes a sum of 5 exponential terms, and contains direction cosines of the individual  $^1\text{H}-^1\text{H}$  vectors with respect to the molecule fixed axes

(Woessner, 1962). The direction of the principal axes of inertia can be determined by diagonalizing the matrix of moments of inertia:

$$\mathbf{I}^D = \mathbf{X}^{-1} \mathbf{I} \mathbf{X} \quad (27)$$

The matrix  $\mathbf{X}$  contains the principal axes of the molecule in the laboratory-fixed axes system. The diagonal matrix  $\mathbf{I}^D$  contains the 3 moments of inertia with respect to the 3 principal axes. This analysis can be performed every MD time step. Consecutively, the direction cosines of every  $^1\text{H}$ - $^1\text{H}$  vector with respect to these 3 axes need to be determined.

A more practical approach when MD data are available, is to calculate  $C(t)$  for each individual  $^1\text{H}$ - $^1\text{H}$  vector directly from Eq. 14, which then implicitly contains the 3 rotational correlation times and the direction cosines. In addition, it contains the effect of internal fluctuations on the reorientation of the  $^1\text{H}$ - $^1\text{H}$  vector, characterized by  $S^2$  and  $\tau_i$ . Thus, rather than entering an overall  $\tau_c$  in the rate matrix, a matrix of  $\tau_c$  values is used leading to different  $J_n(\omega)$  for each  $^1\text{H}$ - $^1\text{H}$  vector.

## IMPLEMENTATION OF CROSREL

The foregoing theory has been implemented in the CROSREL program, for which the flow of execution has been summarized as pseudo-code in Fig. 2. The CROSREL program has been written in FORTRAN-77 and runs on VAX/VMS and SG-IRIS/UNIX computers. Currently the program can handle upto 100 protons. However, incrementing this number is easily possible. The program is started by reading parameter settings and options, including the type of experiment (ROESY/NOESY), which is needed in order to create the proper relaxation matrix  $\mathbf{R}$ . Subsequently,  $^1\text{H}$ - $^1\text{H}$  distances must be provided for which several options are available. On the basis of these data the theoretical peak intensities  $\mathbf{A}^{\text{calc}}$  can be calculated. Several corrections can be applied to  $\mathbf{A}^{\text{calc}}$  prior to comparison with the observed data. The agreement between the experimental and theoretical data is expressed in a weighted Residual ( $R_w$ ) factor. In the following paragraphs the main modules of the program will be discussed in more detail.

### *Input parameters and options*

CROSREL needs to know what kind of NMR experiments (NOESY/ROESY) have been done and, in the case of ROESY, whether correction for direct HOHAHA effects need to be applied. The scaling method must be entered that has been employed to the experimental cross-peak intensities, as well as the molecular tumbling model (isotropic or anisotropic). Within CROSREL a proton-proton distance matrix is used. This matrix can be obtained as a weighted average of different conformations which are entered either as distances or as coordinates. The weights, ( $p_m$ ), also need to be entered. For the calculation of peak intensities CROSREL needs to know  $\tau_c$  and  $R_L$  values and experimental NMR parameters ( $\omega$ ,  $\tau_m$ , carrier frequency and spin-lock field strength). Finally, the user has to select the peaks that are to be used for the comparison between theoretical and experimental intensities.

---

\* The CROSREL program is available from the authors: Leeflang: LEEF@RUUCMR.CHEM.RUU.NL Kroon-Batenburg: BATE@HUTRUU54.BITNET.

```

begin
  open files
  read input parameters
  read model coordinates/distances
  (read  $\tau_c$  factors)
  (read  $S^2$  factors)
  (read chemical shifts)
  (read coupling constants)
  close input files
  (calculate HOHAHA matrix)
  for all  $R_L$  values do
    for all  $\tau_c$  values do
      calculate R matrix
      for all  $\tau_m$  values do
        calculate A matrix
        perform offset correction
        perform scaling
        (perform HOHAHA correction)
        (handle overlap)
        calculate  $R_w$  factor
      od  $\tau_m$  values
      calculate  $R_w$  factor for all  $\tau_m$  values
    od  $\tau_c$  values
  od  $R_L$  values
  finish
  close files
end

```

Fig. 2. Pseudo-code for CROSREL calculations. Underlined statements are only valid for simulation of ROESY spectra peak intensities. Statements in parentheses are optional.

#### NMR- $R_w$ factor

A good way to judge the quality of the CROSREL calculation is a direct comparison of the theoretical  $A^{\text{calc}}(\tau_m)$ , and the observed  $A^{\text{obs}}(\tau_m)$  peak intensities. Several methods have been described which express the quality of fit in one figure (Breg et al. 1989; Borgias et al. 1990; Gonzalez et al. 1991). For CROSREL a weighted R factor similar to X-ray crystallography (Dunitz, 1979) has been chosen, which is rather a measure of relative than of absolute errors as a result of the weights used.

$$R_w = \left\{ \frac{\sum_{i=1}^n \sum_{j=1}^n \sum_{\tau_m}^m w_{ij}(\tau_m) \cdot (A_{ij}^{\text{calc}}(\tau_m) - A_{ij}^{\text{obs}}(\tau_m))^2}{\sum_{i=1}^n \sum_{j=1}^n \sum_{\tau_m}^m w_{ij}(\tau_m) \cdot (A_{ij}^{\text{obs}}(\tau_m))^2} \right\}^{1/2} \quad (28)$$

The weights  $w_{ij}(\tau_m)$  are defined as

$$w_{ij}(\tau_m) = \frac{1}{A_{\text{noise}} + |A_{ij}^{\text{obs}}(\tau_m)|} \quad (29)$$

wherein  $A_{\text{noise}}$  is the estimated background error of peak integration and depends on  $\tau_m$ . The weight function of Eq. 29 emphasizes the contribution of low-intensity (interresidue) cross peaks, which are often determinative for the structure of the molecule. Overestimation of the importance of these cross peaks, however, is prevented by the incorporation of  $A_{\text{noise}}$  in order to set a maximum weight. Further, it should be mentioned that only cross peaks should be used in  $R_w$  factor calculations that have been determined with enough accuracy. A useful criterion in this respect is:

$$|A_{ij}^{\text{obs}}(\tau_m)| > \frac{2}{\sqrt{w_{ij}(\tau_m)}} \quad (30)$$

which is also being used in X-ray crystallography. It is completely up to the CROSREL user to decide which cross peak are to be used for  $R_w$  calculation.

#### *Distance averaging*

In view of the foregoing discussion in the *Internal motion* section, the CROSREL program is capable of performing different distance averaging schemes. MD trajectories of  $^1\text{H}$  positions can be entered into the program and either  $r = [\langle r^{-6} \rangle]^{-1/6}$  or  $r = [\langle r^{-3} \rangle]^{-1/3}$  averaging can be chosen. Then, such a distance matrix can be saved. Distance matrices from different trajectories can subsequently be averaged as  $r = [\langle p_m r^{-6} \rangle_m]^{-1/6}$ , wherein  $p_m$  is the contribution of distance matrix  $m$  in the final distance matrix. Thus only fixed molecular models can be input to the program and only their relative contributions  $p_m$  can be adjusted, although not automatically. This type of averaging is chosen because usually different trajectories represent configurations that are separated by a longer times span than the length of the MD trajectories themselves. These distance matrices can again be entered into CROSREL.

#### *Anisotropic and internal motion*

The general spectral density function used in CROSREL is:

$$J_n(\omega) = r^{-6} \cdot \frac{\tau_c}{1 + n^2 \omega^2 \tau_c^2} \cdot S^2 \quad (31)$$

Either  $\langle r^{-3} \rangle^2$  or  $\langle r^{-6} \rangle$  can be substituted for  $r^{-6}$  in this equation. Isotropic  $\tau_c$  values or individual  $\tau_{\text{fit}}$  values and additionally individual  $S^2$  values can be entered into the program (see *Internal motion* section).

#### *Offset dependence*

When ROESY peak intensities are calculated, the CROSREL program incorporates the offset effect completely in the theoretical matrix  $A$ . In order to perform the offset correction the carrier

frequency, the spectrometer frequency and the complete assignment ( $\delta$ ) must be entered into the program. The experimental data remain unchanged.

### *Scaling of the peak intensities*

The observed diagonal and cross-peak intensities contain an arbitrary scaling factor, due to all sorts of experimental conditions, and therefore have only a relative significance. The calculated peak intensities are in principle related to the initial magnetisation  $\mathbf{M}_0$  at  $\tau_m = 0$ . The sum of peak intensities in one direction (e.g.  $\omega_2 = \text{constant}$ ) is  $\mathbf{M}_0 \cdot \exp(-\tau_m/T_1)$ . In order to make possible comparisons between observed and calculated data, a scale factor has to be applied. Three scaling schemes can be applied in CROSREL:

$$\mathbf{M}_0\text{-scaling: } A'_{ij}(\tau_m) = \frac{A_{ij}(\tau_m)}{\sum_i A_{ij}(\tau_m = 0)} \quad (32)$$

This type of scaling is applied to the observed intensities before they enter the CROSREL program, and to the calculated intensities in the program itself. The scaling is applied to each column  $j$  of data (i.e.  $\omega_2 = \text{constant}$ ). For each proton the value of  $\mathbf{M}_0 = \sum A_{ij}(\tau_m = 0)$  is estimated by extrapolation of the total magnetization ( $\sum A_{ij}(\tau_m)$ ) to  $\tau_m = 0$ . The curve of  $\sum A_{ij}(\tau_m)$  is fitted to  $\mathbf{M}_0 \cdot \exp(-\tau_m/T_1)$ . For the calculated intensities  $\mathbf{M}_0$  would be equal to 1.0,<sup>1</sup> so that this scaling would be unnecessary. However, because of the offset dependence in case of ROESY, a scaling factor of approximately  $\sin^2\theta_j$  appears in  $A_{ij}$ , already at  $\tau_m = 0$ . Since through the scaling a  $\sin^2\theta_j$  factor is removed from the experimental intensities, a similar scaling has also to be applied the calculated intensities. This type of scaling is only applicable in the analysis of build-up series, since  $\mathbf{M}_0$  is estimated by extrapolation. When only one experiment is done, this extrapolation is not possible and  $\mathbf{M}_1$ -scaling and Obs-scaling are alternatives.

$$\mathbf{M}_1\text{-scaling: } A'_{ij}(\tau_m) = \frac{A_{ij}(\tau_m)}{\sum_i A_{ij}(\tau_m)} \quad (33)$$

This type of scaling is applied to the observed intensities before they are entered into the CROSREL program, and to the calculated intensities in the program itself. The total magnetization at each mixing time is scaled to 1.0. This means in fact, that all intensities are multiplied by  $\exp(+\tau_m/T_1)$  (apart from offset factors in case of ROESY). Similar to the  $\mathbf{M}_0$ -scaling method, the calculated  $\mathbf{M}(\tau_m) = \sum A_{ij}(\tau_m)$  is not exactly equal to  $\exp(-\tau_m/T_1)$  in case of ROESY because of offset effects. A danger, however, of this  $\mathbf{M}_1$ -scaling is that the sum of the positive diagonal peak and the negative cross peaks might become close to zero, resulting in large errors in the scaled peak intensities. The CROSREL program also offers a modification of this method that uses the absolute values of the peak intensities to determine the scaling factors.

The idea behind scaling the calculated intensities is that the calculated peak intensities have to resemble the experimental ones as closely as possible. However, because of the arbitrariness of the scaling factor in the observed intensities, the main scaling has to be applied to the observed data.

Most biomolecules have complex  $^1\text{H}$  NMR spectra with much overlap, which prevents the determination of specific diagonal peak intensities. For this reason most NMR conformational analyses are based on cross peaks only. The following scaling scheme can then be used.

$$\text{Obs-scaling: } A_{ij}^{\text{obs}} = \text{o.s.f.} * A_{ij}^{\text{obs}} \quad (34)$$

With Obs-scaling an overall scaling factor, o.s.f., is determined, which scales the observed intensities in order to minimize the differences between observed and calculated peak intensities. Obs-scaling minimizes expression (35).

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{\tau_m}^m (w_{ij}(\tau_m) * A_{ij}^{\text{calc}}(\tau_m) - \text{o.s.f.} * w_{ij}(\tau_m) * A_{ij}^{\text{obs}}(\tau_m)) \quad (35)$$

The weights  $w_{ij}(\tau_m)$  are defined in Eq. 29 and emphasise low-intensity cross peaks. The scaling takes place in the CROSREL program and no scaling of the raw data is needed to prior to the calculation. Either  $M_0$ - or  $M_1$ -scaling are preferable because information of the whole spectrum is used. On the other hand Obs-scaling is more generally applicable. Therefore, it is important to check the compatibility of these 3 scaling methods.

#### *HOHAHA correction*

A general problem in the analysis of ROESY spectra is the occurrence of HOHAHA type magnetization transfer. This transfer can result in spurious cross peaks by relay of cross-relaxed magnetization (Neuhauser and Keeler, 1986), and in attenuation of the cross-peak intensity due to direct HOHAHA transfer between coupled spins that are also closely spaced. The effects of the direct HOHAHA transfer are expected to complicate quantitative analysis of ROESY more than relay effects, since the latter are the products of two transfer steps. Therefore, only the direct HOHAHA transfer is accounted for in CROSREL. Thus, relay effects are neglected. In the *Theory* section a method has been described, introduced by Bax (1988), that estimates the HOHAHA transfer in a two-spin system. Most spin systems in biomolecules consist of more than two spins, and therefore, the CROSREL program must be able to estimate HOHAHA transfer in multi-spin systems. The program assumes that the different HOHAHA transfers are independent and that they only influence each other through the faster decrease of the diagonal magnetization. The multi-spin procedure first calculates raw estimates, based on two-spin interactions without any offset dependence:

$$H_{ii} = 1$$

$$H_{ij} = \frac{s_{ij}^2}{(1 + c_{ij}^2)} \quad (36)$$

In comparing Eq. 36 to Eq. 13, it should be noted that all relevant information is stored in the off-diagonal elements. Then the HOHAHA estimates are normalized by  $\sum H_{ij} = 1$ . The HOHAHA matrix thus obtained can be used to incorporate the HOHAHA effect in the calculated ROESY

peaks as follows:

$$A_{ii}^{\text{HHH}} = A_{ii} \cdot H_{ii}$$

$$A_{ij}^{\text{HHH}} = A_{ij} + A_{ii} \cdot H_{ij} \quad (37)$$

In addition to the complete assignment, which must also be provided for the offset correction, all significant scalar coupling constants ( $J_{ij}$ ) must be entered into CROSREL.

### *Spectral overlap*

Overlap of cross peaks complicates the analysis of peak intensities. The definition of overlap in NOESY spectra is straightforward. In ROESY spectra, however, signals that are almost overlapping and are coupled to each other, should also be assumed to overlap due to the strong HOHAHA transfer that cannot be corrected for. The CROSREL program handles overlapping proton signals as the spectroscopist is forced to handle overlap in the spectra. The diagonal peak and the mutual cross-peak intensities of the overlapping protons are summed and stored as the diagonal intensities of each of these protons. The remaining cross-peak intensities in both rows and columns of the theoretical matrix of these overlapping protons are also added and stored in each column/row. This procedure allows direct comparison with the observed data.

## METHODS

### *NMR spectroscopy*

2D ROESY spectra of methyl(*d*3)  $\beta$ -cellobioside in D<sub>2</sub>O solution were recorded at 283 K at 600 MHz with a Bruker AM-600 spectrometer equipped with a dedicated Aspect 3000 computer (SON hf NMR facility, Nijmegen). A continuous low-power pulse was applied as spin-lock pulse in all experiments. The carrier frequency was placed at 5.75 ppm (left of the spectrum), and the power of the spin-lock pulse corresponded to a 90° pulse width of 100  $\mu$ s. Time Proportional Phase Increments, (TPPI), were used in order to obtain phase-sensitive spectra. Mixing times were 50, 100, 150, 200, 300, or 400 ms.

The processing of the spectra was performed on a  $\mu$ VAX/VMS cluster with the 'TRITON' NMR software package (R. Kaptein, R. Boelens; Department of NMR Spectroscopy, Utrecht University). Prior to Fourier transformation the spectra were multiplied with a  $\pi/2$  shifted sine bell function in both dimensions. Third-order polynomial baseline corrections were applied after each Fourier transformation in order to provide a flat baseplane needed for accurate cross-peak integration. The ROESY cross peaks are integrated by simple summation of the intensities within a defined rectangle around the peak. Overlapping peaks or strongly coupled, almost overlapping peaks (H3, H4, and H5 of both glucose residues) were integrated to obtain only one, overall value.

### *MD simulation*

The molecular dynamics (MD) simulations on methyl  $\beta$ -cellobioside in water were performed using the program GROMOS (Van Gunsteren, 1987) and its standard force-field for carbohydrates (Koehler et al. 1987). The united atom approach was used for aliphatic carbon atoms. Positions of the corresponding hydrogen atoms were calculated after the simulation, using ideal tetra-

hedral geometries and C-H bond distances of 1.1 Å. Methyl protons were not generated because the experimental data are obtained for the methyl deuterated compound. The methyl  $\beta$ -cellobioside molecule was placed in a periodic computational box containing 358 water molecules. The simulation was performed at constant temperature (300 K) and pressure (1 atm). One uninterrupted MD run of 500 ps was performed (Leefflang et al. 1992). Trajectories of the methyl  $\beta$ -cellobioside molecule have been stored every 0.02 ps. The calculations were performed at a local  $\mu$ VAX/VMS cluster and on a Convex 120.

## RESULTS AND DISCUSSION

The aim of CROSREL calculation is to arrive at a model for the molecule under investigation by reproducing the observed NOE/ROE spectra as good as possible. The proposed strategy is to optimize input parameters by evaluating cross peaks of protons at known distances, prior to modifying torsional angles in the model. On the basis of these distances and parameters the effects of overall motion, of scaling methods, and HOHAHA correction can be assessed.

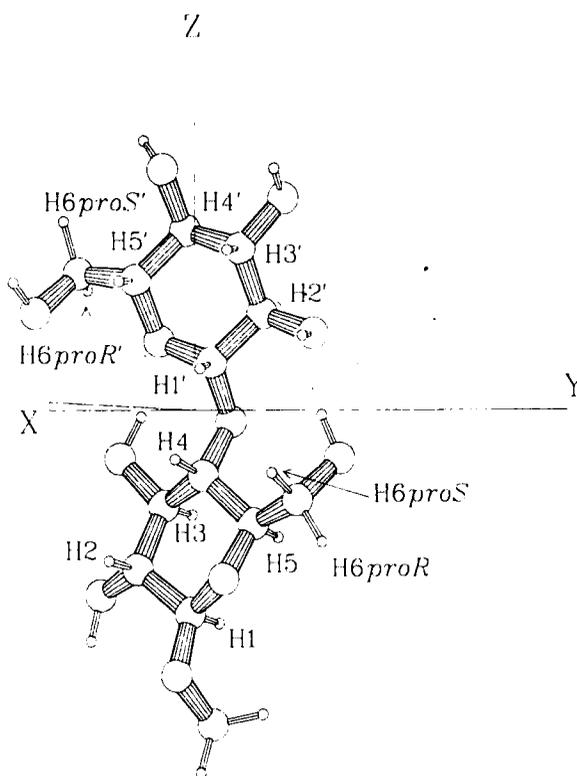


Fig. 3. Molecular model of methyl  $\beta$ -cellobioside. Atom labels and torsional angles follow the JCBN definitions for carbohydrates (IUPAC-IUB JCBN, 1983). The conformation of the hydroxymethyl group is defined by a two-character code (GG, GT, and TG), indicating the orientation of the torsion angles  $\tau(\text{O6-C6-C5-O5})$  and  $\tau(\text{O6-C6-C5-C4})$ , respectively (G means a *gauche* and T a *trans* orientation). Additionally, the orientations of the 3 principal axes of inertia have been indicated.

### Model

A rigorous discussion of conformational aspects obtained from the MD and ROESY data can be found in Kroon-Batenburg et al. (submitted) and Leeftang (1991). Here it suffices to mention that virtually no transitions occurred with respect to the glycosidic torsional angles ( $\phi, \psi$ ), and that several transitions of both hydroxymethyl groups occurred. The proton-proton distances were derived from two parts of the trajectory: one in which both hydroxymethyl groups were in the GG conformation (100 ps of data) and the other with both hydroxymethyl groups in the GT conformation (35 ps of data). The length of these parts of the trajectory may not be ideal, but it suffices for the present purpose, namely to calculate intraresidue proton-proton distances that can be used to test CROSREL. In Fig. 3 a model for methyl  $\beta$ -cellobioside is displayed. In the discussion of the conformation of methyl  $\beta$ -cellobioside a more accurate model will be generated (Kroon-Batenburg et al., submitted; Leeftang, 1991). The two distance matrices from the above trajectories, were each obtained by  $r = [\langle r^{-3} \rangle]^{-1/3}$ . It is estimated that  $\tau_0$  has a value of 70 ps at maximum (see below), and that the internal motions in these trajectories are rather limited ( $S^2 = 0.8-1.0$ ) and have a  $\tau_i$  of approximately 10 ps, so the approximation (2a) for the correlation function (Eq. 22) and spectral densities (Eq. 23) applies. The final distance matrix that is used as input to CROSREL is subsequently obtained by weighted averaging of the two distance matrices by  $r = [\langle r^{-6} \rangle]^{-1/6}$ . This was done because again the internal motions corresponding to the hydroxy-

TABLE I  
INTRARESIDUE DISTANCES (IN Å), ABOVE THE DIAGONAL, AND  $\tau_c$  FACTORS, BELOW THE DIAGONAL, USED FOR CROSREL CALCULATIONS, AS DERIVED FROM MD SIMULATIONS OF METHYL  $\beta$ -CELLOBIOSIDE IN WATER

	H1	H2	H3	H4	H5	H6 <sub>proS</sub>	H6 <sub>proR</sub>
H1	–	3.039	2.657	3.991	2.303	4.210	4.513
H2	0.6150	–	3.013	2.746	3.901	5.143	4.841
H3	0.7575	0.5509	–	3.021	2.470	4.508	4.476
H4	0.7215	0.7779	0.4747	–	3.032	3.232	2.867
H5	0.7586	0.6309	0.5341	0.6155	–	2.485	2.561
H6 <sub>proS</sub>	0.5774	0.6069	0.4780	0.5327	0.4416	–	1.796
H6 <sub>proR</sub>	0.5838	0.5852	0.4540	0.6956	0.4399	0.4636	–
	H1'	H2'	H3'	H4'	H5'	H6 <sub>proS</sub> '	H6 <sub>proR</sub> '
H1'	–	3.041	2.592	3.954	2.342	4.285	4.535
H2'	0.6150	–	3.024	2.712	3.945	5.068	4.652
H3'	0.7575	0.5509	–	3.024	2.606	4.576	4.490
H4'	0.7215	0.7779	0.4747	–	3.035	3.066	2.678
H5'	0.7586	0.6309	0.5341	0.6155	–	2.482	2.585
H6 <sub>proS</sub> '	0.5774	0.6069	0.4780	0.5327	0.4416	–	1.796
H6 <sub>proR</sub> '	0.5838	0.5852	0.4540	0.6956	0.4399	0.4636	–

The distance matrix has been obtained by averaging distances matrices for 2-GG and 2-GT as  $r = [\langle r^{-6} \rangle]^{-1/6}$ . The ratio GG:GT=0.55:0.45 was used. The  $\tau_c$  factors have been obtained by averaging the corresponding scaling factors for each residue. The coupling constants (Hz) entered into the CROSREL program were identical for each residue: – 12.0 (H6<sub>proS</sub>-H6<sub>proR</sub>), 2.3 (H5-H6<sub>proS</sub>), 5.6 (H5-H6<sub>proR</sub>), 8.0 (H1-H2), and 9.0 (H2-H3, H3-H4, H4-H5).

methyl groups were on the same time scale (about 30 ps) as the overall rotation, but now  $S^2$  was smaller ( $\approx 0.5$ ) and approximation (2b) applied. The weights used for GG and GT were 0.55 and 0.45, respectively, which correspond to the average values of the rotamer population distribution of both hydroxymethyl groups, as has been determined from vicinal coupling constants (Leefflang, 1991). Intraresidue distances are listed in the upper triangles of Table 1.

#### *Parameter optimization*

Prior to modifying conformational parameters of the input model supplied to the CROSREL program, two important parameters,  $R_L$  and  $\tau_c$ , have to be determined. In this paper these two parameters are optimized. In a different paper (Kroon-Batenburg et al., submitted; Leefflang, 1991) the input molecular model is adapted so as to find the best agreement with the observed peak intensities, given  $R_L$  and  $\tau_c$ . The leakage rate  $R_L$  cannot be obtained experimentally. The rotational correlation time  $\tau_c$ , however, can be estimated on the basis of  $^{13}\text{C}$  NMR  $T_1$  experiments. The result, however, does not always provide an optimal fit between observed and theoretical data (Breg et al. 1989). Fortunately,  $R_L$  and  $\tau_c$  can be determined with the aid of the CROSREL program itself, using known distances. Obviously, interpretation of NOEs of protons at unknown distances can only be done when the input parameters produce a good fit of NOEs of protons at known distances.

In order to find optimal values for  $\tau_c$  and  $R_L$ , systematic CROSREL grid searches were performed, wherein  $\tau_c$  ranged from 10 to 500 ps with a 10-ps increment, and  $R_L$  ranged from 0.0 to  $1.0\text{ s}^{-1}$  with a  $0.05\text{ s}^{-1}$  increment.  $R_w$  factors are determined on a set of intraresidue peaks, with distances that are obtained from MD simulations, and are considered sufficiently reliable to be used as known distances. Separate CROSREL grid searches were performed with 3 different scaling methods ( $M_0$ -scaling,  $M_1$ -scaling, or Obs-scaling), and with or without correction for the direct HOHAHA effect. In the case of Obs-scaling, the o.s.f. is determined in addition to  $R_L$  and  $\tau_c$ . This procedure was followed for assumed isotropic tumbling with a uniform  $\tau_c$ , and for assumed anisotropic tumbling with relative  $\tau_c$  values ( $\tau_{\text{fit}}$ ) derived from MD simulations (see below). The  $\tau_c$  factors of equivalent  $^1\text{H}$ - $^1\text{H}$  vectors in both residues have been averaged and are listed in the lower triangles of Table 1. In total, 12 grid searches were accomplished. The H3, H4, and H5 atoms of each glucose residue were treated as overlapping. The quality of fit between calculated and experimental data is expressed in the  $R_w$  factor of Eq. 28, based however, on intraresidue ROE interactions only (H1-H3/4/5, H3/4/5-H6proS, H3/4/5-H6proR, and H6proS-H6proR cross peaks of both residues). The noise contribution  $A_{\text{noise}}$  in the weight factors was estimated to 0.5% of the initial magnetization  $M_0$ .

#### *HOHAHA correction*

Despite carefully chosen experimental conditions, HOHAHA effects in ROESY spectra can never be fully excluded, since in principle there is no experimental difference between ROESY and HOHAHA spectroscopy. The CROSREL program performs correction for the intensity attenuation of a cross peak between two coupled protons by HOHAHA magnetization transfer. The HOHAHA effects estimated by the CROSREL program are summarized in Table 2 as normalized cross-peak intensities. The cross and diagonal peaks of the H3, H4, and H5 atoms of each glucose residue are summed, since their resonances overlap or almost overlap. The experimental conditions accomplish a fairly good suppression of HOHAHA type transfer, although the effect

TABLE 2  
 NORMALIZED ESTIMATED HOHAHA EFFECTS (WITHOUT OFFSET EFFECT) IN ROESY SPECTRA OF METHYL(*d*3)  $\beta$ -CELLOBIOSIDE AT 600 MHz AND CARRIER FREQUENCY AT 5.75 ppm AND SPIN-LOCK FIELD STRENGTH OF 2500 Hz (H3, H4, and H5 ATOMS OF BOTH RESIDUES ARE CONSIDERED OVERLAPPING)

	H1	H2	H3/4/5	H6 <i>pro</i> S	H6 <i>pro</i> R
H1	0.9995	0.0005	0	0	0
H2	0.0005	0.9945	0.0101	0	0
H3/4/5	0	0.0050	0.9887	0.0024	0.0037
H6 <i>pro</i> S	0	0	0.0005	0.9537	0.0438
H6 <i>pro</i> R	0	0	0.0008	0.0439	0.9525
	H1'	H2'	H3/4/5'	H6 <i>pro</i> S'	H6 <i>pro</i> R'
H1'	0.9996	0.0004	0	0	0
H2'	0.0004	0.9892	0.0032	0	0
H3/4/5'	0	0.0104	0.9958	0.0002	0.0030
H6 <i>pro</i> S'	0	0	0.0001	0.9669	0.0328
H6 <i>pro</i> R'	0	0	0.0009	0.0329	0.9642

between H6*pro*S and H6*pro*R atoms of both residues is in the order of 4% of  $M_0$ . This result implies an important role of the HOHAHA correction for ROESY peak intensities. The  $R_w$  factors listed in Table 3 indeed are considerably lower when the HOHAHA correction is applied, irrespective of the scaling method or tumbling model. The cause of this large difference in  $R_w$  was evaluated by determining the optima for each mixing time separately. The  $\tau_m$ -specific optima are

TABLE 3  
 CROSREL  $R_L$ ;  $\tau_c$  GRID SEARCH RESULTS FOR A ROESY BUILD-UP SERIES FOR METHYL(*d*3)  $\beta$ -CELLOBIOSIDE<sup>a</sup>

	No HOHAHA correction			HOHAHA correction		
	$R_L$	$\tau_c$	$R_w$	$R_L$	$\tau_c$	$R_w$
<i>M<sub>0</sub>-scaling</i>						
Isotropic	0.45	110	0.1103	0.20	150	0.0647
Anisotropic	0.45	220	0.1001	0.20	290	0.0650
<i>M<sub>1</sub>-scaling</i>						
Isotropic	0	120	0.1033	0	140	0.0773
Anisotropic	0	240	0.1029	0	280	0.0920
<i>Obs-scaling</i>						
Isotropic	1.0	10	0.2961	0.35	120	0.1920
Anisotropic	0.9	10	0.1779	0.0	300	0.2145

<sup>a</sup>  $R_L$  is expressed in  $s^{-1}$  and  $\tau_c$  in ps.  $R_w$  is calculated from intrasidue cross peaks only.  $R_w$  values for  $M_0$ - and  $M_1$ -scaling are based on cross and diagonal peaks, whereas for Obs-scaling only cross peaks are used.

marked in Fig. 4, and are listed in Table 4 together with values  $\langle R_L \rangle$ , and  $\langle \tau_c \rangle$ , and their absolute and relative errors. When the sum of the relative errors is considered a good measure for the deviation of the optima, it is evident that the deviation is significantly less when HOHAHA correction has been applied. It is therefore concluded that the HOHAHA correction is essential for the analysis of ROESY data, even when the experimental conditions are carefully chosen to minimize HOHAHA type transfer.

#### *Analysis of the overall tumbling*

With the final distance matrix (see *Model* section) the ROEs can be calculated. As far as internal rotations are concerned, assuming a rigid isotropically rotating molecule, a  $\tau_c$  value can be obtained by fitting the calculated ROEs to observed ROEs for intraresidue proton–proton distances. This will be referred to as the *isotropic model*. Alternatively,  $\tau_c$  can in principle be obtained from the MD simulations. However, analysis of rotation of the 3 principal axes of inertia, following Eq.

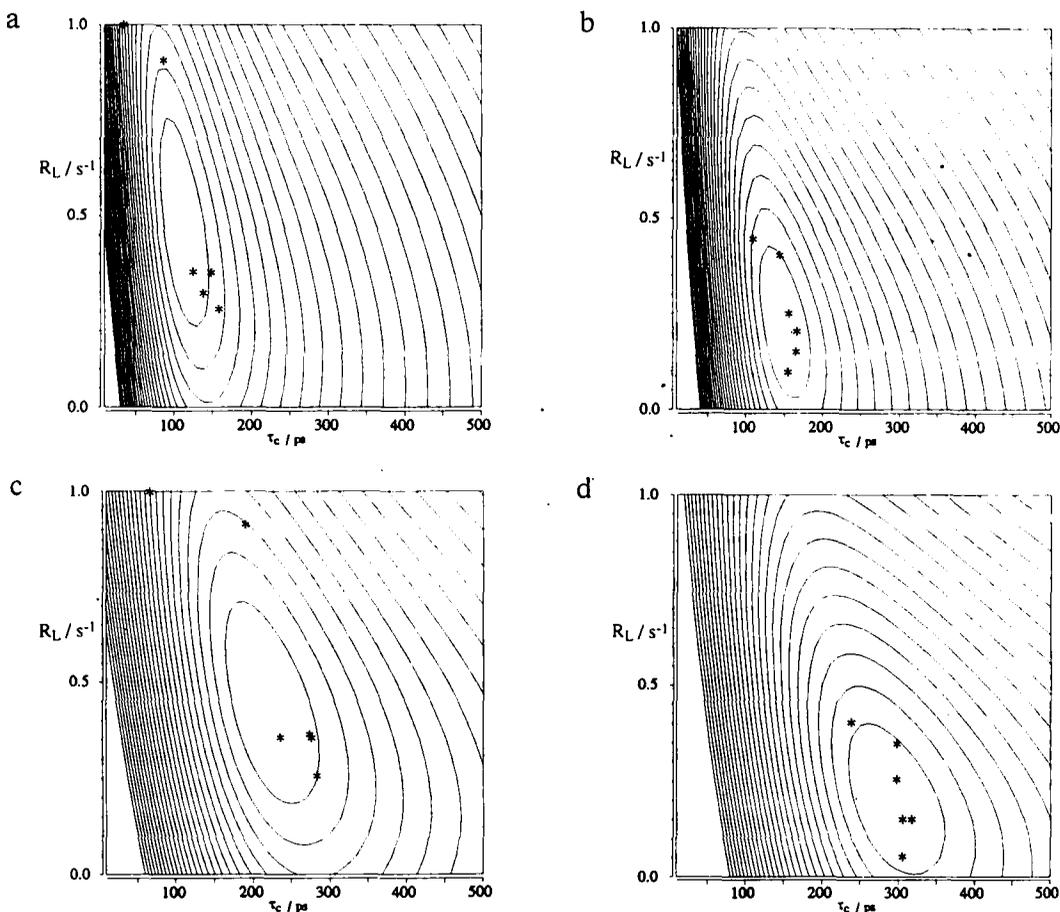


Fig. 4. CROSREL  $R_L$ ;  $\tau_c$  grid search results for a ROESY build up series of methyl( $d_3$ )  $\beta$ -cellobioside using  $M_0$ -scaling. The iso-contour lines are separated by 0.01 in  $R_w$ . (a) Isotropic model without HOHAHA correction; (b) isotropic model with HOHAHA correction; (c) anisotropic model without HOHAHA correction; and (d) anisotropic model with HOHAHA correction. The  $\tau_m$ -dependent optima have been marked with asterisks.

TABLE 4  
 $\tau_m$ -SPECIFIC CROSREL  $R_L$ ;  $\tau_c$  GRID SEARCH RESULTS FOR A ROESY BUILD-UP SERIES FOR METHYL(*d3*)  
 $\beta$ -CELLOBIOSIDE

$M_0$ -scaling	No HOHAHA correction			HOHAHA correction		
	$R_L$	$\tau_c$	$R_w$	$R_L$	$\tau_c$	$R_w$
<i>Isotropic</i>						
$\tau_m=0.05$	1.00	20	0.111	0.45	110	0.073
$\tau_m=0.10$	0.90	90	0.073	0.40	140	0.049
$\tau_m=0.15$	0.35	120	0.075	0.10	150	0.052
$\tau_m=0.20$	0.40	130	0.071	0.20	160	0.054
$\tau_m=0.30$	0.40	140	0.069	0.25	150	0.059
$\tau_m=0.40$	0.25	150	0.085	0.15	160	0.075
<..>	0.55	108	-	0.26	145	-
$\sigma_{n-1}$	0.32	48	-	0.14	19	-
$\sigma_{n-1}/<..>$	0.57	0.44	-	0.54	0.13	-
<i>Anisotropic</i>						
$\tau_m=0.05$	1.00	60	0.107	0.40	230	0.073
$\tau_m=0.10$	0.90	180	0.070	0.35	290	0.054
$\tau_m=0.15$	0.35	230	0.073	0.05	300	0.056
$\tau_m=0.20$	0.35	270	0.070	0.15	310	0.057
$\tau_m=0.30$	0.35	270	0.065	0.25	290	0.055
$\tau_m=0.40$	0.25	280	0.086	0.15	300	0.074
<..>	0.53	215	-	0.23	287	-
$\sigma_{n-1}$	0.33	85	-	0.13	29	-
$\sigma_{n-1}/<..>$	0.62	0.40	-	0.57	0.10	-

<sup>a</sup>  $R_L$  is expressed in  $s^{-1}$  and  $\tau_c$  in ps.  $R_w$  is calculated on the basis of intraresidue cross peaks only.

27 and calculating  $C(t)$  from Eq. 14 for each of these axes, revealed that the rotation is essentially anisotropic. The correlation functions for the molecular x-, y- and z-axes are shown in Fig. 5. The orientation of the axes is shown in Fig. 3. Rotation around one axis is equivalent to the rotation of the axes perpendicular to this axis. It can be seen that  $C(t)$  drops slower for the z than for the other two axes. These results lead to a picture in which the methyl  $\beta$ -cellobioside molecule behaves like a cylinder, rotating fast around the long z-axis and slower around the two other axes. The correlation functions, although averaged over 500 ps, show a lot of noise.  $C(t)$  is shown only up to 100 ps, but then the effect of noise is already visible. Fits of  $C(t)$  to  $\exp(-t/\tau_c)$  for data between 0 and 50 ps, lead to  $\tau_c$  values for the x-, y- and z-axes of 40, 38 and 65 ps, respectively. This means that a  $^1H$ - $^1H$  vector aligned along the z-axis would experience a  $\tau_c$  that is almost twice as large as  $^1H$ - $^1H$  vector aligned along the x- or y-axis. This led us to introduce anisotropy in the spectral density functions. As discussed before it is the most practical to evaluate  $C(t)$  for every individual  $^1H$ - $^1H$  vector directly from Eq. 14, thus incorporating both anisotropy and internal motion. The  $\tau_c$  values found as  $\tau_{fit}$  values for every  $^1H$ - $^1H$  vector were used as relative  $\tau_c$  factors in CROSREL, thus still leaving room to fit a  $\tau_c$  value by comparing calculated with observed ROEs for intraresidue distances. This model will be referred to as the *anisotropic model*. Application of anisotropic

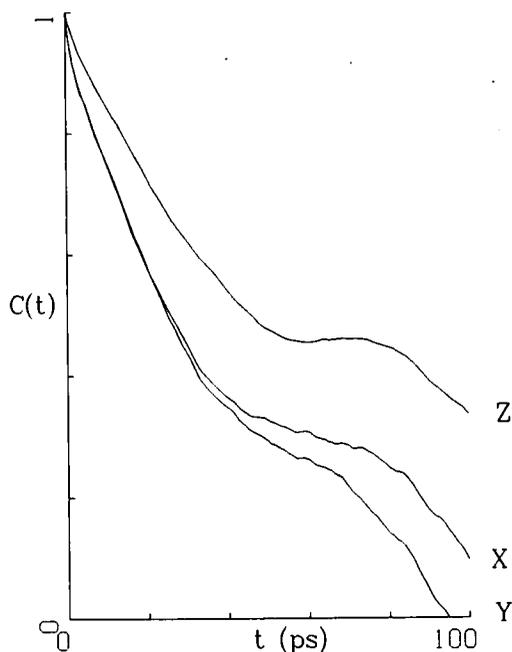


Fig. 5. Graphical representation of the correlation functions of the 3 principal axes of inertia of methyl  $\beta$ -cellobioside in water, as determined from MD simulations. The x-axis is assigned to the axis with the largest moment of inertia and the z-axis is that with the smallest.

$\tau_c$  factors in the CROSREL grid searches did not yield the expected improvement, as is evident from the results listed in Table 3. The  $\tau_c$  factors obtained for each  $^1\text{H}$ - $^1\text{H}$  vector are listed in Table 1. Multiplying these factors with 65 ps leads to the actually obtained  $\tau_c$  values, which range from 23.8 to 71.9 ps. Since most of the shorter  $^1\text{H}$ - $^1\text{H}$  vectors have  $\tau_c$  factors of about 0.5, this leads to  $\tau_c$  values obtained by a fit to ROEs that are about twice as large as the  $\tau_c$  from the isotropic model (see Tables 1 and 3). In the case of the isotropic model,  $\tau_c$  obtained from the ROESY peak intensities is 2–3 times larger than that obtained from the MD simulation. For the anisotropic model these  $\tau_c$ 's are 4–5 times larger. It seems, therefore, that both the overall and internal motions are somewhat exaggerated in the MD simulation. First it should be noted that specific MD parameters may have an effect on the correlation function, and therefore on the  $\tau_c$ . These effects, however, are unknown. More likely causes for the observed  $\tau_c$  discrepancies are the different temperatures at which the ROESY experiments have been carried out (283 K) and at which the MD simulations have been performed (300 K). It is evident that at lower temperatures the mobility of both the solvent and the solute will also be lower, which yields larger  $\tau_c$  values. An additional cause is the usage of  $\text{H}_2\text{O}$  in the MD simulations, whereas  $\text{D}_2\text{O}$  has been used in the ROESY experiments. Therefore, the molecular masses of the solvent and of methyl  $\beta$ -cellobioside will be higher in the ROESY experiments, due to exchanged hydroxyl protons, which both will result in larger  $\tau_c$  values. The disagreement of theory and experiment with respect to  $\tau_c$  is not worrisome, since only relative mobilities from MD are intended to be used.

### Scaling methods

$M_0$ -scaling is regarded to be the optimal scaling method, because all spectral information is used in the analysis.  $M_0$ -scaling, however, will often not be possible, due to lack of measuring time and/or the complexity of the spectra, disallowing the acquisition of build-up series and/or specific diagonal peak integration, respectively. The first alternative is the  $M_1$ -scaling, where the integrated column intensities are set to unity for each mixing time. An advantage of  $M_1$ -scaling is that the effect of  $R_L$  is fully cancelled, so that only  $\tau_c$  has to be optimized. Therefore, the CROSREL calculations are performed with  $R_L = 0$ . Although the obtained  $R_w$  factors are slightly higher than those for  $M_0$ -scaling, the overall outcome is similar, maintaining a preference for the isotropic model. The  $\tau_c$  value of the optimum, however, has shifted slightly to smaller values.

Obs-scaling is the second alternative, which does not require the measurement of diagonal peaks. The  $R_w$  factors obtained for Obs-scaling, as given in Table 3, are significantly higher than those for scaling methods that include high-intensity diagonal peaks. This is a direct effect of the lack of high-intensity peaks, and is not an indication of low quality of the scaling method. A

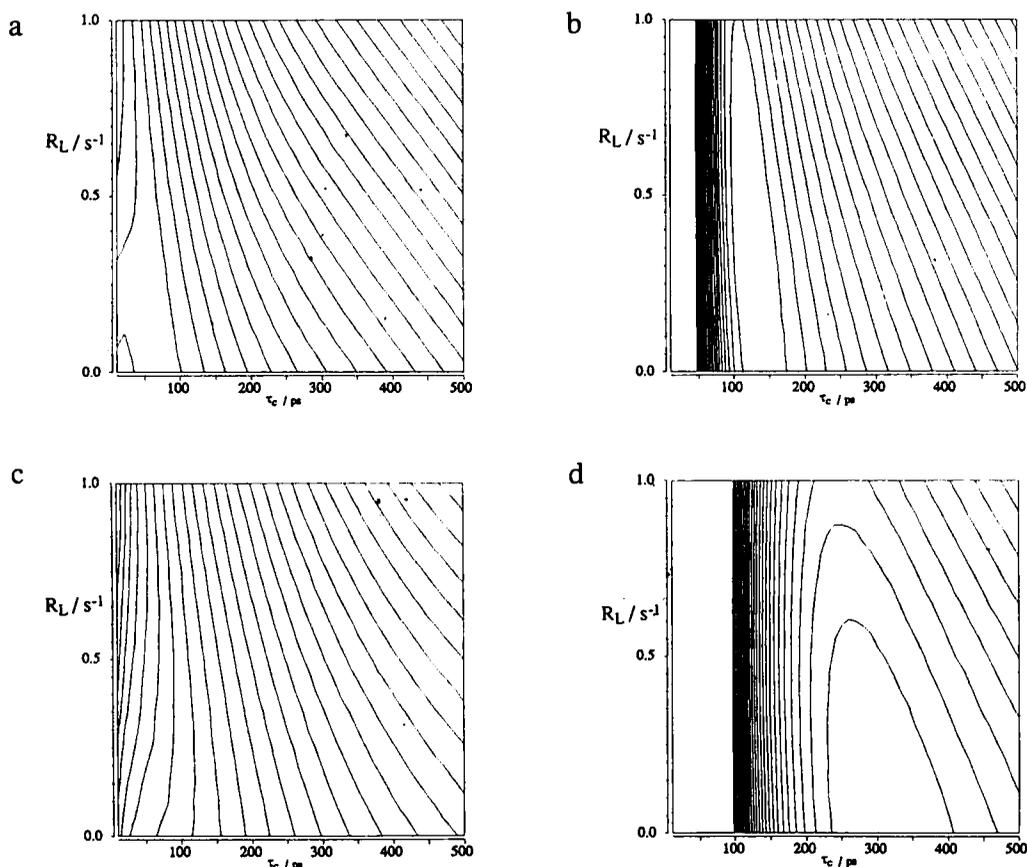


Fig. 6. CROSREL  $R_L$ ;  $\tau_c$  grid search results for a ROESY build-up series of methyl( $d_3$ )  $\beta$ -cellobioside using Obs-scaling. The iso-contour lines are separated by 0.01 in  $R_w$ . (a) Isotropic model without HOHAHA correction; (b) isotropic model with HOHAHA correction; (c) anisotropic model without HOHAHA correction; and (d) anisotropic model with HOHAHA correction.

graphical representation of the Obs-scaling grid search is given in Fig. 6. The grid search using the anisotropic  $\tau_c$  factors without HOHAHA correction yields the lowest  $R_w$  factor. Evaluation of this result revealed a danger of the Obs-scaling method. The correlation between  $R_L$ ,  $\tau_c$  and the o.s.f. is large. In the case described above, the erroneous  $R_L$  and  $\tau_c$  values are more than compensated by the o.s.f. that aims at minimizing the  $R_w$  factor. Overcompensation can be recognized by  $\tau_c$  tending to extremely low values, and/or large  $R_L$  values. The  $R_L$  and  $\tau_c$  values for the isotropic model, without HOHAHA correction, also suffer from the large correlation with the o.s.f., but without the extremely low  $R_w$  factor. The fact that the position of the  $R_L$ ;  $\tau_c$  optimum is  $\tau_m$  dependent when no correction for the HOHAHA effect has been applied is a likely cause of this problem. In the case when HOHAHA correction has been applied, the preference for the isotropical model using Obs-scaling is even more profound than for the other scaling methods. Again slight  $\tau_c$  shifts of the optima are observed, compared to  $M_0$ -scaling, with an unchanged  $R_L$ .

It is concluded that the compatibility of the scaling methods offered by CROSREL is high, so that any method can be used that best suits the available experimental data.

## CONCLUSIONS

A method is introduced that is suitable for the analysis of NOESY and ROESY peak intensities, which is implemented in the CROSREL program. The agreement between calculated and experimental peak intensities is expressed by a weighted  $R_w$  factor, with weights to attenuate the contribution of intense peaks.

The correction for direct HOHAHA magnetization transfer, although an estimation, improves the agreement between calculated and experimental intensities significantly, and is regarded obligatory for the analysis of ROESY data.

Three different scaling options are available. Expression of all peaks relative to the equilibrium magnetization  $M_0$  and  $\tau_m = 0$  ( $M_0$ -scaling) is regarded to be the optimal scaling method. The use of this scaling method, however, will often be impossible, since it requires the measurement of individual diagonal peak intensities and the acquisition of a build-up series in order to estimate  $M_0$ . The requirement of diagonal peak integration also holds for  $M_1$ -scaling, wherein all peaks are expressed relatively to the integrated column intensity for each mixing time, but one single ROESY experiment is sufficient for  $M_1$ -scaling. The third option, Obs-scaling, can be used in general for complex biomolecules, wherefore only cross peaks can be measured. Application of Obs-scaling has shown to give erroneous results, when the positions of  $\tau_m$ -specific optima are largely scattered in the  $R_L$ ;  $\tau_c$  grid. This was the case for the grid search for methyl(*d3*)  $\beta$ -cellobioside without HOHAHA correction. Comparison of the different scaling methods revealed a good agreement between the results obtained for the different scaling methods. This implies that Obs-scaling can be used for biomolecules with complex NMR spectra when correction for HOHAHA has been applied. When it remains impossible to find a good  $R_L$ ;  $\tau_c$  optimum, a reasonable, fixed  $\tau_c$  value must be used.

The implementation of anisotropy and internal motion in CROSREL seems to work correctly, although the present data do not allow a conclusion as to whether the methyl  $\beta$ -cellobioside molecule behaves anisotropically.

It has been shown that the developed full relaxation matrix approach, implemented in the CROSREL program, yields reliable, theoretical ROESY peak intensities. This allows quantitative analysis of ROESY spectra, which has long been regarded to be impossible.

## ACKNOWLEDGEMENTS

The authors thank Prof. Dr. R. Kaptein, and Prof. Dr. J.F.G. Vliegthart and Drs. A. Bonvin for valuable suggestions, Dr. A. van Kuik for the use of the plotting software, and Prof. Dr. W.F. van Gunsteren for the use of the GROMOS program package. This work was supported by the *Netherlands Program for Innovation Oriented Carbohydrate Research (IOP-k)* with financial aid from the Ministry of Economic Affairs and the Ministry of Agriculture, Nature Management and Fisheries, and by the *Netherlands Foundation for Chemical Research (SON)* with financial aid from the *Netherlands Organization for Scientific Research (NWO)*.

## REFERENCES

- Bax, A. (1988) *J. Magn. Reson.*, **77**, 134–147.
- Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **63**, 207–213.
- Bloembergen, N., Purcell, E.M. and Pound, R.V. (1948) *Phys. Rev.*, **73**, 679–712.
- Boelens, R., Koning, T.M.G. and Kaptein, R. (1988) *J. Mol. Struct.*, **173**, 299–311.
- Boelens, R., Koning, T.M.G., van der Marel, G.A., van Boom, J.H. and Kaptein, R. (1989) *J. Magn. Reson.*, **82**, 290–308.
- Borgias, B.A., Gochin, M., Kerwood, D.J. and James, T.L. (1990) *Prog. NMR Spectrosc.*, **22**, 83–100.
- Bothner-By, A.A. and Noggle, J.H. (1979) *J. Am. Chem. Soc.*, **101**, 5152–5155.
- Bothner-By, A.A., Stephens, R.L., Lee, J.-M., Warren, C.D. and Jeanloz, R.W. (1984) *J. Am. Chem. Soc.*, **106**, 811–813.
- Breg, J., Kroon-Batenburg, L.M.J., Strecker, G., Montreuil, J. and Vliegthart, J.F.G. (1989) *Eur. J. Biochem.*, **178**, 727–737.
- Clore, G.M. and Gronenborn, A.M. (1985) *FEBS Lett.*, **179**, 187–198.
- Dunitz, J.D. (1979) *X-Ray Analysis and the Structures of Organic Molecules*, Cornell University Press, Ithaca, NY.
- Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.
- Farmer II, B.T. and Brown, L.R. (1987) *J. Magn. Reson.*, **72**, 197–202.
- Fejzo, J., Zolnat, Z., Macura, S. and Markley, J.L. (1989) *J. Magn. Reson.*, **82**, 518–528.
- Gonzalez, C., Rullman, J.A.C., Boelens, R. and Kaptein, R. (1991) *J. Magn. Reson.*, **91**, 659–664.
- Griesinger, C. and Ernst, R.R. (1987) *J. Magn. Reson.*, **75**, 261–271.
- IUPAC-IUBJCBN (1983) *Eur. J. Biochem.*, **131**, 5–7.
- Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.
- Keepers, J.W. and James, R.L. (1984) *J. Magn. Reson.*, **57**, 404–426.
- Koehler, J., Saenger, W. and Van Gunsteren, W.F. (1987) *Eur. Biophys. J.*, **15**, 197–210.
- Koning, T.M.G., Boelens, R. and Kaptein, R. (1990) *J. Magn. Reson.*, **90**, 111–123.
- Kroon-Batenburg, L.M.J., Kroon, J., Leeftang, B.R. and Vliegthart, J.F.G. (1992) *Carbohydr. Res.*, submitted.
- Kumar, A., Wagner, G., Ernst, R.R. and Wüthrich, K. (1981) *J. Am. Chem. Soc.*, **103**, 3654–3658.
- Leeftang, B.R. (1991) *Conformational Analysis of Oligosaccharides*, Ph.D. Thesis, Utrecht.
- Leeftang, B.R., Kroon-Batenburg, L.M.J., Van Eyck, B.P., Kroon, J. and Vliegthart, J.F.G. (1992) *Carbohydr. Res.*, **230**, 41–61.
- Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4546–4559.
- Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, **41**, 95–117.
- Neuhaus, D. and Keeler, J. (1986) *J. Magn. Reson.*, **68**, 568–574.
- Noggle, J.H. and Schirmer, R.E. (1971) *The Nuclear Overhauser Effect*, Academic Press, New York, NY.
- Olejniczak, E.T., Dobson, C.M., Karplus, M. and Levy, R.M. (1984) *J. Am. Chem. Soc.*, **106**, 1923–1930.
- Olejniczak, E.T., Gampe Jr., R.T. and Fesik, S.W. (1986) *J. Magn. Reson.*, **67**, 28–41.
- Scarsdale, J.N., Yu, R.K. and Prestegard, J.H. (1986) *J. Am. Chem. Soc.*, **108**, 6778–6784.
- Solomon, I. (1955) *Phys. Rev.*, **99**, 559–565.
- Steele, W.A. (1976) *Advances in Chemical Physics, Vol. 34*, Wiley, New York, NY, pp. 1–104.
- Summers, M.F., South, T.L., Kim, B. and Hare, D.R. (1990) *Biochemistry*, **29**, 329–340.
- Tropp, J. (1980) *J. Chem. Phys.*, **72**, 6035–6043.
- Van Gunsteren, W.F. (1987) *GROMOS, Groningen Molecular Simulation Package*, University of Groningen, The Netherlands.
- Woessner, D.E. (1962) *J. Chem. Phys.*, **37**, 647–654.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.